

# Abstract

There has been a vast expansion of data usage in recent years. The requirements of database systems to provide a variety of information has resulted in many more types of database engines and approaches (such as cloud computing). A once simple management task has become much more complex. Challenges exist for database managers to make the best choices of practices and procedures to satisfy the requirements of organisations.

This research is aimed at understanding how the management of database systems is undertaken, how best practices and procedures form a part of the management process, and the complex nature of database systems. The study examined the adoption of best practices and how the complex interactions between components of the database system affect management and performance.

The research followed a mixed methods approach, using sequential explanatory design. The quantitative research phase, using an online survey, highlighted the breadth of issues relevant to database management. It concluded that existing practices and procedures were not optimal, and revealed some of the complexities. Based on the findings from the survey the qualitative research phase that followed utilized information from the quantitative survey to seek understanding of key areas, through a number of focus groups.

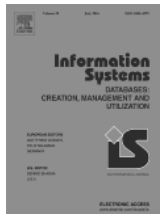
As part of this research, an innovative method was developed in which thematic analysis of the resulting data was deepened through the use of systems thinking and diagramming. Taking this holistic approach to database systems enabled a different understanding of best practices and the complexity of database systems. A 'blueprint', called a CODEX, was drawn up to support improvement and innovation

*Abstract*

of database systems. Based on a comprehensive assessment of the individual causal interactions between data components, a data map detailed the complex interactions.

## Published Papers

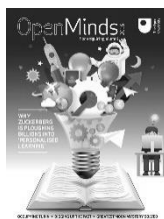
The paper(s) already published from the work are:



Holt, V. *et al.*, 2015. The usage of best practices and procedures in the database community. *Information Systems*, April 2015 49, pp.163–181. DOI: <http://dx.doi.org/10.1016/j.is.2014.12.004>

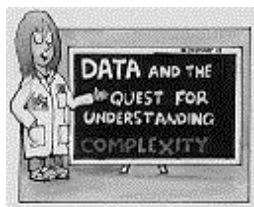
### Impact

Public communication of research ideas.



Thanks to OpenMinds magazine for publishing my creative approach, <http://bit.ly/2aD81rv> , to explain to the rest of the world how exciting and important my research is.

Thanks to Microsoft Technet UK for publishing 2 blogs about my research and my Data character. Thanks to Andrew Fryer for creating the comic strip from my storyboard.



A research précis  
<http://bit.ly/1yMvzzS>



A research journey  
<http://bit.ly/1VNaO33>

# Acknowledgements

I would like to thank everyone who has contributed to enable me to complete this thesis.



Thanks to the Association of Open University Graduates (AOUG) Trustees of the Foundation for Education for presenting the Will Swann Award for Innovation and Knowledge Development to me.



Thanks to my tutors, Magnus Ramage and Karen Kear, for all their help through the PhD and for nominating my research for the AOUG Will Swann Award for Innovation and Knowledge Development.



A special thanks to Joan Holt, David Holt and Roger Holt for their inspiration, support and encouragement throughout the PhD process.

# Contents

Abstract .....	1
Published Papers.....	3
Acknowledgements.....	4
Contents .....	5
Figures .....	14
Tables.....	20
List of Definitions .....	22
Chapter 1: Introduction .....	31
1.1 Databases Today.....	31
1.1.1 Database Systems: Issues and Problems .....	36
1.1.2 Vignette – Ecosystems of Evolving Database Landscape.....	37
1.1.3 Management Frameworks .....	43
1.1.4 Rich Picture .....	44
1.2 Purpose of the Research .....	45
1.3 The System of Interest.....	49
1.3.1 The Technical DBMS.....	50
1.3.2 People System .....	54
1.3.3 Operating Model.....	55
1.4 Research Questions .....	56
1.5 Research Approach .....	59

*Table of Contents*

1.6	Significance of the Study.....	60
1.7	Overview of Thesis Chapters .....	61
Chapter 2: Literature Review of Database Management in Practice.....		64
2.1	Introduction .....	64
2.2	Best Practice.....	64
2.3	Systems Thinking.....	71
2.4	Organizational Management .....	76
2.4.1	Decision Making .....	79
2.4.2	Culture and Conflict .....	81
2.4.3	Learning .....	83
2.5	Information Systems as Adapted to Database Management .....	85
2.6	Database Systems .....	87
2.7	Database Technical Facets.....	96
2.7.1	Architecture (Structure Pillar).....	96
2.7.2	Access and Control (Structure Pillar) .....	100
2.7.3	Maintenance (Life Support Pillar).....	102
2.7.4	Resilience and Conservation (Business Level Pillar) .....	105
2.7.5	Data (Business Level Pillar).....	106
2.7.6	Change (Improvement and Innovation Pillar).....	110
2.7.7	Forecasting (Improvement and Innovation Pillar).....	111
2.8	Database Management Lifecycle Frameworks .....	113
2.8.1	Architectural Frameworks .....	113

2.8.2	Agile Management of Databases.....	115
2.8.3	Database Management Service.....	117
2.8.4	Data Management Framework .....	119
2.9	Improvement and Innovation.....	121
2.10	Summary.....	127
Chapter 3:	Research Design.....	130
3.1	Introduction.....	130
3.2	The Research Strategy .....	131
3.3	Mixed Method Summary .....	137
3.4	Linking the Research Questions to the Research Methods .....	138
3.5	Quantitative Design.....	139
3.5.1	Quantitative Sampling .....	140
3.5.2	Triangulation for the Quantitative Phase.....	141
3.5.3	Quantitative Data Collection Method: Survey.....	143
3.5.4	Quantitative Data Analysis.....	147
3.6	Connecting Quantitative and Qualitative Phases .....	149
3.6.1	Triangulation.....	151
3.7	Qualitative Design.....	152
3.7.1	Qualitative Sampling.....	152
3.7.2	Qualitative Triangulation.....	154
3.7.3	Qualitative Data Collection Method.....	154
3.7.4	Qualitative Data Analysis.....	158

*Table of Contents*

3.8	Thematic Analysis .....	159
3.8.1	Codes Defined .....	161
3.8.2	Themes Defined .....	162
3.9	Defined Data Analysis Process and Method used .....	163
3.9.1	First Coding Cycle .....	164
3.9.2	Transitional Process .....	170
3.9.3	Synthesis: Systems Thinking .....	174
3.10	Summary .....	179
Chapter 4:	Quantitative Survey Findings on the Utilization of Best Practices ..	180
4.1	Introduction .....	180
4.2	Survey Findings .....	181
4.2.1	Demographics .....	183
4.2.2	Respondents' Organizations .....	186
4.2.3	Understanding Best Practice.....	188
4.2.4	Control of Best Practices .....	192
4.2.5	Database Demographics .....	194
4.2.6	Database Servers .....	197
4.2.7	Training .....	200
4.2.8	Database Architecture, Design and Development.....	204
4.2.9	Database Technical Practices.....	209
4.2.10	Database Operations.....	214
4.2.11	Cloud Databases .....	217



4.2.12	Data Management.....	221
4.2.13	Application Centric .....	226
4.2.14	Change Management.....	229
4.2.15	Organizational Culture.....	231
4.2.16	Improvement Methods.....	234
4.2.17	Future Vision.....	236
4.3	Connecting Quantitative and Qualitative phases.....	237
4.3.1	Importance of Best Practices .....	239
4.3.2	Selection of Database Engines.....	240
4.3.3	Requirements Gathering and Design.....	241
4.3.4	Database Lifecycle Management.....	242
4.3.5	Technical Layers .....	243
4.3.6	Managing Cloud Databases .....	244
4.3.7	Complexity Compromising Implementation.....	246
4.3.8	Creation and Control of Best Practice.....	247
4.3.9	Cross Boundary Communication .....	248
4.3.10	Strategic Planning .....	249
4.4	Summary .....	250
Chapter 5:	Qualitative Findings, from Analysis to Synthesis .....	252
5.1	Introduction.....	252
5.2	Qualitative Analysis Process.....	253
5.3	Participant Demographics.....	255

*Table of Contents*

5.4	First Cycle of Coding.....	256
5.4.1	Familiarizing Yourself with the Data.....	257
5.4.2	Generating Initial Codes and manual coding to a repository .....	257
5.4.3	Searching for Themes.....	260
5.4.4	Reviewing Potential Themes.....	269
5.4.5	Defining and Naming Themes.....	273
5.5	Transitional Process.....	278
5.5.1	Code Landscaping.....	279
5.5.2	Code Relations .....	281
5.5.3	Operational Model Diagram .....	289
5.6	Synthesis: Systems Thinking.....	291
5.6.1	Systems Map of the Codes.....	293
5.6.2	Architectural Subsystem .....	296
5.6.3	Database Management Complexity .....	305
5.6.4	Best Practice .....	307
5.6.5	Management.....	312
5.6.6	Technical .....	317
5.6.7	Requirements and Architecture.....	324
5.6.8	Understanding, Knowledge and Skills.....	331
5.6.9	Business and Change.....	335
5.6.10	Global Influence Diagram .....	341
5.7	Summary.....	344

Chapter 6: Discussion of Database Management and the Complexity of Delivering a Best Practice Solution ..... 345

6.1 Introduction ..... 345

6.2 Best Practice Usage ..... 346

6.2.1 Best Practice ..... 347

6.2.2 Database Management and Data Management ..... 349

6.2.3 Operational Management ..... 351

6.2.4 Architecture, Design and Development..... 353

6.2.5 Cloud..... 355

6.2.6 Database Engines ..... 356

6.2.7 Management Approaches..... 357

6.2.8 Summary ..... 358

6.3 Complex Interactions in the Management of Database Systems..... 360

6.3.1 Technical System ..... 361

6.3.2 Architectural Subsystem ..... 363

6.3.3 App Dev Subsystem ..... 365

6.3.4 Operational Subsystem ..... 366

6.3.5 People System ..... 368

6.3.6 Knowledge Subsystem ..... 373

6.3.7 Business System ..... 374

6.3.8 Management System..... 376

6.3.9 Summary of the Complex Interaction Findings ..... 378

*Table of Contents*

6.4	Adoption of Best Practice Affected by Complex Interactions .....	379
6.4.1	Adoption of Best Practices .....	383
6.4.2	Changing Best Practice .....	387
6.4.3	Summary .....	390
6.5	Improvement and Innovation .....	391
6.5.1	Application of Lessons from the Research .....	394
6.5.2	CODEX Blueprint.....	397
6.5.3	The Working CODEX.....	408
6.6	Summary.....	413
Chapter 7:	Research Conclusions.....	415
7.1	Introduction .....	415
7.2	Contribution to Knowledge .....	416
7.3	Summary of Key Findings .....	418
7.3.1	Best Practices and Procedures Utilised by the Database Community 418	
7.3.2	Complex Interactions of the Management of Database Systems .....	420
7.3.3	Adoption of best practices and procedures affected by the complex interactions .....	421
7.3.4	Contribution to Improvement and Innovation.....	424
7.4	The CODEX .....	426
7.5	Implications for Method .....	429
7.6	Future Work .....	429

7.6.1	Prediction using Machine Learning.....	431
7.6.2	Networks through Graph Theory.....	432
7.6.3	Interdisciplinary Data Science Complexity Prediction .....	433
7.7	Conclusions .....	434
Appendix A:	Quantitative Questions.....	437
Appendix B:	Qualitative Questions .....	447
Appendix C:	Word Frequency Count .....	448
Appendix D:	Qualitative Question Spray Diagrams .....	450
Appendix E:	Code Book Code Summary.....	460
Appendix F:	Data Map .....	462
References	.....	465

# Figures

Figure 1.1 Rich picture: database management .....	45
Figure 1.2 The system of interest as initially conceived .....	50
Figure 1.3 The technical DBMS .....	51
Figure 2.2 Components of an information system .....	86
Figure 3.1 Sequential explanatory design .....	133
Figure 3.2 Method design .....	137
Figure 3.3 Qualitative research process .....	164
Figure 3.4 First coding cycle .....	164
Figure 3.5 Format for a spray diagram .....	167
Figure 3.6 Candidate overarching themes, themes and subthemes .....	169
Figure 3.7 Transitional process .....	170
Figure 3.8 Code relations .....	173
Figure 3.9 Synthesis: systems thinking .....	175
Figure 3.10 Format for a systems map .....	176
Figure 3.11 Component influence diagram .....	177
Figure 3.12 Format for an influence diagram .....	178
Figure 4.1 Size of organization's workforce .....	184
Figure 4.2 The number of database administrators in different sizes of respondents' organizations .....	185
Figure 4.3 Experience in the field .....	186
Figure 4.4 Organization's number of database servers .....	186
Figure 4.5 People administering the databases .....	187

Figure 4.6 Time spent managing database servers ..... 188

Figure 4.7 Frequency table of best practice ..... 189

Figure 4.8 Where do you find best practice..... 190

Figure 4.9 Organization following best practice..... 191

Figure 4.10 Issues with best practices ..... 191

Figure 4.11 Control of best practices ..... 193

Figure 4.12 Who controls database choices ..... 194

Figure 4.13 Size of the single largest database ..... 195

Figure 4.14 Type of database engine used..... 195

Figure 4.15 Database applications used ..... 196

Figure 4.16 Database environments ..... 197

Figure 4.17 Database platforms used ..... 198

Figure 4.18 Service availability ..... 200

Figure 4.19 Receipt of database training ..... 201

Figure 4.20 Encouragement for taking certifications ..... 201

Figure 4.21 Type of database training received for the length of time working in the field..... 202

Figure 4.22 Opportunity for formal training..... 203

Figure 4.23 Opportunity to attend conferences ..... 204

Figure 4.24 Architecture frameworks ..... 205

Figure 4.25 Processes at the architectural stage ..... 206

Figure 4.26 Design processes ..... 206

Figure 4.27 Development methodologies..... 207

Figure 4.28 Development stage processes ..... 208

*Table of Figures*

Figure 4.29 Development methodologies and the process for requirements gathering .....	209
Figure 4.30 Servers managed.....	209
Figure 4.31 Installation and configuration.....	210
Figure 4.32 Security policies enforced .....	210
Figure 4.33 Practices and procedures for database management .....	211
Figure 4.34 Storage types used .....	212
Figure 4.35 Practice followed for database storage configuration .....	212
Figure 4.36 Practices and procedures for availability .....	213
Figure 4.37 For the platforms used is database management abstracted for the hardware layer? .....	214
Figure 4.38 IT Service management framework.....	215
Figure 4.39 Problem management method .....	216
Figure 4.40 Frequent malfunctions.....	216
Figure 4.41 Long term fixes for regular issues.....	217
Figure 4.42 Forms of cloud database usage .....	218
Figure 4.43 Cloud database service usage .....	219
Figure 4.44 The types of cloud database services used for the type of environment .....	220
Figure 4.45 Practices and procedures to manage cloud.....	220
Figure 4.46 Database software patching policy .....	221
Figure 4.47 Current practices and procedures for data management.....	222
Figure 4.48 Practice and procedures for data management.....	223
Figure 4.49 Data requirements driving database management .....	223



Figure 4.50 Transfer data between servers..... 224

Figure 4.51 Database product selection constrained by employee skillset for data requirements, driving database management ..... 225

Figure 4.52 Legal procedures followed for data with historical data policies..... 226

Figure 4.53 Practices and procedures for the main application ..... 227

Figure 4.54 Different management practices for different database products..... 228

Figure 4.55 Practices and procedures for big data ..... 229

Figure 4.56 Practices and procedures for change management ..... 230

Figure 4.57 Approximate database changes a week..... 230

Figure 4.58 Changes not following policies and procedures ..... 231

Figure 4.59 Communication and business practices ..... 232

Figure 4.60 Working team practices..... 233

Figure 4.61 Database management visibility..... 233

Figure 4.62 Database product practices ..... 234

Figure 4.63 Do you have an improvement method to follow for database management ..... 235

Figure 4.64 Improvement method to follow for database management ..... 235

Figure 4.65 Business vision of database management ..... 237

Figure 5.1 Qualitative analysis process..... 254

Figure 5.2 First coding cycle ..... 256

Figure 5.3 A subset of the initial raw data for Best Practice within Questions 1 and 2 ..... 259

Figure 5.4 Spray diagram of Question 1 codes and themes from the data corpus 261

Figure 5.5 Distribution of codes within all 10 questions with greater than 1 occurrence ..... 264

*Table of Figures*

Figure 5.6 Best practice example thematic map.....	271
Figure 5.7 Transitional process .....	278
Figure 5.8 Word cloud created from the entire qualitative text (data corpus) .....	279
Figure 5.9 Code relations.....	285
Figure 5.10 Operational model diagram .....	290
Figure 5.11 Synthesis: systems thinking .....	292
Figure 5.12 Systems map of the management of database systems.....	293
Figure 5.13 Technical system .....	295
Figure 5.14 Architectural subsystem .....	296
Figure 5.15 Application development (app dev) subsystem.....	298
Figure 5.16 Operational subsystem.....	299
Figure 5.17 People system.....	300
Figure 5.18 Knowledge subsystem .....	301
Figure 5.19 Business system .....	302
Figure 5.20 Management system.....	303
Figure 5.21 Best practice influence diagram .....	307
Figure 5.22 Management influence diagram .....	312
Figure 5.23 Technical influence diagram.....	318
Figure 5.24 Requirements and architecture influence diagram.....	324
Figure 5.25 Understanding, knowledge and skills influence diagram.....	331
Figure 5.26 Business and change influence diagram .....	336
Figure 5.27 The database system influence diagram consolidated .....	342
Figure 6.1 The technical system .....	363

Figure 6.2 The architectural subsystem .....	365
Figure 6.3 The app dev subsystem .....	366
Figure 6.4 The operational subsystem .....	368
Figure 6.5 The people system .....	373
Figure 6.6 The knowledge subsystem .....	374
Figure 6.7 The business system .....	375
Figure 6.8 The management system.....	378
Figure 6.9 The whole database system influence diagram.....	381
Figure 6.10 Spray diagram of a business vision of database management.....	394
Figure 6.11 Most prevalent components .....	400
Figure 6.12 The systems map incorporating data from the code relations chart and showing (by colour) the CODEX groupings.....	401
Figure 6.13 The database system CODEX .....	404
Figure 6.14 Codex component linkage example of three inputs scenarios .....	411
Figure 7.1 Graph of component complexity.....	431

## Tables

Table 3.1 Aspects considered in planning mixed methods design.....	135
Table 3.2 Research questions and research method .....	138
Table 3.3 Likert Scale .....	146
Table 3.4 Survey question types .....	146
Table 3.5 Qualitative sampling selection used .....	152
Table 3.6 Focus group demographics .....	157
Table 3.7 Phases of thematic analysis .....	160
Table 3.8 Descriptive code summarizing the primary topic.....	162
Table 3.9 Adapted coding strategies and methods.....	163
Table 3.10 Interconnections and prevalence.....	173
Table 4.1 Usage on-premises database software .....	198
Table 4.2 Cloud database services used.....	199
Table 5.1 A data item from the data corpus.....	257
Table 5.2 Code book early generation of codes and themes.....	268
Table 5.3 Combined themes for the best practice example .....	270
Table 5.4 Data map - An example displaying influences between components from the data corpus .....	283
Table 5.5 Data map top influences.....	284
Table 5.6 Findings from code landscaping, data map and code relations.....	288
Table 5.7 Data map systems summary showing the total number of interconnections .....	289
Table 5.8 Total number of occurrences of influences .....	343

Table 6.1 Most important areas to add improvement for database management .	391
Table 6.2 Summary from Table 5.13 with the quantitative results. ....	399
Table 6.3 Creation of the CODEX.....	403
Table 6.4 Data input features.....	409
Table 6.5 Component influences from the scenario .....	410

## List of Definitions

This section provides an explanation of terms used in the thesis.

<b>Term</b>	<b>Definition</b>
Actors	The people who carry out the activity in the system.
Application Centric Management	Area focusing on the application at the foundation and the control of the fundamental parts that require resources and services for the architecture.
Best Practices	The working definition in this thesis for best practice is: a recommended practice for carrying out actions for desirable outcomes, rather than always being the best way of doing something.
Big Data	A general term used to describe the large volume of unstructured and semi-structured data that cannot be processed using conventional methods. It could be complex data.
Blueprint	The reproduction of the database system, documenting the complexity of the architecture, the technical components, business and people utilizing technology. A blueprint allows a rapid and accurate documented reproduction to be created.

Term	Definition
Capta	Data that is relevant and is to be collected. Capta is the results of collecting selective data for attention.
Cloud Computing	“Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.” (Mell & Grance 2011a).
Code	A short word or phrase used to describe part of the data.
Code Landscaping	<i>Code Landscaping</i> integrates textual and visual methods to see both the forest and the trees [...] Internet tools such as <i>Wordle</i> ( <a href="http://www.wordle.net">www.wordle.net</a> ) enable you to cut and paste large amounts of text into a field.
Complex interactions	Complex interactions for the purpose of the research are three or more interactions that are linked with a single task or component.
DBMS	Database management system (DBMS) is defined by Date as “ <i>Basically, it is nothing more than a computer record keeping system: this is, a system whose</i>

List of Definitions

Term	Definition
	<p><i>overall purpose is to record and maintain information.</i></p> <p><i>The information concerned can be anything that is deemed to be of significance to the organization the system is serving – anything, in other words, that may be necessary to the decision-making processes involved in the management of that organization.”</i></p> <p>(1981, p.3).</p>
Data	Factual information, facts.
Data Corpus	Refers to all data collected in the research project.
Data Extract	This is an individual coded chunk of data, which has been identified within the data item.
Data Item	An individual unit of data (e.g. an interview).
Data lineage	The data life cycle that includes where the data originates and where it is moved over time.
Data Set	All the data items from the corpus that are used in the analysis.
Database as a Service (DBaaS)	An architectural and operational approach enabling IT providers to deliver database functionality as a service to one or more consumers - defined by Oracle (2011, p.5).



Term	Definition
Database Engine	<p>“A database engine (or storage engine) is the underlying software component that a database management system (DBMS) uses to create, read, update and delete (CRUD) data from a database. [...]</p> <p>The term "database engine" is frequently used interchangeably with "database server" or "database management system" (Wikipedia 2016)</p>
Database Management	<p>To ensure the data stored in the database is maintained, optimised for performance, available when required and secure.</p>
Database System	<p>The larger holistic system containing the software called a database management system (DBMS), application centric components, technical features, cultural factors and current paradigms.</p>
Digraph	<p>A directed graph (or digraph) is a graph that has a set of vertices that is connected by edges.</p>
Explanatory research	<p>The term <i>explanatory research</i> implies that the research in question is intended to explain, rather than simply to describe, the phenomena studied.</p>
Graph Database	<p>A Graph Database is a type of NoSQL Database E.G. Neo4j and Graph Engine.</p>

List of Definitions

<b>Term</b>	<b>Definition</b>
Graph	Graph or digraph is a visual representation of a social network, where actors are represented as nodes or vertices and the ties are represented as lines, also called edges or arcs.
Hadoop	A framework for processing large distributed data sets.
Key-Value	A Key-Value Database is a type of NoSQL Database E.G. Apache Cassandra, HDFS, Apache HBase, Voldemort and Dynamo.
Knowledge	Is what is learnt, larger, longer living structures of meaningful facts.
Information	What is said or recorded, meaningful facts.
Improvement	A thing that makes something better or is better than something else (Oxford University Press 2016).
Innovate	Make changes in something established, especially by introducing new methods, ideas, or products (Oxford University Press 2016).
Influence Diagram	An influence diagram represents the main structural features of a situation and the important relationships that exist among them. It presents an overview of

Term	Definition
	areas of activity, their main interrelationships and it is used either to explore those interrelationships.
In Vivo Coding	The practice of assigning a label to a section of data, such as an interview transcript, using a word or short phrase taken from that section of the data.
On Premises	In house stand-alone software installation.
Operational Model diagram	An abstract and ideally visual representation (model) of how an organization delivers value to its customers or beneficiaries. The elements that make up an operating model are often people, process and technology.
Platform	The platform consists of the hardware (such as CPU, RAM, Storage), whether it is Physical, Virtual or Cloud, the Operating System and all other components that are not the database engine.
Practices	A repeated exercise in or performance of an activity or skill so as to acquire or maintain proficiency in it (Oxford University Press 2016).
Procedures	An established or official way of doing something and methodology a system of methods used in a

List of Definitions

Term	Definition
	particular area of activity (Oxford University Press 2016).
Processes	A series of actions or steps taken in order to achieve a particular end (Oxford University Press 2016).
Management	The application of skill, or care in the manipulation, use, treatment, or control (of things or persons), or in the conduct (of an enterprise, operation) (Oxford University Press 2016).
Mixed Method Research	<i>Mixed methods research</i> includes the mixing of quantitative and qualitative data.
Node	In mathematical graph theory a vertices or node is the fundamental unit of which graphs are formed.
NoSQL	Was originally for a non-relational database whose data storage mechanism is not relational.
Relational Database	A collection of data items organized as a set of formally-described tables from which data can be accessed or reassembled in many different ways without having to reorganize the database tables.

Term	Definition
Service Management	<p>“A set of specialized organizational capabilities for providing value to customers in the form of service”</p> <p>Cartlidge <i>et al.</i> (2007, p.6).</p>
Spray Diagram	<p><i>Spray diagrams</i> are mainly used for representing the structure of an argument, to encapsulate the relationships between the ideas of others or for note taking.</p>
Structured Data	<p>Generally resides within relational database tables. It is organised data that has an identifiable structure.</p>
Synthesis	<p>Putting things together.</p>
Systems Map	<p>This is a visual representation of a system. A systems map is a snapshot of the system and environment at a point in time.</p>
Thematic Analysis	<p>A qualitative analytic method for: ‘identifying, analysing and reporting patterns (themes) within data. It minimally Organises and describes your data set in (rich) detail. However, frequently it goes further than this, and interprets various aspects of the research topic. (Braun &amp; Clarke 2006, p.79) In summary searching across a data set to find repeated patterns of meaning.</p>

*List of Definitions*

<b>Term</b>	<b>Definition</b>
Theme	A theme captures something important about the data in relation to the research question and represents some level of patterned response or meaning within the data set (Braun & Clarke 2006, p.82).
Unstructured Data	Objects that have no identifiable structure. They can be textual, images, audio, video, graphics, social media messages and other types.
Visualization	Creating images or diagrams to communicate results.

# Chapter 1: Introduction

## 1.1 Databases Today

Database management is the set of administrative tasks associated with the storage, modification and retrieval of data held within a Database Management System (DBMS). Organizations require impeccable database management in order to maintain a high quality of data, and for that data to be secure and available whenever and wherever it is required. The data for governments, banks, financial institutions, health organizations and other types of organizations must also satisfy statutory legal requirements.

A DBMS has long been the principle technology used by organizations to store data and the technical layers discussed later are based on this design. The DBMS is a central part of the holistic Database System. The term 'Database System' referred to throughout this research is a larger holistic system containing software, technical features, cultural factors, paradigms and application centric components. Today there are a number of DBMS available, each of which has certain additional features. Although each works automatically, the database administrator (DBA) needs to be able to make an informed choice and to be aware of the advantages and limitations of each, also to follow best practice in usage for trouble free operation.

Database management has evolved over the last five decades since the first functioning prototype DBMS (Haigh 2012) to become an integral part of most organizations' business. Global organizations today cannot operate without a functioning database. To assist with the management of these databases, DBMS

vendors and influential people in the field have specified best practices to be followed for many areas. These best practices come in many forms, from whitepapers and scripts for database configuration or application deployments to blogs. Organizations increasingly realize the value of the data that they hold and are beginning to draw more benefit from its analysis and mining (Anon 2010). Some examples of usage include shopping history to predict purchases (Cuddeford-Jones 2013), social media to predict trends (Schoen *et al.* 2013) and disaster relief systems using distributed data management systems (Gao *et al.* 2011).

The trend of rapidly increasing data volumes is being driven not only by the requirements of government and business to store more information, but also by media organisations digitising film and television for use by individuals in the home. Unstructured data, objects that have no identifiable structure e.g. textual, images, audio, video, were not previously considered within the database community or in the DBMS; however this view has now changed within the industry.

International collaborative science projects, such as the Large Hadron Collider sensor data from CERN (Segal *et al.* 2000), and the storage of the human genome to aid medical research (Ballew *et al.* 1998), generate large volumes of scientific, astronomic, and meteorological data, as do other data intensive scientific discoveries (Hey *et al.* 2009; Hensley *et al.* 2014). The volume of data is estimated to be almost doubling in size every two years and to increase tenfold from 4.4 zettabytes in 2014 to 44 zettabytes in 2020 (Turner *et al.* 2014). This is both structured and unstructured data. The term 'Big Data' has come into common use to refer to very large volumes of data which have variety (different types of data), are subject to velocity (the speed of constantly changing data) and have value (the process of discovering hidden value) (Hashem *et al.* 2015, p.100).



The Lowell Report, a summary of a gathering of academic database researchers' discussions on the state of database research, stated that "*Database needs are changing, driven by the internet and increasing amounts of scientific and sensor data.*" (Abiteboul *et al.* 2005, p.111). Thus database changes and increasing data volumes (Abadi *et al.* 2016) that are created from new scientific tools and the management, migration and consolidation of manual records and systems, means those methods that were used to manage small volumes of data will no longer work effectively for large volumes. As an example, to back up a small (500MB) database or tune indexes may take five minutes, but to back up and tune indexes for a large (500GB) database may take over 24 hours. During these administration tasks users still need to be able to work without the processing power being reduced to an ineffective level. New and changing operating models add complexity to an already complicated activity.

Changes in culture and everyday life have brought about the sharing of more information, and this has radically changed the usage of databases (Mckendrick 2015). Rapid improvements in technology, hardware and software result in systems subject to continuous, rapid advances.

The DBMS itself is constructed of many components which can be considered to form a layered technical system. The layered technical system can result in disparate organizational teams managing each layer independently. These teams have different sets of goals, with a variety of approaches which mean that many problems in operation can occur due to the interconnections. Many organizations demand low cost infrastructure without jeopardizing functionality or operational ability. This in turn has increased the challenges of overall management and has added to the complexity of the system.

Database architecture, design and development are the foundation of any well designed DBMS. The database is continually evolving and adapting to the demands of the users, organizations and the global environment. Also, the cost of managing databases has escalated in line with the increase in the amount of data being stored and the complexity of the tasks. A task to troubleshoot database performance can be complicated by its geographical location, the storage tier, the database engine vendor provider, the type of workload, the volume of data and whether the system is interconnected to other systems. Better IT and organizational performance has been seen through the implementation of DevOps technique (IT Revolution 2015). DevOps is short for development and operations and the IBM Corporation (2014) define it as follows:

*“DevOps is an essential enterprise capability for continuous software delivery that enables organizations to seize market opportunities and reduce time to customer feedback”*

DevOps enables the adoption of agile, lean practices through automation tools and seeks to enhance collaboration between operations and development teams.

The introduction of new technologies to try to address the increasing volumes of data is coming to the fore. The emergence of cloud computing and virtualisation which use shared resources provides a financial incentive for organizations to change the way data and databases are managed. The volume of data leads to a need to deliver content, extract, transform and load data, validate data, provide storage, process experimental empirical observations, and secure data in databases. The data stored in the database is critical to organizational aspects which require database management methods to grow and evolve. The scope of heterogeneous data and the new philosophy of managing ubiquitous data-driven

environments, changes the current requirements for database management. To meet the needs of the public, users and organizations, effective lifecycle management of the component parts of the database is required. The characteristics of efficiency, resilience, access control and persistence exemplified in Silberschatz *et al.* (1991; 1995) are intrinsic to the nature of database management. The emergence of new technology, changing organizational goals, company culture, technologists' views on which technology to use and what management techniques to follow, vendor application centric views where the vendors want to promote the use of a particular configuration or a part of their application, current management practices, technical layers, database innovation and database administrators in the database community are a few of the components which all form part of the complex system.

The management of the database and the data contained within it are often, but not always, undertaken by different teams. Two separate functions were identified by Kahn (1983, p.794) as database administration and data administration. However, they have many interconnected components. Organizations' business requirements for data collection and manipulation appear to be driving what sort of database management is required. The fields are gradually merging, and Mullins (2012) proposed data administration practices and procedures to address this, arguing that *"when database administration is treated as a management discipline, the treatment of data within your organization will improve"* (Mullins 2012, p.9).

It is common that certain practices and procedures are recommended for use in a particular sector of industry, for example, banking. These recommendations aim to provide a good, reliable standard for use in operations and to give confidence to customers. When practices and procedures are set for each new business, it is of value to explore whether they are followed, whether their usage is documented and

whether these practices and procedures are sometimes found to be too unwieldy in practice. Best practices and procedures are intended to suggest a way of controlling data to fulfil requirements expediently. However, the research presented in this thesis questions their effectiveness.

The working definition in this thesis for best practice is: a recommended practice for carrying out actions for desirable outcomes, rather than always being the best way of doing something. Best practices are defined by the owners of a particular task. In the end-to-end management of database systems there can be many stakeholders who set these best practices and thus conflict may arise from their different perspectives.

### **1.1.1 Database Systems: Issues and Problems**

Complexity science, as defined by Johnson (2009, p.3), is the study of the phenomena which emerge from a collection of interacting objects. The interaction within the DBMS and external factors form the holistic Database System that makes this a complex environment. This complex environment with split responsibilities has caused many well publicised problems. Some examples of government IT disasters involving databases are:

- Fire control, intended to replace 46 fire control centres in England with 9 regional sites – scrapped in 2010
- In 2006 an electoral register database was intended – cancelled in 2011
- In 2002, an NHS National programme to connect 30,000 GP records with 300 hospital records was proposed and didn't proceed as intended, with data security risks and ever spiralling cost.

Some key findings for the NHS National programme, suggested by Maughan (2010), were a lack of good consultation with stakeholders, lack of time at the start

to fully consider how it was to be implemented, additional requirements being added at a later stage and the use of different service providers.

An early report by Blasis (1977) highlighted problems in DBA teams with administration, organizational issues, new technology introduction, control and technical configuration.

### **1.1.2 Vignette – Ecosystems of Evolving Database Landscape**

The following vignette illustrates some of the possible complications experienced by a fictitious database manager placed in a new hypothetical situation. The vignette is derived from numerous discussions with people working in the field.

A database manager (DM) was hired to help with the evolution of the business strategy.

The IT director welcomed the DM explaining the challenge the business faced was that the infrastructure was not performing well, running out of storage space, and the application technology was outdated and costly to maintain.

The DM was assigned the task to migrate from an old database system to a new system as part of a larger IT system of change. The DM became acquainted with the Database Administration (DBA) team and was keen to understand their current way of working, their best practices, procedures and processes. The team also explained the challenges they have while the company was undergoing the IT system of change which planned to update the current infrastructure and ecommerce applications to reduce the cost of IT provision. The DBAs complained about the lack of consultation and disagreed with the business on the order of system change.

Next on the DM's agenda was to understand the exact requirements of the migration project. The project was to organise transition between systems, with minimum or no downtime and no data loss. From the details shared by the team, this agenda could be

rather hard to achieve. The DM asked questions about a lookup website and an analytics system which it appeared had been forgotten, as they needed connectivity from the ecommerce application.

The platform for the new system had been already selected by the business. The business chose the Cloud as the new platform to save money on hosting physical servers. The DM was keen to locate any technical, functional and non-functional requirements collection to validate the choice of infrastructure. The DBAs thought there was a document drawn up by another team but had not seen the proposal. The DM communicated with the solution architect who had provided an outline architectural design which, although it covered the application design well, had little reference to the supporting database. Key technical details had not yet been distributed amongst the teams. For the database servers it was key to understand the requirements, to know what physical attributes the infrastructure required, the application usage, the storage required and any software version and edition choices.

The DM was disappointed to have no control or input into the new system chosen, although keen to share knowledge and skills and avoid failures such as those described above, which could be very serious for the whole business. The mention of the Cloud worried the DM with regard to security of the data and meeting legislative requirements. The DM raised the cloud security issues with the IT manager. The primary data was held on a system they did not control and to protect it to a certain level meant selecting a certain cloud package. If the data was lost this could affect the operational ability or result in a loss of data reputation. The DM was concerned also about recovery of the data on a system they did not control. This had to be checked carefully, although the DM understood the benefit of using the cloud for scale out, predictable performance, availability and near-zero maintenance. The relational database-as-a-service was also easily accessible.

The business wanted to use Windows SQL Azure Database. The option the business wanted to use was the Basic level designed for light transactional workload. The business as it grows would consider upgrading to standard, mid-level predictable performance transaction rates with the added business continuity feature but only if it was actually needed. The developers wrote and deployed the new application code. The application was crashing intermittently and investigation was undertaken. During testing it worked fine every time. The testing was carried out on local servers. The application code was coded in an older style and not designed for a cloud based application where it needed to be coded with retry logic. The DM raised this issue with the development team and their manager who complained there was not enough time to rewrite some of the application. The development team manager raised this issue to the business about the lack of notice from the DM to fix the issue.

The second part of the solution was the integration of another system which was hosted by another provider on a virtual machine (VM). The VM required network access and agreement from the networks team that the traffic volume would not cause any other business issues.

At this point the IT director insisted on the need for speed to get the new system up and running before the year end. This resulted in team conflict and exacerbated low staff morale caused by the continual time pressures and high level of stress.

Shortly after the system went live an incident occurred on the live finance system. The data imported from the cloud system to the on premises system and external VM had caused the system to run out of space. The alert thresholds were configured incorrectly for the combined data volume. Emergency action was taken by speaking to the Storage team and the Systems Administration team to get permission to extend the disks to increase the storage size. The physical storage disks where the virtual disks resided

were full so the virtual disks could not be expanded without being migrated to larger physical storage disks. Once that was undertaken the virtual disks were expanded. On investigation it appeared that the incident was caused by lack of capacity management at either the Storage Area Network (SAN) level or at the server virtual level disks. The alert level thresholds for the monitoring were also set at a level which would result in late notification. The support teams did not know about the connectivity as the new set up was not documented. The very poor level of communication between the teams had been causing many problems. Also the poor level of staff skills in the new virtualizations and cloud technology had become very evident.

Later the business wanted analytics to monitor what sales had been made. There was no further budget or plan to purchase more on-premises servers or cloud services so sharing with another system was the only option. This was configured. DBAs suggested sharing existing servers which were used solely by another department. The other department was not told of this, due to a quick fix which was delivered to the business at speed. The other department was then told the additional databases were added to their servers. They complained to the business that customers had been affected by the slow running servers and that security could be compromised having internal and external facing databases on the same server. More resources could not be added immediately as the server and storage team were not asked for the extra resources in advance. They needed to purchase more resources. The security team was asked to carry out an audit and they made recommendations to add encryption, which was unplanned work to fit in by both the DBAs and developers. The performance of the server degraded further and other departments complained. The results were taking too long to process. A few days later there was the realisation that this new analytics system had to be available as part of the business continuity plans. This requirement was missed. The databases on the database server were currently protected by a high availability technology (HA) at the



application tier rather than the storage tier, due to the location and type server. This did not meet the business requirements which were for geo-locational HA. The technical teams realised this but the business did not stipulate this as a requirement so the technical teams did not mention it.

The DBAs with Business Intelligence (BI) team members, were to manage the data migration. This required scripts to be written and checked into a source control system; also functional and regression testing needed to be carried out. All customers needed to be informed when parts of the system were unavailable or being updated, which could partially affect service. The customers needed to test the application in advance. A few customers complained afterwards that some of their data was not correct. The application data quality only went through minimum checking. The DM said he was responsible for the hardware not data. It turns out there was no data governance and the customer had not actually completed the testing thoroughly.

Once the applications were in operations it appeared there were some more configuration issues required. The operations team did not receive adequate hand over and training on the new technology and caused an outage whilst making the changes. There was a continual need to update and change the DBMS over time. This would need to be addressed. Training to improve staff skills was also of key importance for best operational efficiency.

The vignette raises numerous issues:

- Poor communication between actors (the people who carry out the activity in the system)
- Technical problems due to interconnectedness not known by other technologists and lack of suitable procedures – education and managers not up to date

- Variety of people involved
- Organizational issues due to time pressures on actors
- Initial specifications not clear – poor communications between customers and the organization
- Customer changing requirements
- Security insufficient due to developer's limited knowledge of requirements and what is possible
- Merging problems and change not understood by management at an early stage
- Financial pressures on purchasing hardware and staff costs
- Organizational culture within each group
- The quality of the work completed by the administrators
- Poor clarity of who controls the system
- Lack of planning, testing and design
- The adoption of new technology required changes
- The increasing volumes of data and type of data change management processes
- Resource and capacity planning was not undertaken
- Staff needed to learn new skills

These types of issues exist and are continually multiplying. There exist further problems that can occur: errors created by those who have access to the data, current and new practices within the database community and environmental factors such as government legislation.

### 1.1.3 Management Frameworks

A number of management frameworks have been drawn up to assist with the setting up and support of new IT systems. Architectural frameworks are overarching and can be used in developing a wide variety of systems to suit different organisations. They provide standards for planning, designing and other general aspects of IT systems. An example of a widely used framework is the Open Group Architectural Framework (TOGAF). TOGAF is a framework that can encompass infrastructure, processes and information technology services for designing, planning, implementing, and governing an enterprise architecture across multiple groups.

Scrum and Kanban for database development and database management are agile methods that offer strategies to be used in practice to manage these areas. Scrum is agile software development that has small teams working together on a predefined set of tasks assigned to a sprint period. Kanban is a just in time technique for managing the software development process.

Service Management Frameworks for managing operations in database systems are also widely used. They offer many aspects of control through standard processes which can be used by a wide variety of organisations. The IT Infrastructure Library (ITIL), is widely adopted throughout the world and is a framework for best practices within IT service management. It specifically focuses on governance, service delivery and continual improvement of services. ITIL offers knowledge to organisations that certain procedures need to be followed but the specifics of what tasks need to be completed at the database level are unclear. The stated working definition in this thesis for best practice is a recommended practice for carrying out actions for desirable outcomes, rather than always being the best way of doing something. The DBAs need to manage databases at a granular level,

assessing components within the system. This can affect database systems and have a profound effect on how they are managed.

The use of appropriate practices and procedures can have a significant impact on the availability, recoverability and quality of data used in the operations of businesses. The diversity of an organization's fields of operation, strategies and practices can lead to a variety of practices and procedures.

Certain practices can be considered best practice. Best practices are frequently described as recommended practices for carrying out actions for desirable outcomes. Best practices drive operational excellence and effectiveness (Dembowski 2013).

The complex landscape presented relating to the management of database systems could be improved with a better understanding of the best practices and procedures that are utilized by the database community. However, that raises the question as to whether the adoption of best practice is constrained by the many interactions between the interconnected aspects of the management of database systems.

#### **1.1.4 Rich Picture**

The messy complicated human system described above can be depicted through a representation called a 'rich picture'; a technique developed by Checkland (1999).

The rich picture contains pictorial symbols of the situation, cartoons, sketches and relationships that represent the situation as seen from the artist's point of view. The rich picture can be used to illustrate relationships, connections, influences, cause and effect; it can provide insight. Other elements such as character, points of view and prejudices may be shown (Reynolds et al. 2014a). The rich picture in Figure 1.1 helps to show the potential complexities of database management. Monk and Howard (1998) argued that the use of the rich picture encourages user centred

design that focuses attention and is an informal versatile technique to serve as starting point for design processes. Figure 1.1 shows the technical component of the DBMS, the design, the usage, the management, the frameworks, the data and the storage provision. Each of the components has many parts and Figure 1.1 helps to show that it is not just the technical provision but the people and culture that are a part of database management.

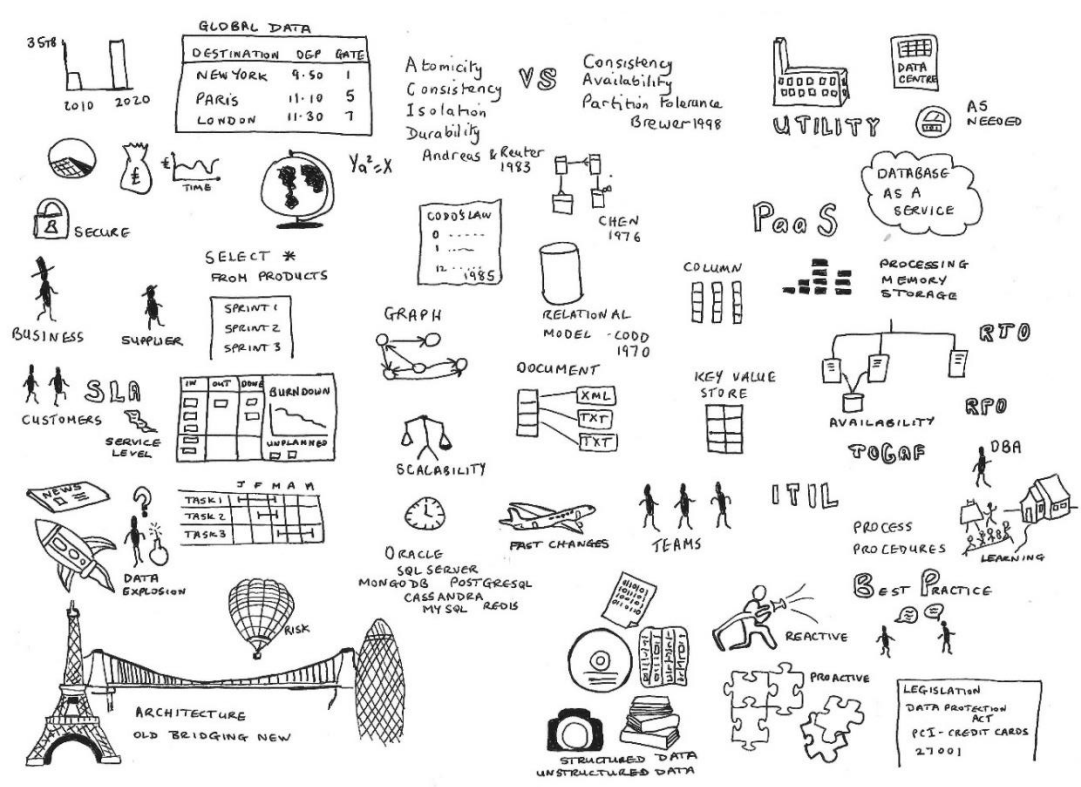


Figure 1.1 Rich picture: database management

## 1.2 Purpose of the Research

The main purpose of this research is to investigate whether there are ways to improve and innovate in the management of database systems. There are many current complexities related to data management and data administration, and as Aiken *et al.* (2011) suggest, data management is still evolving. The Claremont report (Agrawal *et al.* 2009, p.65) on database research highlighted concerns regarding

the increasing technical scope, processes and keeping track of the field that is important to the community. Other surveys previously undertaken highlighted the rise of database administration, with an unclear direction of the future path (McCririck & Goldstein 1980; Gillenson 1982; Gillenson 1985; Gillenson 1991; Aiken *et al.* 2011; Mckendrick 2013).

A way of investigating complex systems is through systems thinking. Systems thinking, as defined by Checkland (1999, p.318), is:

*“An epistemology which, when applied to human activity is based upon the four basic ideas: emergence, hierarchy, communication, and control as characteristics of systems. When applied to natural or designed systems the crucial characteristic is the emergent properties of the whole.”*

In order to look at the overall database management system, adopting a systems approach would be advantageous. A systems methodology identifies the parts of the system as interdependent and inter-connected. The complex interactions of the components within the DBMS and external factors are an integral aspect of the management of the database system. Johnson (2009, pp.3–4) discussed complexity science as *“the study of the phenomena which emerge from a collection of interacting objects”*. These interactions have emergent behaviour and may be competing for resources that need management. Johnson (2009, p.4) calls this part of *“complexity in action”*. Complex interactions for the purpose of the research are three or more interactions that are linked with a single task or component.

The management of database systems covers not only the limited technical management but the whole system. The characteristics of systems, as defined by Von Bertalanffy (1969) and later Checkland (1999) are that the recognisable whole

consists of multiple components. As Schein (1980) stated, there is a need to look at relationships between systems and their environment. The environment is a conceptual area outside the system boundary which may affect the system.

Database management is affected by the environment. Database management should consider the pluralistic environment, diversity in people, technology and processes. This complexity potentially creates many outcomes and variants when managing database systems, with potential emergence in the chaotic system, as defined by Gleick (1998).

The scope of this research is to explore the database management, practices, procedures and interactions, and not to carry out a technical study relating to the various pieces of software in the market.

The key significance of this research is:

- understanding what practices and procedures are used for managing databases
- understanding the interactions between the components, the relationships of the complex parts of the system and how these affect the management of database systems
- reflecting on these outcomes to identify possible improvements in the system which may result in the emergence of innovation

The emergence of a new method or change in process could be thought of as innovative. Over the years numerous suggestions for improvements in database design and operation have been made and new tools and methods of operation have been devised. However various complex issues remain, with the potential to produce chaotic results.

Lewin (1993) argues that complexity science has shown organizations to be complex adaptive systems:

*“we have seen, complex adaptive systems are composed of a diversity of agents that interact with each other , mutually affect each other , and in so doing generate novel, emergent , behaviour for the system as a whole”* Lewin (1993, p.198).

This richness of interactions, Waldrop (1992) states, leads to the systems as a whole undergoing self-organization in a spontaneous manner. Self-organization may occur to some extent in database management systems. Bullock and Cliff (2005) discuss the interconnected complexity of systems that Information and Communication Technology (ICT) practitioners create and their inability to predict emergent interactions of components. Capra and Luisi (2014) add that human organizations have two types of structures: designed and emergent. Designed structures provide rules whereas emergent structures provide novelty, creativity and flexibility.

Improvement and innovation within database management would allow the ‘adoption of new practices by people in a community’ Denning (2002, p.313).

Utterback (1996, p.18) suggests using old capabilities to create innovation, claiming that there are rarely any new ideas, just modifications of known concepts and ideas.

Once a database is in operation people may not interact directly, but instead interact remotely though the medium of the database. The database structure and resources must be sufficiently capable of providing an operational system.

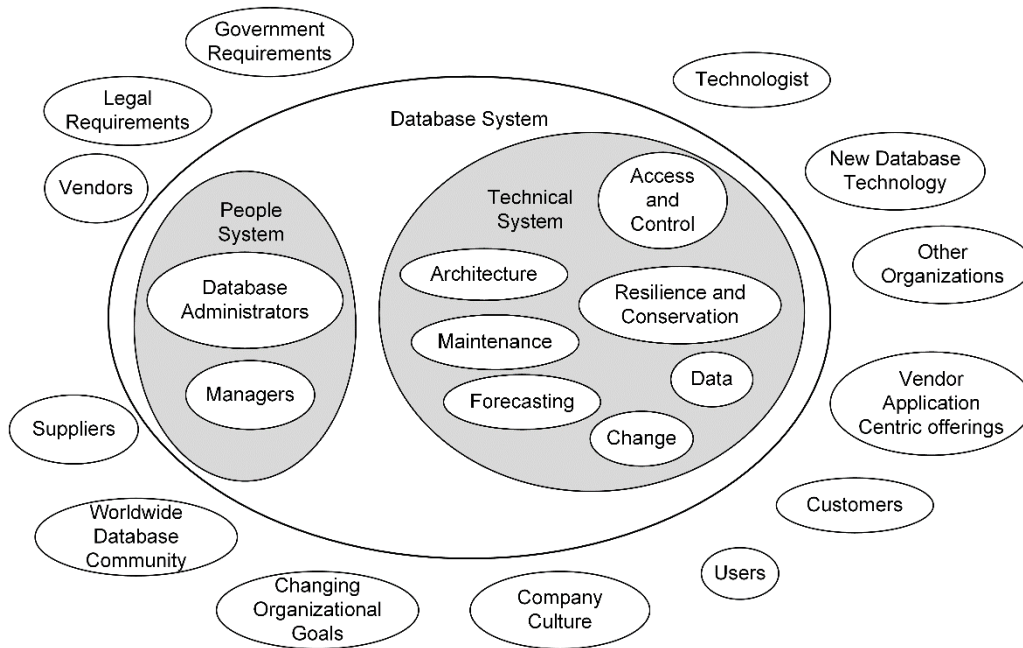
Communication between stakeholders should take place at the planning and documentation stages. Moreover the actors affect the outcomes and development.



This study is multi-disciplinary, looking at management of database systems, organizational management, socio-technical issues and complex systems. The experiences of people working in the field are key to understanding more fully the operation of a database system. There is no one study or group of studies, to date, that cross the database technical field areas and the organization and systems operation. Artus (2008) saw database research as complex and above all 'an interdisciplinary enterprise'.

### 1.3 The System of Interest

The parts of the database system defined in this research are the actors (the people who carry out the activity in the system), the organization and culture, the vendors (proprietary software designers), the hardware, the database tools for management, the DBAs and the technologists in connected fields. The components outside of the system (the environment), all have an influence on the system. The interactions between the internal and external community also affect database management. Figure 1.2 shows a holistic high level view of the database system. The boundaries in the context of this research have initially been placed around: the technical components of the database system; and the principle people involved in database management. These subsystems form the database system.



**Figure 1.2 The system of interest as initially conceived**

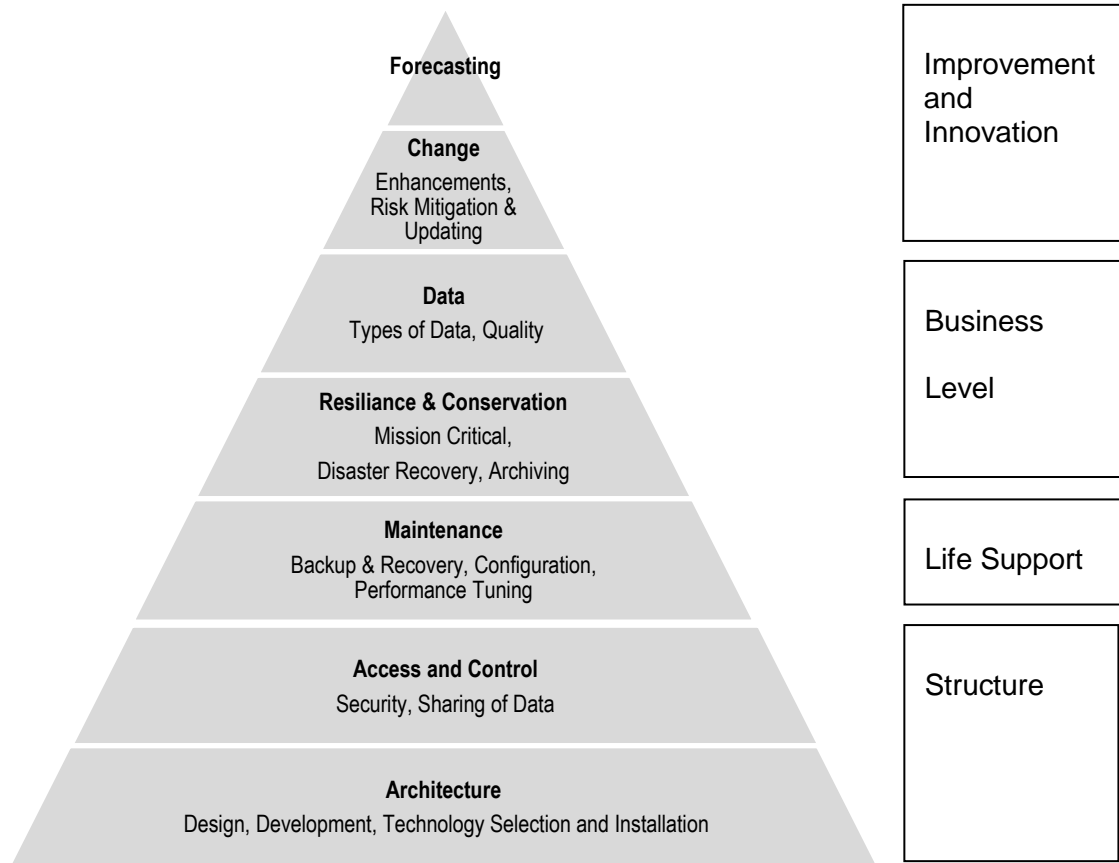
This initial system of interest placed many components from the rich picture, (Figure 1.1), outside the database system boundary. These components from a priori knowledge were thought to affect the database system but were not considered a part of it.

### 1.3.1 The Technical DBMS

The database system itself is constructed of many components. The many facets of this layered technical system contribute to the database system as a whole. The database is continually evolving and adapting to the demands of the users, organizations and environment but it is required to be available to share the content.

The anatomy of the DBMS technical structure is shown in Figure 1.3 – a suggested development of the information systems diagram in the Introduction to Information Systems by J O'Brien (1998, p.14). The model shown in Figure 1.3 was used to

help understand the parts of the technical DBMS and to illustrate how each part of the hierarchy builds on the others to form the holistic set of requirements.



**Figure 1.3 The technical DBMS based on J O'Brien (1998, p.14)**

This diagram suggests a hierarchy within the DBMS. Best practices and procedures exist in each of the layers and are often affected by other layers, often involving different teams of people.

The structure pillar consists of the architectural (design, development, technology selection, installation) and access and control layers (security, sharing of data, privacy, encryption). Database infrastructure architecture, design and development are the foundation stones of any database system. Database architecture consists of the database software and that of the user defined application databases.

Security of data is a primary concern of the database administrator. The protection of the asset is critical for organizations and users.

*“As organizations increase their adoption of database systems as the key data management technology for day-to-day operations and decision making, the security of data managed by these systems becomes crucial. Damage and misuse of data affect not only a single user or application, but may have disastrous consequences on the entire organization.” Bertino et al (2005, p.2).*

The life support pillar for the database is provided by the DBA who maintains the database server ecosystem. This layer includes the maintenance that is undertaken on the system, such as backup, recovery, configuration and performance tuning. Database maintenance is required to keep the database operating. Maintenance tasks have evolved around knowing what type of failures can occur.

The business level pillar includes a resilience and conservation layer (mission-critical 24/7 databases for high availability, disaster recovery, archiving of data) and a data layer (types of data, quality, transformation, governance). The demand on data availability and resilience is a critical factor when providing agile data anytime. The need to reduce complexity of applications to provide mission-critical secure systems which are scalable, auditable and recoverable is of key importance. Data is defined as

*“Data are raw facts or observations, typically about physical phenomena or business transactions [...] Thus data are usually subjected to value-added process [...] (1) its form is aggregated, manipulated, and organized; (2) its content is analyzed and evaluated; and (3) it is placed in a proper context for a human user.” (O’Brien 1998, p.23)*

The last two hierarchical layers are incorporated in the improvement and innovation pillar: change (enhancement, risk mitigation, updating, and documentation) and forecasting (trend and pattern analysis, capacity management, reporting visualisation, cloud).

An emerging feature of the database is that the state continually changes as changes in the real world are reflected within it. The changes are permanent or temporary and the database is continually evolving and growing (Mullins 2012, p.243).

Forecasting database systems' capacity, monitoring and performance to show trends, and reporting these trends, ensures systems are proactively maintained. Capacity management of a database system should be planned ahead, to identify the needs of the database to deal with the current workload on the systems and to continue to keep dealing with the workload:

*“This is a unique opportunity for a fundamental ‘reformation’ of the notion of data management, not as a single system but as a set of services that can be embedded, as needed, in many computing contexts.” Agrawal et al (2009, p.61)*

Best Practices are a part of the management of database systems. Best practices and procedures could be intended as a means to maintain consistent quality, but they might not actually achieve this, and are often defined by the software vendors, in whitepapers, on community experts' field notes, by organizations' documented procedures or by current usage.

### 1.3.2 People System

The database administrators and managers are all affected by the company culture. The perspective of a database administrator/manager was taken in deciding the operating model, described in the next section. The end users and vendors as depicted in Figure 1.2, are placed outside the initial boundary because they are not involved with the day-to-day task management. The implications of choosing this as the system could be that the initial assumptions are misleading, and the investigation might miss the collection of vital information that could elucidate the management of database systems. In addition the process of taking this boundary choice could result in an incomplete picture of the complexity. If it is found that users, and/or vendors have a role in the management of database systems, the research boundary may need to be changed. Also the researcher's experience as a database administrator may have influenced the situation and set misleading starting assumptions. The rich picture in Figure 1.1 raises various questions.

Cultural issues identified for investigation alongside the technical components are:

- The code of practice and governance consisting of rules, policies, processes and methods of handling data, by the actors involved
- The company perception of how database administration should be carried out rather than simply to rely on the DBA's knowledge of certain requirements
- The organizational management structure which dictates who controls the DBAs
- The customers' requirements which the management service should provide
- The financial controls over how much the company is prepared to pay for the software, how much training they are prepared to pay for and whether there is any budget for progression to new versions of software and hardware

- How proactive are businesses, to anticipate future events and plan for them, rather than being reactive and only dealing with events and problems when they occur
- Do people from within the database community liaise with each other to form opinions, treatment for issues and even potentially bring about changes to products? What are the attitudes and beliefs of DBAs?
- The influence of vendors, suppliers and competitors can bring about change and influence the future of tools and services
- The impact of interconnected teams such as Change Managers, Developers and Service Managers on the usability of the database system
- Marketplace changes from a local model to global model looking to incorporate cloud computing but also the change in usage patterns with availability being 24/7 for around the world access
- Environmental considerations, required when looking at datacentre power consumption
- The business has to consider changes to the business model and the risk involved with that. The government, legal and political issues also need to be considered

### **1.3.3 Operating Model**

Management of database systems begins by consideration of the required outputs and how these can be achieved. An operating model would need to be decided which would set up suitable frameworks and practices and procedures to be followed.

The operating cycle is influenced by the people throughout: staff skills, the level of communication between the teams and the culture of each team. Organisation

culture is affected by a variety of influences, such as sets of values and current and past events. The performance of the people is vitally important throughout for successful operations.

The DBMS does not automatically correct itself and people involved in operations need to inform managers of issues with the current operating system if improvements are to be made. In this way best practices evolve for a database system but these can be subject to change over time as new developments take place. For this reason the experiences of people working in the database field need to be sought.

## 1.4 Research Questions

The purpose of the research was to investigate the management of database systems to avoid the failures experienced in the past and to be effective, efficient and perform well. Before improvement can take place it is necessary to understand whether best practices and procedures are used and to further understand what complex interactions exist in the system. Best practices are discussed in the paragraph preceding Figure 1.1.1 as a recommended practice for carrying out actions for desirable outcomes rather than always being the best way of doing something.

The vignette given in this chapter is one of many similar scenarios which affect database management. The proposed research questions address the issues surrounding this complex system. The questions were both procedural and content related.

The research questions are significant and important to help understand the perspectives of the actors and stakeholders, the ecosystems and the complexity



involved. Answering these questions could lead to insights into the system not studied before as a whole and highlight how to achieve better performance during a lifecycle of operation. The ultimate aim is to help improve the quality of database administration management methods. Databases are key to everyday life and successful management of these is of critical importance.

### Question 1

*To what extent are best practices and procedures utilised by the database community?*

This question elucidated the current practices related to:

- the design of the system
- communication between all involved
- overall control of the system set up
- management of the DBMS
- the types of data
- the security of the information
- the accuracy of the data
- DBMS end to end lifecycle
- the continual use of database practices and procedures
- the combination of database components used and the relationships between the parts of the database system
- what is missing from the management toolset

Question 2

*What are the complex interactions that are an integral part of the management of database systems?*

This question examined the number and type of interactions now essential in the operation of databases today. The question focused on software, hardware, management and interactions. 'The complex interactions' are the interconnections between the attitudes, skills, knowledge and perceptions of management, customers, vendors, other staff, suppliers, database community, IT progression, competitors, other IT fields, developers, change analysts, service managers, social attitudes, beliefs and world markets. Then there are other factors that interact such as cost, market place changes from local to global, from private data to shared data to big data volume and increases in all types of data.

Question 3

*Is the adoption of best practices and procedures affected by the complex interactions that are an integral part of the management of database systems?*

The third question examined the relationship between best practices and the complex interactions. Complex interactions may affect best practices used within management of database systems and the connections may result in unpredictable outcomes.

The management of the database system has many interconnected parts. Holistic methodologies are practices and procedures which could be used in the management of the whole database system including taking into account the

interactions between the many facets. This question revealed further effects of the complexity of database systems.

#### Question 4

*How can a better understanding of the complex interactions contribute to improvement and innovation?*

The fourth research question took a forward looking approach gained from the increased understanding of the complex interactions. It aimed to suggest approaches for improvement and innovation within the management of database systems. By delving into the systemic problems surrounding existing systems suggestions for improvement were indicated.

## 1.5 Research Approach

The research aimed to establish the current position of best practice usage and to understand why best practices were either used or not used. In addition the complex interactions were examined to identify the actual interactions with an aim to improve the management of database systems. To address this need the research used a mixed methods approach (Creswell & Plano Clark 2011b). This uses a combination of quantitative and qualitative data collection methods. The research approach chosen could result in producing a more complete picture from the complementary methods. The specific research design used was sequential explanatory design. That design starts with quantitative data collection, followed by qualitative data collection and analysis. The quantitative stage was undertaken through a survey and the qualitative stage utilised focus groups.

The data analysis of the qualitative stage firstly used thematic analysis which then went through a transitional stage to the final synthesis stage utilising systems thinking. The use of systems thinking in the process and analysis of the research disclosed emergent properties of the whole system.

## 1.6 Significance of the Study

The management of database systems can often have problems. Changing technical scope, the growth in database communities and the requirements of database management highlight the lack of available management methods that are specific to database systems. The management methods used do not fully consider the effect of interaction between the components (Figure 1.2), technical factors and database culture (Figure 1.1). This research conveys an academic approach to a very large part of information technology that appears to have received little study. It offers insight into systems currently practiced in the IT world in respect of databases: thus many new and current database managers could better understand what is involved and be guided by the information revealed. The ontologies are the types and interrelationships that exist within the database system.

The state of a database system changes continuously, often with unpredictable and uncertain outcomes. The understanding of new technology and how the emergence of environmental factors has affected provision is extremely important. System control depends on a complete knowledge of the benefits and limitations of the technology and the complex interactions within the system. The quality and reliability of the work carried out is critical to good performance.

This phenomenon has not been widely researched to date, but a deeper understanding of it would be both academically innovative and beneficial to all organisations that have to manage database systems.

## 1.7 Overview of Thesis Chapters

Chapter 2 *Literature Review of Database Management in Practice* examines the existing body of knowledge in key areas that relate to the management of database systems. The problem domain crosses many fields and as such the literature is an amalgam of research areas that affect the study. It crosses database theory and practice, organizational theory, management theory, complexity theory and system theory. This body of knowledge provides a review of the diverse fields that are brought together in the research: organizational management, technical fields, DBMS knowledge, best practice methods, well known frameworks such as ITIL, TOGAF and Agile and information systems. This chapter also shows how the diverse fields underpin this research.

Chapter 3 *Research Design* provides a summary of the methodology used in this research. To gain a better understanding of the current situation of database practices through the lifecycle, statistical trends from industry practice were insufficient. Insight into the social problems encountered provided a way to add stories to the trends. For this reason mixed methods research provided the structure to this investigation. The mixed method research approach, presented by Creswell (Creswell 2009) and Creswell & Plano Clark (2011b), combines both quantitative and qualitative methods to investigate a single problem. The chapter is divided into three areas: defining the mixed methods approach of sequential explanatory design, then followed by sections on quantitative and qualitative methods for analysis. The concluding part of the method used systems thinking for synthesis.

Chapter 4 *Quantitative Survey Findings on the Utilization of Best Practices* provides a summary of the results and findings from the large scale quantitative research survey. This included the most relevant results from the survey findings. The second part of the chapter provides a section which connects the quantitative and qualitative phases. The section explains the interesting points that led to the qualitative questions, used later for focus groups. The survey results were mainly related to the first research question.

Chapter 5 *Qualitative Findings, from Analysis to Synthesis* provides the outcomes of the analysis based on examples of the qualitative data from the focus group research, traversing through the method. The research questions addressed in this chapter are the second and third questions. The analysis traverses through 3 stages, the first coding cycle, the transitional process and concluding with the synthesis systems thinking stage. The first coding cycle includes the thematic analysis method. The transitional process uses various tools to change the focus of the analysis from the in-depth view to a holistic view. The final stage is the synthesis, using systems thinking. System thinking changes how the components are viewed and helps to gain a holistic understanding of the data.

Chapter 6 *Discussion of Database Management and the Complexity of Delivering a Best Practice Solution* provides a discussion of the research, drawing together the qualitative and quantitative analysis. The chapter provides a synthesis that captures both trends and details of the complex situations. The discussion illustrates the complexities involved in the management of database systems through the use of textual quotes and diagrams. The bridge between the quantitative and qualitative data analyses to gain deeper insight is discussed. It is concluded that the use of best practices and procedures are not always successful in fulfilling all requirements of database management quickly, reliably and with ease. A new blueprint for

database management is presented: the CODEX (Control Of Data EXpediently).

The CODEX is an agile innovative way to look at managing database systems which has collected together components under the headings of: Control; Control of Operations; Data; EXpediency and X (unpredictable events). The components are all interconnected together through the system, and need to be considered carefully together. The CODEX is about creating continuously changing best practice.

Chapter 7 *Research Conclusions* summarises the outcomes of the research and provides interpretations and recommendations. It discusses: what has been learnt from the methods; the CODEX; and potential further research. The potential next stage of the research would be to look at machine learning algorithms or graph theory for use with the CODEX to help improvement in the management of database systems.

## Chapter 2: Literature Review of Database Management in Practice

### 2.1 Introduction

The goal of this chapter is to explore the existing body of knowledge relating to the management of database systems. The diverse fields which come together in database systems need to be considered as part of this research. This chapter introduces best practice, which is continually referred to by vendors and organizations as a way to manage database systems. It then moves on to discuss system thinking as a holistic method and way of examining database management. Organizational management and information systems set the scene for database management within the organization. The discussion then progresses to the database system followed by the technical system and the frameworks that can be used to help manage the current system. The chapter concludes with a discussion on improvement and innovation and how this can be achieved. This interdisciplinary field has many lessons that can be drawn together to improve database management.

### 2.2 Best Practice

Best Practice is a pervasive term that means different things to different people. The working definition in this thesis for best practice is defined in section 1.1, as a recommended practice for carrying out actions for desirable outcomes. Best Practice has been defined in various ways (Dembowski 2013; Wellstein & Kieser 2011; Sanwal 2008). Dani *et al.* (2006) stipulate



*“A best practice is simply a process or a methodology that represents the most effective way of achieving a specific objective”.*

Jarrar & Zairi (2000) state that the term Best Practice is often used within organizations to depict leadership and is recognised as the best way to achieve superior results. In the glossary of benchmarking terms (American Productivity and Quality Centre 1999) cited in (Jarrar & Zairi 2000, p.S734) best practices were defined

*“Those practices that have been shown to produce superior results; selected by a systematic process; and judged as exemplary, good, or successfully demonstrated. Best practices are then adapted to a particular organisation”*

Many different situations require different best practices and with new technology evolving ‘best’ is a moving target (Jarrar & Zairi 2000).

Markus (2011, p.4) argued that the cultures and practices that develop over time in organizations have changed to become “off-the-shelf” services labelled best practice standards, which organizations needed to adopt and understand. Markus argued the change from unique coded management ideas for handling packages to standard software with relentless upgrades requires knowledge development and standard practices.

Sanwell (2008) stated that the use of best practices are affected by certain beliefs:

- Best practices help make decisions quickly in a complex uncertain world
- Best practices are easier because they have been proven by other organizations who also operate with complex and uncertain elements.

- Management understanding of other organizations in the field are organizational specific. Best practices are often developed later and often already behind leading organizations.
- Value must be gained from best practices as other experts, consultants and vendors share them for current trends
- Best practices can improve performance

Falconer (2010) argued to the contrary that best practice exacerbates failure:

*“Best practice is flawed because it acts as a placeholder for proper management practice, displacing accountability for effectiveness and fit. Best practice is flawed, further, because it supplants strategy, adopting solutions out of convenience or copying them reactively, and supplants innovation, allowing “the best we know about”, “the best we’ve come across”, or even “the best we’ve done before” to be adequate. Best practice considers the world predictable, and discounts the emergence of better, novel ideas” (Falconer 2010, p.754)*

Falconer thought that problem situations are being incorrectly handled due to best practices replacing analysis.

Sanwell (2008) pointed out that changing these best practices in the multidimensional world requires consideration of organizational culture and behaviour, organization processes and organizational systems.

As Gonnering stated,

*“Best Practices” can serve as a beginning but adaptation will most likely be necessary. Outcome is an emergent property, and the organization that has taken the time to learn the methodology of improvement will*

*reap the benefits. The “continuous” in “continuous quality improvement” depends upon rapid-cycle, small-scale serial innovation and not a static and dogmatic adherence to past processes.” (2011, p.100)*

Gonnering argued that complex problems using best practices failed to have positive outcomes and forced the complex systems to become chaotic.

Bretschneider *et al.* (2004) highlighted three important characteristics of best practice: a comparative process, with action, and linked to an outcome or goal.

Nattermann (2000) suggested best practice might be the most widely used management tool in business and important for improving operational efficiency, but for strategic decision making, best practices might not be the best way forward to increase profit margins. Best practices management could be used to benchmark performance, with certain benchmarks being required to demonstrate best practices.

The core or classic best practices utilised within the database community have been developed through the sharing of knowledge, experience and actual outcomes across the sector. The improvement of these best practices were raised by Gratton & Ghoshal (2005) with the term “a signature process”, a process that envelops the company’s character and idiosyncratic nature. This signature process could advance the company although it required careful adaptation and alignment to business goals to succeed. However the allure of classic best practices that were clear, logical and easy to understand were the ones shared within the database community, the body of knowledge often yielding optimal results (Tucker *et al.* 2007). Some best practices were tightly coupled with their organizations and inseparable from the context (Becker 2004).

Jarrar & Zairi (2000) identified three types of best practice: proven best practice across organizations, good practice techniques for an organization, and unproven good ideas based on intuition. There were drawbacks with unproven ideas that could be a matter of luck and the lack of information to reduce the risk, lack of situational context, application criteria or success measure (Falconer 2011). This serendipitous discovery could lead to ease of deployment and innovation.

The Cynefin framework (Snowden & Boone 2007) classified and ordered simple systems in the domain of best practices. In an earlier paper in the chaos domain Kurtz & Snowden (2003) argued that applying best practices probably caused the chaos in the first place. They argued that different contexts use different management responses and that there are different tools for the management of complex contexts. The best practices domain is based on cause and effect relationships that have simple contexts, often within areas that do not change frequently.

Wagner and Newell (2011, p.400) stated that

*“The best way of operationalizing a process in one context and at one point in time may be different in another context and time”*

They contended that there is no such thing as best practice, as knowledge is created by engagement in a practice. Practice is always changing and emergent with inconsistencies in the same practice, with best practice being defined locally. Wagner and Newell (2011, p.401) suggested a move to negotiated practice with a cooperative approach to best practice adoption. Their aim was to smooth out complex implementation through compromise. They concluded that highlighting problems with identifying best practice (due to it being an interactive process based on learning through implementation with information systems) sometimes required

customisation to work well. This approach was also adopted by Avgerou & Land (1992) with their notion of 'appropriate' context specific practice, where information systems innovation looked for

*"best practice, or suitable new organizational form for the information age"* (Avgerou 2011, p.650).

Avgerou drew together organizational and information systems to develop a framework which had one key tenet of a knowledge management system or a best practice solution to help address static and commoditized technology.

Best practices and procedures were continually developed by database software providers (e.g. Microsoft, Oracle and MongoDB) to enable the management of database systems to be carried out to the highest standards. The procedures were based on formal rules the business world defined which were sometimes called standard operating procedures (Becker 2004). Best practices were defined by the software providers as exemplary tested designs for certain configurations or ways of doing things. They were multi-faceted and resided in varying layers from architectural design, through development, to operational management.

The management of database systems utilizes best practices and procedures provided by software providers and often industry best practices shared by the community. McGregor (2007) argued that this rarely leads to great customer service. McGregor's (2007) idea that "Next Practice" was the future of continually analysing and looking for positive quality products and service in other organizations, would bring ideas and innovation to improve the business. There was an aspiration to improve database management and improve business processes to provide good quality service when managing IT projects and database systems. Best practices might not however be the best solution. Sanwell (2008) raised some

key issues with using processes and strategies created by other organizations, and did not believe that following these would create a better organization or bring about improvement.

Within database systems there are various types of practices and procedures that need to be incorporated within change processes. Savage (2014, p.17) stated Stonebraker thought “in memory” database engines will take over online transactional processing systems (OLTP). Savage (2014, p.16) shared Stonebraker’s views on the database world, that it could be divided into three types: OLTP, data warehouses and everything else (Hadoop, graph databases). This was likely to mean three or more database management and best practices models were required.

Best practices operate at different levels within the sphere of database management. There are technology best practices which deal with specific tasks for deployment of databases onto servers or into the cloud; and management best practices which relate to higher level functions and overall processes. In addition there are best practices which are defined by software vendors for their own products. As technology and management change, in the world market, and more is understood about certain areas, best practices change. Thus best practices are replaced with new best practices. The large collection of best practices created are likely to be defined and owned by a multitude of people. This can cause problems with conflicting best practices. Sometimes there is a mismatch between best practices and a compromise needs to be found where possible.

Best practices are intended to be useful for technical solutions to help people provide the required results. They aim to provide a useful guide on what management need to do to perform certain tasks. Best practices are sometimes

adapted from vendor or industry defined best practices for nonstandard configurations or different business scenarios. However, sometimes communication is lacking between the management requirements, the vendors' practices and the technology tasks. Different teams may each create best practice, in places where the technology overlaps, which are not shared. There are therefore limitations to the usage of best practices. The best practices presented are significantly different for ILTM, CMM and ILTIL. There are many different types of tasks from in depth technical ones to higher level models that combined can produce a well-managed database system. Each task, model or part of the database system will have its own best practice, which aims to achieve those reliable results. These best practices at different levels may, in practice, sometimes be in conflict. This discussion on best practice has shown there are many diverse views on the usability and definition of best practice. The working definition in this thesis for best practice is: a recommended practice for carrying out actions for desirable outcomes, rather than always being the best way of doing something.

## 2.3 Systems Thinking

Systems thinking is used within this research to advance understanding of the operation of databases. Navigating through the ubiquitous database system incorporating management best practices, new technology and application information systems and the database ecosystem led to the adoption of a systems thinking approach. Systems thinking has only infrequently been applied to the management of database systems. A system has been defined by a number of people and the definitions by Ackoff, Checkland and Senge set the scene for this research. Ackoff defined it as

*“A system is a set of two or more elements that satisfies the following three conditions.*

*(1) The behavior of each element has an effect on the behavior of the whole.*

*(2) The behavior of the elements and their effects on the whole are interdependent. This condition implies that the way each element behaves and the way it affects the whole depends on how at least one other element behaves.*

*(3) However subgroups of the elements are formed, each has an effect on the behavior of the whole and none has an independent effect on it.”*

*(Ackoff 1981a, p.15)*

and Checkland defined a system as:

*“A model of a whole entity; when applied to human activity, the model is characterised fundamentally in terms of hierarchical structure, emergent properties, communication, and control. An observer may choose to relate this model to real-world activity. When application to natural or man-made entities, the crucial characteristic is the emergent properties of the whole.” (Checkland 1999, pp.317–318)*

Using systems thinking, it is possible to examine the interconnected components of information systems. Systems thinking is:

*“The discipline for seeing wholes. It is a framework for seeing interrelationships rather than things, for seeing patterns of change rather than static “snapshots” [...] Today we need systems thinking more than ever because we are becoming overwhelmed by complexity. Perhaps*



*for the first time in history, human kind has the capacity to create far more information than anyone can absorb, to foster far greater interdependency than anyone can manage, and to accelerate change far faster than anyone's ability to keep pace [...] Systems thinking is a discipline for seeing the "structures" that underlie complex situations"*  
(Senge 1990, pp.68–9)

Senge (1990, pp.57–67) had a different perspective on organisations. He argued the problems we see today have come from our past solutions - compensating feedback could result in more energy being used to improve the situation. Cause and effect if understood correctly could bring improvement.

Systems Thinking is the change to synthesis (putting things together as wholes) from "Machine Age" analytic thinking (taking things apart to reduce focus) (Ackoff 1981a, pp.16–17). This new approach to organizational management was started by Bertalanffy (1969) who argued for a General Systems Theory (GST) across all systems. Capra and Luisi (2014, p.80) stated that "*systems thinking is inherently multidisciplinary*" and that qualities and patterns in non linear dynamics were a characteristic of system thinking (Capra & Luisi 2014, p.114). Rousseau and Wilby (2014) argued that the complex challenges facing design and management could only be overcome using a systemic transdisciplinarity approach where it is possible to :

*"see across the boundaries between the disciplines and therefore reveal the impact of local interventions on the neighbouring and global systems. "* (Rousseau & Wilby 2014, p.674)

Emergence has led to further database development where large volumes of data of various types are involved. This is called big data and as yet has not been precisely

defined although Hashem *et al.* (2015, p.100) proposed an enhanced definition of big data based on observation and analysis; examples of big data were social media sites, banking, scientific data from CERN and NASA. Clearly the rigid rules of some database structures were not appropriate and a new set of tools and management techniques were required. Gray cited in (Hey *et al.* 2009) discussed a fourth paradigm for data exploration following empirical, theoretical and computational paradigms. This was raised due to the huge increase in volumes of data. Bell *et al.* (2009, p.1298) raised issues that the database community's speed of advance was due to the database skill set, workflow management, visualization, and cloud computing technologies. Management of the abundant complex data has led to the emergence of the data intensive paradigm, discussed in 'Science 2020' (Franklin *et al.* 2005) and later expanded upon by Buchan *et al.*, cited in (Hey *et al.* 2009, pp.91–97). Data was now both structured and unstructured and the technology of home and the workplace was merging.

There might be an impact on database administration methods where no feedback took place from all the ecosystems involved. Reflection in action could help improve the situation but it has limits. It was further reflection on past action that allowed change and improvement. Capra and Luisi (2014, pp.362–3) maintained that there were ecological patterns and processes that were fundamental in systemic understanding that when ignored can cause problems in a globally interconnected world. They also postulated that large social institutions were subscribing to an out of date world view:

*“These systemic problems, they require systems solutions; and since only the viable solutions are those that are ecologically sustainable, they must incorporate the basic principles of ecology, or principles of sustainability.”* (Capra & Luisi 2014, p.363)

Thus systems thinking should look holistically at the elements incorporating hierarchical structure, emergent properties, communication and control. Morgan (1986, p.47) states that living systems should not isolate themselves from the diversity in the environment. Requisite variety, originally proposed by Ashby (1956), states that the internal control mechanism must be as diverse as the system's environment, to ensure the complexity and distinct nature of the environment is not lost and to prevent atrophy from occurring. To enable systems to be successful in dealing with changes of the environment, an appropriate level of variety must be incorporated into internal controls. In database systems, if a customer wants a change that affects the database, all the connected teams such as service desk, database administrators, change teams, need to be informed and current practices followed.

A model which could be undertaken in the service sector to enhance organizational resilience is the Vanguard model (O'Donovan 2014; Jaaron & Backhouse 2014) originally proposed by Seddon (Seddon 2003; Seddon 2008). This model was based on systems thinking and could improve efficiency and effectiveness through the organizational structure and employees' commitment. The Vanguard method could help organizations change from a command and control system to a systems approach.

In summary systems thinking takes into account characteristic structures, emergent properties, communication and control which help with understanding complexity and patterns of change. It is a transdisciplinary approach that incorporates ecosystems with human perspectives, culture, business and technology, and can help to improve efficiency and effectiveness, which are of key importance in organizational contexts. Thus systems thinking is an extremely useful tool for advancing understanding of the operation of database systems, leading to future

improvements. The next section examines areas of organizational management connected with managing databases.

## 2.4 Organizational Management

The operation of organizations has been studied over many years and the theories are still very relevant to with the operation of databases.

Within the classical school of management Taylor's Theory of Scientific Management discussed specialists and division of work. Frederick W. Taylor (1947), cited in Pugh (1990, p.179) was the father of Scientific Management.

Taylor is relevant to the management of database systems as tasks require in depth technical knowledge and as such the work is divided between the relevant technical people. Taylorism is also found in particular work contexts where the operation of machine-like precision is required (Bell & Martin 2012, p.107). An example is the management of technical support staff who answer database support calls and are required to resolve these database incidents and service requests in the quickest timeframes. Bell and Martin (2012, p.111) argue that Taylor's scientific methods determine the component order in tasks and balance between workers. The teams working on database management support have tasks split between workers, sometimes based on the difficulty of the tasks.

Taylorism has been widely criticised, as discussed by Morgan (1986, p.35) who states that this mechanistic model has both strengths and limitations. When straight forward tasks and exactly the same product output is required and accuracy is important, the mechanistic approach can work well. Morgan (1986) argues that the centralization of design and development of products and services, together with controlled decentralized implementation, has worked to great effect with the

Taylorist approach. Database systems use in part pre-defined standards, documented performance targets and run books that set out precise tasks that must be carried out in order to achieve the desired state.

Morgan (1986) discusses the limitations of Taylorism: difficulties in adapting to changing circumstances quickly; and employees may not be given opportunities to innovate. Morgan (1986) argues that mechanistic approaches do not mobilize human capacities or allow the strengths and potential of people to be built upon in organizations. Garud and Kotha (1994, p.671) argued that Taylor was appropriate for the mass production era but that flexibility is required for rapid change and the increasing variety of products, which are critical to an organization's survival.

Greenwood (1981, p.225) suggested that many organizations which have service related database teams use Peter Drucker's Management by Objectives philosophy (Drucker 2007, pp.84–94) allowing superior and subordinate managers to reach common goals. This allowed technical specialists to bring their existing and new ideas of improvement and innovation to the table when reflecting on any emerging facts identified as a result of performance evaluation. Technical specialists were adept at both the technical level and following best practice.

Organizations can vary significantly in size from a few people to large multinationals. The activities undertaken by these organizations are diverse and global. Most organizations rely on databases and require them to be managed to ensure the data contained within them is accurate, available, secure and recoverable. Some organisations might not fully appreciate the significance of this data to the organisation and Peters cited Mosley *et al.* thus:

*“Organizations that do not understand the overwhelming importance of managing data and information as tangible assets in the new economy will not survive.” (2009, p.1)*

Data is used for a multitude of purposes for critical applications. An ever increasing number of tools are available to manipulate data. Key changes include the volume of data, mapping data lineage (its origins and migration) and Internet of Things (IoT). IoT was coined in 1999 by Kevin Ashton to denote a network of physical objects that communicate and interact with the environment using technology (Gartner Inc. 2013). This increases the complexity for managing database systems. There are many challenges for database management discussed by Cooper & James (2009). Aiken (2016) discussed transformation of the organization to a data driven culture which requires technology, people and process.

Databases for many organisations are critical to everyday business activity whether they are used for internal application management (e.g. finance and HR) or external facing applications. Stein *et al.* (2015) argued that organizational goals are achieved through the effective use of IT.

Organizational management of database systems might be located in a part of the organizational structure considered best for producing successful outcomes, based on the team interactions and communications necessary. This is underpinned by organization theory, which examines the core concepts of strategy, goals, technology, social dynamics, culture, change, learning, decision making, politics; and environmental factors which can lead to complexity, change and uncertainty.

It is not possible to fully understand database systems management without understanding the core concepts of organizations. The organization's strategy and decision making could be influential factors affecting the management of database

systems. The culture of the organization could affect database management through the strategy and decision making.

Communication between all the departments, stakeholders and team members is required when managing data and databases. Communicativeness, Ackoff (1999, p.128) stated, is an “essential property of good management”. Organizational management of database systems can often suffer due to different management viewpoints. Sometimes conflict could occur between any of these parties which, without good communication between them, might affect overall management of database systems. A final key organizational component is learning. Technology is advancing and changing all the time, requiring organizations and people to continually learn new concepts and methods.

The following sections discuss: decision making, a key element relating to how management is carried out and whether best practices are used; culture and conflict, which are prevalent within the parts of organizations relating to database management; and learning, an important factor given continual technological change.

### **2.4.1 Decision Making**

Decision making is interwoven into organizational management. Decision making is based on judgment (Drucker 2007, pp.183–197) and there are four types: generic; answered through rules; identifying the boundary conditions that need to be met through clear specification; turning the decisions into action and ensuring the effectiveness through feedback. There have been a number of studies into organization decision making theory. A theory of the effects of advanced information technologies on organizational design, intelligence, and decision making (Huber 1990), showing the need for critical empirical investigation, focused on:

*“technology-prompted changes in organizational design that affect the quality and timeliness of intelligence and decision making, as contrasted with those that affect the production of goods and services”*

Pugh *et al.* discussed March’s organizational learning:

*“An organization is a collection of choices looking for problems, issues and feelings looking for decision situations in which they might be aired, solutions looking for issues to which they might be the answers, and decision-makers looking for work.”* (1989, p.117)

Child (1983) discussed control as an important factor relating to decision making and at what point decision making should be delegated. He also looked at practices and procedures debating how formalised these should be and how much supervision there should be to ensure that a predictable consistent behaviour was preserved.

*“there is somewhat more appreciation that effective control requires a positive commitment from employees if instructions are to be followed and accurate feedback secured on results”* (Child 1983, p.134)

Drummond (2014) concluded that complex projects rarely go smoothly. She raised the need to educate managers to prevent inconsistency without which there may be ruinous consequences for conflicted managers unwilling to admit that projects have failed, who reinvest in tools too late. Drummond (2002) argued that ‘in organisations, assumptions have a habit of hardening into fact’ (2002, p.237) and that therefore it is that the “knowns” that are the most dangerous; based on an analysis of the Barings clearing bank collapse, she argued that, although decisions



with uncertainty imply risks, significant risks can also arise in situations where people - wrongly - feel certain.

## 2.4.2 Culture and Conflict

Database culture is closely coupled with organizational culture. 'Database culture' is the attitudes and norms of the people working on databases in that particular organisation. The established culture can affect database quality and availability. An example of practices and habits that are characteristic of a particular database culture could be that some changes may be carried out under the radar to avoid going through all the extra management controls and approvals. Another example of database culture could be where database changes to fix a problem are made in a reactive manner rather having a proactive change policy.

Organizational cultures are distinctive and can be varied. Handy shared the view that organizations were affected by both past and present events, the people, their values, beliefs and type of work (1985, p.185). Organizations were also affected by the environment such as "*economic, legal, social and ethical factors*" (Pettinger 2000, p.21).

Katz and Kahn's archetype model (1978, cited in Pettinger 2000, p.22) linked the organizational, technological and other holistic factors together. The global organization displays the complexity and interconnectedness of the database system with its environment. In this research, the DBMS was included in the technology layer whereas the holistic organizational culture rests at the centre.

Schein (2010) described culture as an abstract concept, which had been used in many contexts. He argued for the use of deeper anthropological models. The observable events Schein discussed formalised skills that could be used to manage database systems.

The model of culture Schein refers to includes behavioural interaction, espoused values, philosophy, skills, climate and rules (Schein 2010, pp.14–15). The societies and customs were embedded into the people who manage the database systems.

Schein argued that

*“The culture of a group can now be defined as a pattern of shared basic assumptions learned by a group as it solved its problems of external adaptation and internal integration, which has worked well enough to be considered valid and, therefore, to be taught to new members as the correct way to perceive, think, and feel in relation to those problems.”*

(2010, p.18)

These group dynamics can be seen in action within database management teams. Schein (2010, p.24) developed three levels of culture matching the three levels of a database system. The first was the artifacts that could include architecture of physical environments, its technological language and non-technological language, technology and products. In the second level, espoused beliefs and values were obtained from shared knowledge, dealing with tasks, issues and problems and reflection. Thirdly, basic assumptions became reality when hypotheses were proven as working solutions.

Traditional database management did not use agile methods; however in development teams agile methods have become the prominent way of working. A new climate and culture combining these DevOps, development for operational tasks, evolved over the last few years to enhance technological management, speed and accuracy. The State of DevOps Report (IT Revolution 2015) measured cultures and organizational performance and used Westrum’s (2004) three cultures model: pathological (power orientated), bureaucratic (rule oriented) and generative

(performance orientated). The report described information flow as critical to the successful implementation of resilient systems at scale.

Conflict arose from different values, different viewpoints, issues and political behaviour. Hatch (1997) stated that cooperation was required in organizations as a classical management view. Conflict could disrupt this. Hatch (1997, p.301) defined organizational conflict as

*“An overt struggle between two or more groups in an organizations. It is usually centred on some state or condition that favors one social actor [...] also frequently explained in terms of interference [...] when the activities of one social actor are perceived as interfering with the outcome or efforts of other social actors.”*

Job satisfaction and stress is a factor of the data professional's role. A survey (Mckendrick 2014c) stated that there was too much firefighting and not enough time for innovations - managing increasing workloads and complexity was challenging.

Organizational hurdles for distributed DBMS are the same as those raised for the database systems, as stated by Gordon (1992, p.339). Further organizational issues include culture, structure and top management's attitudes to new technology.

### **2.4.3 Learning**

Organizations grow and learn continuously as they evolve (Senge 1990, p.14). How the organizations learn can differ between different organizations. Organizational Learning was an influencing factor in how database systems were managed. A learning cycle was identified by Kolb (1984, p.33) as an adaptive holistic process. Kolb created a model of four elements: concrete experience, observation and

reflection, the formation of abstract concepts and testing in new situations (Kolb *et al.* 1971, p.40).

Argyris and Schön's (1978, pp.8–29) investigations into organizational learning looked at the mismatch of expected outcome and the correction of errors through single loop and double loop learning, where organizational strategies, assumptions, frameworks, collaborative enquiry and (in the case of double loop learning) the restructuring of organizational norms, were also restructured. Both Kolb and Argyris's learning methods could help identify the development methods utilised (or not) for both existing and new database processes.

The development of databases to be self-reproducing and self-managing brought self-learning to the database platform. Handy (1989, p.46) produced a learning wheel which was continually applied to the progression of management of database systems. This in conjunction with Deming and Kolb's cycle demonstrate how learning took place within the database system. For best practices to be effective they need to be transferred between people. This could be done through the use of an internal knowledge base (Zairi & Whymark 2000) although Jarrar & Zairi (2000) stated the transfer of best practice learning from knowledge bases was a complex process.

The organizational management section discussed various approaches to organizations, each of which represent a facet of the management of database systems in practice. These include the mechanistic approach to management of Taylor, which despite the weaknesses discussed by Morgan (1986) remains central to the way call-centres are managed across a range of organizations; Drucker's (2007) management by objectives; Peters & Waterman's (1982) case for raising the importance of management of data for organizations; Ackoff's (1999) argument for

good communication; Drummond's (2002) types of decision making and Child (1983), who stated control was important. There is a new emerging culture of DevOps, where collaboration and communication provide an agile relationship between development and IT operations. Westrum's (2004) typology of organizational culture identified a way to shape the performance of an organization. The section concluded with a brief discussion on how organizational learning takes place. The research reported in this thesis was influenced by all these approaches, given the need to consider a breadth of socio-technical aspects in order to fully understand the complexity of the database systems field. The organizational management approaches discussed in the section all relate to different areas connected to the management of databases.

## 2.5 Information Systems as Adapted to Database Management

The section introduces information systems and shows how they are connected to database systems and their management. Database systems are a type of information system. Information Systems are integrated components (Figure 2.1) combining hardware, software, infrastructure, data resources, people, processes and organizations. Information Systems help organizations thrive, have efficient decision making and can help generate innovative ideas.

Information Systems began to emerge as a discipline in business schools in the first era of Hirschheim & Klein's (2011, p.20) model, c.1964-74. Since then technological development along with system development led to the creation of separate information system departments (1985-94) responsible for "*maintaining organisation wide data for future needs*" (Hirschheim & Klein 2011, p.36). The fourth era (1995 onwards) was for information systems technology and the business environment. There has also been a shift for organizations to "*provide better services to their*

customers" (Hirschheim & Klein 2011, p.42) in this global internet age of widely distributed technologies and data.

The Information Systems conceptual framework in Figure 2.1 shows the core components and activities which are part of information systems. Data is a core component within the system which connects hardware, software and people.

O'Brien (1998) described this model as the relationship between "*people, hardware, software, data and networking*". It is within this backdrop that database systems are situated.

Figure 2.1 has been removed from the electronic copy as this image is copyright protected by McGraw-Hill.

**Figure 2.1 Components of an information system  
(O'Brien 1998, p.20)**

Information Systems is a multi-disciplinary field which has been defined as:

*"A combination of two primary fields: computer science and management, with a host of supporting discipline e.g. psychology,*

*sociology, statistics, political science, economics, philosophy and mathematics. IS is concerned not only with the development of new information technologies but also with questions such as: how they can best be applied, how they should be managed, and what their wider implications are” Boland and Hirschheim (1985, vii) cited in (Checkland & Holwell 1998, p.10)*

It was this definition which drew this research to seek a methodology for dealing with information systems. Checkland & Holwell raised the point that there was a lack of any clear structure for information systems (Checkland & Holwell 1998, p.29). Information systems act as a bridge between the technical components, the data and the system. Systems thinking is a way of thinking about the information systems field (Checkland 2011, p.94).

## 2.6 Database Systems

A database system was defined in Chapter 1 as the larger holistic system including the DBMS. The boundary, as expressed in Chapter 1, was extended to take in data warehousing and big data. This system contained emergent properties which could be outcomes from a complex system. Complexity, which exists within the database system, has many different and connected parts and the factors of the complicated process or situation were not easy to analyse or understand (Anon 2014). This section draws together the technical DBMS, systems, ecosystems, emergence and complexity to drive improvement and innovation.

Haigh (2006) thought the DBMS to be the foundation of every modern business. The concept of management information system (MIS) was conceived in the 1960's, and was promoted by people such as computer experts, consultants,

manufacturers. Haigh (2006, p.33) termed them “Systems Men” (corporate staff specialists in administrative management). Conceptual development and technological innovation were running in parallel. The DBMS became the new term to integrate pools of data and all aspects of management which Haigh (2006) described as a “totally integrated management of information systems”. Deutsch (1970, pp.19–28) argues that there have been four data revolutions in terms of the type and quantity of data to be handled:

- One consisted of disjointed facts and figures, such as the book by Sir William Petty’s entitled *Political Arithmetick* (1683-1689).
- The second around 1840 was mainly historical data, types of societies, stages of society, locating theory and economic geography. The efforts of social scientists, such as Herbert Spencer in developing more detailed theories was a later step within this stage.
- The third began around 1935 with new collection methods of partial and sectoral data, using new quantitative methods for organising and interpreting the data. Data was becoming relational and ordered due to the double discovery of data collection methods and more rigorous techniques for putting existing data in relation to each other.
- The fourth data revolution about 1970 marked the increase of multiple methods and complex data bases.

As illustrated by the four data revolutions, data has always been an important part of everyday life and it penetrates all areas in society. The storage of facts and figures, societal, geographic and economic data, data collection and transformation, and rigorous validation techniques have led to an increasing number of tools and database engines. Each innovation expands data types, societal impact and widens the tool set for managing database systems. Hence, the complexity of management



of data has increased. The speed and type of data revolutions has been rapidly increasing since 2000. The revolutions that have been notable are: the Big Data revolution, the NoSQL data revolution, the Open data revolution, the analytical data revolution and 2017 the algorithmic data revolution with artificial intelligence. These technological revolutions impact all types of social areas from human development (DNA), scientific data and citizen data.

With the invention of the World Wide Web, by Tim Berners-Lee, the range of data, knowledge and information required by society had been magnified. For many organizations, as Woody (2002, p.4) stated about DBMS, access methods by software to the data were all that needed be considered, with the actual data storage, management and retrieval being abstracted. Codd (1971, p.377) stated that the creation of a DBMS required knowledge of: the data, the storage, the environment, governance and its intended usage. Codd also suggested that the internal representation of data in the database be extracted from the use of large data banks. This is contrasted with the additional priorities of security, cloud and big data (Mckendrick 2015) that data driven organizations are now facing.

Administration of data within the DBMS originally emerged informally and Lyon described the role of database administrator (DBA). The DBA must

*"at once be technically qualified, if not inventive [...] he must encourage the users to work with him willingly and yet he will be forced to rule against their pet projects; he must be all things to all people at all times"*

(Lyon 1971, p.12)

A disjointed database culture was described by Nolan (1973, p.80), with the utilization within organizations of fractured data repositories (different databases for each team or group).

Date (1981) defined DBMS as a computer record-keeping system to envelop organizational processes and deliverables which fed back into improvement and innovation. Stonebraker (2016) argued that customers depended on a DBMS to be bulletproof and commercialization added software challenges.

The culture of the DBMS has been defined from its earliest creation and evolved through commercialization. Bergin *et al.* discussed using “*data as a corporate resource*” (2009, p.26) to raise the image of computing within the managerial world. The idea of “data as a resource” recently became a reality with the sharing of corporate data. Current examples of this are Microsoft’s Windows Azure Marketplace and Ordnance Survey open data, which provides mapping datasets for Great Britain. Using these datasets, among others, customers can subscribe and use preconfigured datasets.

Bergin *et al.* (2009, p.33) highlighted the efficiencies of scale when shared databases are used. Bergin *et al.* (2009, p.33) also raised potential issues relating to the data owner, the storage format, data access and maintenance of the shared data.

Many DBMS are relational and Codd (1985) set out rules to determine if a software database management system was fully relational. Understanding and belief in what could be achieved through the use of relational databases was important. King’s (2015) survey findings found that traditional DBMS were still primarily used with no plans in the next few years to use Hadoop (a framework for processing large distributed data sets) or NoSQL (a non-relational database). Sivarajah *et al.* (2017) discussed the broad challenges of big data, as data challenges relating to the characteristics of the data, process challenges for the capture, integration and

transformation for analysis, and management challenges such as security, governance and ethics.

There are various types of new databases that have departed from the transactional model now classified as NoSQL. These databases require slightly different management techniques and are linked with the traditional databases. Data warehouses are managed by many of the same DBMS tools as transactional systems, albeit with some additional tools. Cloud managed database systems often have separated services for data warehousing due to scalability and differing performance requirements. Although there are different technical challenges for relational, data warehousing, NoSQL and big data platforms, they are all used for managing data. A further development, Stonebraker (2012) argued, are the New SQL systems which provide high scalability and preserve ACID (Atomicity, Consistency, Isolation, Durability) properties of database transactions. Hayashi (1992) classified the DBMS into five areas: organization and personal training, hardware and software, data management, user services and evaluation. The findings from the latest survey report on the state of DBaaS by IBM (2016) stated that database administration was a key challenge facing enterprises.

Diversity in tools, data, database administration and knowledge are some of the challenges facing the management of database systems.

*“Today’s data driven world involves a richer variety of data types, shapes, and sizes than traditional enterprise data, which is stored in a data warehouse optimized for analysis tasks. Today, data is often stored in different representations managed by different software systems with different application programming interfaces, query processors, and*

*analysis tools. It seems unlikely a single, one-size-fits-all, big data system will suffice for this degree of diversity.” (Abadi et al. 2016, p.96)*

In addition to technical and data components management through practices, procedures and governance are important. Juiz & Toomey (2015, p.60) argued that IT governance is sometimes confused with internal IT management. Governance is linked to the business and the newer model Juiz & Toomey described should also cover the broader external business agenda. Governance should no longer focus purely on technology although it affected all processes, accountability and best practices relating to the management of database systems.

The ubiquitous management of the database system required the understanding of many overlapping areas. This entanglement of database complexity and databases is irrevocably bound to organizations and people, and an ability to understand the linkage would enhance their management. This may lead to the improvement and innovation of database management. Determining exactly the areas for improvement and innovation could help clarify future possibilities for research, for which Artus (2008) highlighted the key skills in which the database researcher should be educated: constructing and managing databases, sociological, psychological, statistical and informatics.

The increasing number of components and development of database systems has been described in an entire suite of database management surveys. An initial workshop, with the Laguna Beach participants (Bernstein *et al.* 1989), was held on future directions in DBMS research, discussing very broad requirements of database systems. Silberschatz *et al.* (1991) discussed the characteristic requirements of database systems at the first National Science Foundation

workshop, these being efficiency, reliability, access control and persistence. This early paper shared a moral from Relational DBMS that:

*“when the relational data model was first proposed, it was regarded as an elegant theoretical construct but implementable only as a toy.”*

(Silberschatz *et al.* 1991, p.113)

A second National Science Foundation workshop (Silberschatz *et al.* 1995) raised the issue of the drastically changing landscape of database systems. The NSF workshop argued that the scope, magnitude and complexity of the database systems had expanded significantly. Trends mentioned affecting database research included technology and database architectural issues. The demands of the changing world had pushed database technology to the limits. Other components were quality of service, distribution of information, degree of autonomy, data integration, data warehouse, workload management and ease of use.

Another workshop (Silberschatz *et al.* 1996) delving into the technologies that were part of the complex system:

*“The autonomy of information sites makes it impossible for any centralized authority to mandate standardization.”* (Silberschatz *et al.*

1996, p.773)

Many areas were raised to be studied in the research area and concluded that research needed to be broadened and new problems might have completely unidentifiable solutions. The Asilomar report (Bernstein *et al.* 1998) set out research activities and set a 10 year goal:

*“The Information Utility: Make it easy for everyone to store, organize, access, and analyse the majority of human information online.”*

(Bernstein *et al.* 1998, p.80)

This highlighted the potential usage of database systems today. Databases are apparently a simple structure of tables and objects but can be seen to be complex, drawing on both epistemological and ontological perspectives. From an epistemological perspective, we might say that the system is complex through a study of the methods, scope and validity of the database system. Ontologically, we could consider how entities relate to each other and the nature of the database world. This approach queries what the databases consist of and what operations are contained within them, thus more than simple table structures. The progression of these workshops highlighted the increasing complexity and growth of database systems today. Database systems were continually pulled in multiple directions by the owner of the systems requirements, user requirements, the database architect, developer, administrator, the environment and best practice.

It is important to consider all parts of the ecosystem and perspectives. Database administration had evolved into its own field and could be considered application centric. A database application centric approach has its own perspective. Instead of viewing the application from the outside in, it focuses on inside out. The application centric model starts with the application and considers factors that will improve the whole DBMS rather than just one component within the DBMS. The shift to a database centred approach was first presented by Bachman (1973) where he compared the shift from computer centred to database centred viewpoint, to that of Copernicus when he argued that the earth revolved around the sun rather than the earth being at the centre of the universe. Bachman (1973) thought the first step was to learn the “rules of the road” to be able to navigate the database information

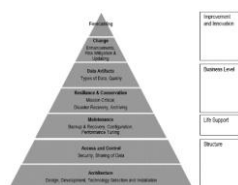
space. Haigh (2009) stated the prophecy took several decades to happen. This reformation of thought navigating through the database system still resonates today.

We return briefly to systems thinking as applied to database systems. Emergence as defined by Bertalanffy (1969, p.55) identified the whole as being more than the sum of all the parts and Kauffman (1995, p.24) gave an example of this as 'life' being more than a collection of molecules. This definition can be applied to the database system being more than the sum of the ecosystems. There are many interconnected parts within the database system and their interactions are unknown. Chaos theory explained that small changes of the initial state can have a profound effect on the output, which has unpredictable and complex outcomes. Johnson (2009, p.40) argued that chaos is an example of non-linear dynamics and where the outputs of the systems vary erratically and seem to be in a random way. Gleick's (1998, p.5) insight into chaos argued that this changed how decisions were made. Understanding the parts, Capra (1997, p.29) argued, enables understanding of the whole complex system. The database revolution had not yet reached a point where a grand unified theory existed, something Gleick (1998, p.7) described as the holy grail of science. Johnson (2009, p.39) discussed complex systems as a tendency to move between different types of output, making it appear complicated.

Every database was yet to have a place in the universe where pure data governed the theoretical improvements and innovation. Following Gleick's (1998, p.308) discussions and applying them to databases, database entropy increases the longer the databases exist, even when they are administered well and the data ordered. Databases need to be looked at holistically to understand the emergent behaviour from evolution.

The implications for this study of this aspect of the literature, are that this is a wide ranging field which covers many aspects of the socio-technical environment affecting all database operations and management. This diverse complex environment is rapidly changing and has many aspects that need to be considered. System thinking enables a holistic view of the situation to help with understanding complexity.

## 2.7 Database Technical Facets



The DBMS itself is constructed of many components. The anatomy of the database system was shown in Figure 1.3 and is discussed in this section.

This section covers the conception of the database technical design to demonstrate the initial management processes required. The four pillars of structure, life support, business level, and improvement and innovation, are discussed. Within the pillars there are seven layers each of which are discussed in turn: architecture; access and control; maintenance; resilience and conservation; data; change and forecasting.

### 2.7.1 Architecture (Structure Pillar)

The initial stage of architecture design is the requirements analysis and specification. This is the blueprint of the product which sets out the scope of the development that the customer ratifies. Functionality, data structures and non-functional specification like security, performance requirements, usability and documentation for the development team are included in the architecture design stage. Storey & Goldstein (1993, p.25) referred to design as an artistic and intuitive process; however there was a requirements engineering process which should be



followed. This process, as defined by Kotonya and Sommerville (1998, pp.116–117), has four stages: mutable requirement; emergent requirements; consequential requirements and compatibility requirements, due to evolution. These stages evolve as the assumptions are further understood.

DBA team functions often have to support databases where they were not involved in requirements gathering. Blasis (1977, p.232), in a very early report, mentioned issues that are still current, where teams were not included in the initial planning and design and the implementation of the database left to the later stages. The DBA would only be expected to sort out database access.

Database architecture and development were the foundation stones of any database system which provided functionality for applications such as financial operations, HR systems, ecommerce systems, science research analytics, intranet and internet sites. There has been an explosion in database architecture types, designs and methods. Gray (2004) discussed revolutionary database architecture changes including self-managing, self-healing and continuous availability of information. Hellerstein *et al.* (2007, pp.142–3) suggested database management systems architecture was not as well understood, documented or communicated as it should have been for applied database systems.

Connolly and Begg (1995, p.417) highlighted three phases of database design: conceptual, logical and physical. These three stages were essential for database systems following best practices. Software architecture was defined by Quatrani (2000, p.153) and looked at the behaviour of the system including physical attributes and collaborative elements. Shu *et al.* (1983, p.161) discussed the difficulties in database design and the collection of relevant information. Shu *et al.* separate this information into two categories, those processes which used data and

the data used by processes and points out that the process required and the data requirements needed to be specified at the outset.

Society's demands for a broad spectrum of functionality and types of data have led to a variety of database systems. Examples of the advanced database systems hybrid types were given by Lungu *et al.* (2009, pp.94–95), who also identified key users in database design (p.92). There might be additional stakeholders such as project managers and government. DBMS designs can be problematic due to the difficulty of including all the process requirements of the application and user data requirements (Batini *et al.* 1986, p.325).

It is necessary to decide which database management system is the right tool for the functionality required for the development. Yuhanna *et al.* (2009, p.2) stated that it was important to have the right tool for the right job. The Forester report (2009) listed the enterprise level contenders in the industry highlighting the offerings, suitability and issues for different functionality. Once the systems have been designed and created, the system's requirements involve not only the emergent properties of the physical design but also reliability, maintainability, performance, usability and security (Kotonya & Sommerville 1998, p.13). Olofson (2015) discussed the need to integrate transactional operations with analytics to create one platform that deals with blended database functionality and mixed data usage.

Scalability is a key consideration when designing new database systems, and Hull (2013, pp.54–58) discussed technology and maintenance obstacles for achieving scalable systems. Hull described ten obstacles to scaling beyond optimisation speed, which could be avoided by best practices.

The database landscape is currently in a state of flux with a shift towards Database as a Service (DBaaS), an emerging model where an individual database is migrated

into a cloud infrastructure in which multiple databases share resources, or models where infrastructure was sacrificed to use that provided by cloud vendors. DBaaS as defined by Oracle (2011, p.5) enabled database functionality to be offered to more than one consumer as a service. DBaaS was built on cloud computing through virtualization. It created cost savings as the new environment allowed databases to share platforms rather than on the traditional model which often consisted of one database per customer per database server. Mell & Grance (2011b, p.2) believed services could be rapidly provided with a minimal level of management. The 451 group believed cloud was a model for service delivery and consumption (2011, p.1). This provisioning of “on demand” cloud service provided elasticity, rapid provisioning, multi tenancy, scalability and a pay-per-usage basis, creating a different architecture on which to host database systems.

Oracle (2011, p.9) argued that DBaaS was a paradigm shift that impacted the organizational landscape, which affected the deployed ecosystem. Improvements allowed self service provisioning, utilization transparency and the satisfaction of strategic business goals such as control, flexibility and agility.

An emerging market for database systems is cloud computing (Otey 2010). The cloud database service allows for elastic provision of resources and chargeback for actual resource usage. The Berkeley report, ‘Above the Clouds’ (Armbrust *et al.* 2009, pp.14–18) discussed ten obstacles and opportunities within cloud computing, these being: availability of service; data lock-in; data confidentiality and auditability, data transfer bottlenecks; performance unpredictability; scalable storage; bugs in large-scale distributed systems; scaling quickly; reputation fate sharing (where bad behaviour of a customer can affect the reputation of the entire cloud) and software licencing. Some of these areas still have a profound effect on the adoption of cloud services.

Aslett (2015) argued that vendors' DBaaS strategy was important and traditional concerns about security and interoperability were being balanced with the DBaaS offerings

## 2.7.2 Access and Control (Structure Pillar)

Security of data is a primary concern of the database administrator and requires management processes to be in place. The protection of the data asset is critical for the organizations and the users. Data security is key to avoid misuse and damage, disastrous consequences for the organization as a whole (Bertino & Sandhu 2005, p.2). Information security regulations were introduced by the government to manage many new threats such as terrorism, globalization and growth of the internet (Kayworth & Whitten 2010, p.163).

Database security breaches occur as a result of cybercrime. The malicious attacks might aim: to try to gather sensitive information, manipulate database information, change system level commands, cause 'denial of service' attacks, or create blind attacks that create a database user to observe or change data in the database which the user was unauthorised to view or amend. These attacks might embarrass the organizations or lead to financial loss, loss of stakeholders' confidence, identity theft and defence risks. Bertino & Sandhu (2005, p.2) stated a complete solution had three requirements (protecting against unauthorised disclosure, improper modification of data and recovery of systems after errors or attacks), thus protecting the quality of the data. Organizations follow some or all of the regulations for security standards: ISO27001, Data Protection Act (DPA), Payment Card Industry Data Security Standard (PCIDSS) and General Data Protection Regulation (GDPR).

Anderson *et al.* (2009, pp.8–9) highlighted the sheer volume of databases that control our lives. A number of losses of data have already occurred, such as credit

card records (Jewell 2004) and personal and financial data losses (The Guardian 2015). Losses of data have also occurred when physical copies were required in other locations.

There are a number of areas where there were data security risks. The Data Security Survey (Mckendrick 2014a) reported that the highest risk to enterprises was human error. Wagner and Dittmar (2006, p.8) expanded on this, highlighting risks of weak control, fatigued, distracted and malicious operators, and the implementation of manual processes. Alzain & Pardede (2011) highlighted security risks for organizational data privacy. Agrawal *et al.* (2009) discussed further the data outsourcing paradigm that had technical and economic advantages and proposed a scalable security algorithm.

Kayworth and Whitten (2010, pp.163–164) discussed the primary objects for a security strategy which affect any organization: to balance security and business needs, to ensure compliance and cultural fit. These affect database systems management practices and procedures. Database security frameworks were a core requirement within operations to ensure the protection of data was maintained. Pavlou and Snodgrass (2008, p.30:3) highlighted the need to find out “what data was altered”. A database server has various layers of security. This determined what activities the user was allowed to do. Imran & Hyder (2009) mentioned that some access was only required for certain time periods so decisions on how to manage that needed specifying. They also raised the disparity between the aspects and features contained in secure databases.

User applications accessing databases require developers to be security aware when writing the applications and Magnabosco (2009, p.89) expanded on the use of encryption and cryptographic keys for providing data security. Auditing the database

servers regularly helped mitigate the risk and added data protection. Wagner & Dittmar (2006, p.1) noted that with the Sarbanes-Oxley Act of 2002, auditing of security events became a requirement to restore confidence in reporting and stop fraud for investors. They also identified unexpected benefits through increased standardisation, documentation and automation.

Auditing of databases is a process which would ensure that no user accounts would be left enabled on the servers when they were no longer required, and would also allow permission for the database objects to be checked. This continual observation of action (Liu & Huang 2009, p.982) requires formal procedures to be aware of events within the DBMS.

Recent events where Barclays Bank details were stolen and sold (BBC News 2014) highlight the need for database forensics. This is the application of computer investigation and analysis techniques to gather database evidence suitable for presentation in a court of law, for which Fowler (2008) presented a suitable methodology. Standardization and a best practice method are required.

### **2.7.3 Maintenance (Life Support Pillar)**

Database maintenance is required to ensure the database system is configured, recoverable and has good performance. This section delves into this need but first clarifies the nature of database administration:

*“The term database administration is often confused with the term data administration, but these are two distinct organizational activities. Database administration is primarily a technical function and often supports data administration. Data administration is the establishment and enforcement of the policies and procedures for managing the company's data as a corporate resource.”* (Kahn & Garceau 1985, p.88).

The life support system for the database is provided by the people who maintain the database servers. It is prepared for known failures such as the database becoming corrupt, the disks becoming full, performance degradation and data recovery following a catastrophe. Haerder and Reuter (1983, pp.290–291) mentioned three types of failure: transaction failure, system failure, and media failure. Mullins (2012, p.409) gave examples of database structure integrity and semantic data integrity as a type of failure. To ensure reliability of the maintenance standards, these should be documented, regular tasks automated and the server should manage itself as much as possible. In survey findings Blasis (1977, p.233) reported that standardization and documentation was not always completed if at all.

The backup strategy should be documented to meet organizational and legal requirements to allow recovery up to the point in time of a failure, for recovery or for historic data retrieval. Ideally all maintenance should be performed following best practice. To reduce the risk of failure, it is best to have a maintenance strategy and plan (McGehee 2009, p.18), which was divided into small tasks, automated where possible. Database backups are considered to be a part of the maintenance plan. A key issue documented by McBath (2002, p.4) was that backups are the last chance of recovery. It was not sufficient to only have backups without any other process in place to test the validity of these backups once they have been taken. When catastrophic failure occurs, McBath (2002, p.4) pointed out the consequences of not being able to restore from backups and the reasoning behind having a second check.

For any database, automatic alerting is required to protect the database against failure. Monitoring the error logs for warnings and errors allows alerts, a response to an event (Knight *et al.* 2007, p.113), to be raised to the operators. Alerts can be for failed database jobs (Woody 2002, p.200), hardware failure or when certain

performance conditions occur. In certain circumstances automatic fixing of the errors can occur.

A large section of maintenance involves dealing with performance tuning which requires detailed analysis of hardware, software and application. As databases age, the volume of data often increases and the number of tasks carried out multiplies, so regular reviewing of performance becomes required. Whalen *et al* (2001, p.6) identified a series of steps to help solve performance problems. Fragmentation of database files and indexes occurred over time through user updates (Sokut & Iyer 2009, p.14:3) and reduced the performance of the server. Performance tuning was discussed by Haerder & Reuter (1983, pp.289–290) and Bolton *et al.* (2010, p.357), although King's (2015) survey findings stated database performance tuning was often done manually.

Consistency within database management can be obtained through a run book (Woody 2003). Run books help define, build, orchestrate, manage and report on workflows to help in the management of databases; and can help management through documentation of best practices.

A survey on IT resource strategies (Mckendrick 2014b) showed that the database management activities taking up IT budgets are upgrades, patching, availability, making copies of database information, security performance tuning and diagnosis. King (2015) stated performance (system and data availability diagnostics, optimising and tuning) and maintenance (backups, alerts, integrity checks, defragmentation) were where DBAs spent most of their time. Jones (2006) argued that maintaining legacy software could be labour intensive.



#### 2.7.4 Resilience and Conservation (Business Level Pillar)

Availability of the data is maintained through good quality maintenance and additional features such as high availability, disaster recovery or archiving. Requirements specification for high availability and disaster recovery need to ensure the design met the customers' needs and budget. Mullins (2012, p.822) defined availability as the percentage time that the data resource is accessible compared to expected accessibility which ITIL (Information Technology Infrastructure Library) also include as a key metric for managing the operational toolset. Extreme fault tolerance systems such as those developed by NASA engineers require the highest availability.

Resilience is defined as the capacity of ecosystems and populations to return to a previous state after they have been disturbed. Planning gives organizations a chance of survival after a natural or human-induced disaster, system failure or infrastructure failure. Resilient systems are required to stop loss of revenues, loss of productivity, and loss of the company's reputation. The requirement for resilient disaster recovery (Choy *et al.* 2000, p.277) was determined by the organization's need to protect against site failure, network failure, storage failure or data loss caused by faulty code. All of these eventualities should be covered within Business Continuity Planning. Armour (2015) discussed the business continuity revolution and argued that best practices are wrong because no other approaches were compared and that there was a lack of empirical evidence confirming that best practices improved recoverability and preparedness. He concluded that organizations may fare better but highlighted the lack of communication being a critical gap.

Data and database archiving have different properties. Data archiving, Müller (2009), argued archiving is the storage of data sets and electronic media over time,

although Mullins observed that archiving is required for recovery purposes. Mullins (2012, p.821) defined database archiving as the process of removing data from the operational databases. The archiving, preservation, best practice and management of digital data was being researched holistically by the Digital Curation Centre (DCC). They were looking to preserve data for future generations. There was the need for standards to be put in place to assist with conserving data on a huge scale, the complexity of which needed to be addressed. Berman (2008, p.50) compared data preservation to that of physical infrastructure where this ought to be stable, predictable, cost effective and sustainable. Will the data after 30 years mean anything or has the structure, the purpose of data, the meta-data been catalogued to go with the backups? The implications of saving our history for the future were high. Moore (2010, p.189) argued that the retention of data could enable new discoveries. After the data has met certain standards it is archived for other researchers to use (Ailamaki *et al.* 2010, p.73).

The data explosion of digital data poses new issues for storage and transportation of data globally for analysis for many organizations such as the European Centre for Nuclear Research (CERN) (Viekzke 2009). Berriman *et al.* (2011) discussed the tools required for astronomy data to survive in the archive by looking at emerging technologies, compute infrastructure, cultural changes and educational changes. Thus the need for reliable, resilient, conserved data at speed with a never ending volume of new data that was easily accessible required new systems and procedures to be established to deal with the holistic nature of data.

### **2.7.5 Data (Business Level Pillar)**

Data is the fundamental asset required for a database system to exist. Database systems provide a storage medium for data and provide secure access to that data whenever required. The Economist report (Anon 2010, p.11) on data talked about

the immense volumes of data produced and pointed out the problem of accessing the relevant data easily and quickly. O'Brien described the process whereby data was processed to become information that human users could understand.

Checkland and Holwell (1998, p.90) suggested this then progressed from facts, data selection 'capta' and information to knowledge. Aiken (2016, p.22) shared a data management practices hierarchy stating that five data practices (data governance, data quality, data management strategy, data platform / architecture and data operations) only work well when they are all applied together.

Data driven decision making in businesses is increasing with big data. Davenport (2014, p.18) argued big data changed technologies and management processes. Kimball (2011, p.3), who defined one of the classic data warehousing models, discussed the evolving role of the data warehouse with the classification of three seismic events shifting the provision from that of historical data, customer behaviour data, to massive quantities of machine generated unstructured data. Data was used by organizations to describe user's lives and habits. Research into the data research lifecycle was undertaken as part of Institutional Data Management Blueprint (IDMB) project (Takeda *et al.* 2010), to help researchers with data curation and data management. The IDMB Project argued that for both current and future demand, no coherent data management approach existed (Takeda *et al.* 2010, p.3).

Magnabosco identified three other types of data: personal, identifiable and sensitive (2009, pp.20–21). The explosion of ubiquitous data on the web had infinite possibilities for the database system and research field (Agrawal *et al.* 2009, p.57). Blackburn (2001, p.1) discussed the ethical environment and Goguen (1999) argued there were ethical issues with data collection and usage which were exemplified with the data access layer structure.

Colwell cited in Denning (2002, p.16) discussed issues on the avalanche of data particularly in scientific research and how the presentation of this volume of data visually helped people understand the complexity. Data factories helped reduce complexity through the orchestration of data services, transformation of all types of data, storage movement, management of datasets and data lineage.

Managing data required some level of governance i.e. authority, and control planning, monitoring and enforcement (Mosley *et al.* 2009, p.39). Jagadish *et al.* (2014, pp.86–94) mentioned there were technical challenges to be addressed throughout the database systems lifecycle. The survey findings of Mckendrick (2015, p.23) reported

*“Real time enterprises need highly responsive data environments [...] Management and monitoring are essential to keeping data environments responsive.”*

A process to transform data meant that data input by the user or data sent by other systems could be manipulated to comply with the organizational rules, database rules and structure. It could help improve the quality of the data. Sharing of data between databases required an analysis of the data migration needs and identification of the requirements. Landrum (2009, p.66) provided a summary of the questions that needed to be answered for requirements gathering. He also mentioned the need to check that versions and editions of the software used were the same. However, some data sets were predefined and shared like the mapping and geographic data from Ordnance Survey. For the data to have reached the database some kind of data validation against business rules was likely to have taken place (Loshin 2009, p.100).

There are many long term (Kahn 1983) and current complexities related to data management and data administration, and as Aiken *et al.* (2011) suggested, data management is still evolving. The Claremont report on database research (Agrawal *et al.* 2009, p.65) highlighted concerns that were important to the community regarding the increasing technical scope, processes and keeping track of the field that was important to the community. Other surveys previously undertaken provided some insight, and highlighted the rise of database administration, with an unclear direction of the future path (McCririck & Goldstein 1980; Gillenson 1982; Gillenson 1985; Gillenson 1991; Aiken *et al.* 2011; Mckendrick 2013).

Aiken *et al.* (2007) carried out a study to look at improving organizational data management practices by creating a roadmap. Their suggestion for improvement highlighted the need for a formal feedback loop, noting that data management was focused on business areas rather than from an enterprise perspective and changing the perception of viewing data as an asset rather than a maintenance cost. The Beckman Report (Abadi *et al.* 2014, p.61) argued that the database community were now part of a “data management game” .

Villar and Kushner (2010, p.25) stated that organizations’ behaviour was like humans when achieving goals and mapped the hierarchy of data needs on to Maslow’s stages of human development. The aim of this comparison was to provide data quality and analytical data whilst providing some immediate results and a documented roadmap.

Berman (2008, p.50) discussed the need to have stable, predicable data infrastructure which was cost effective and sustainable. Ailamaki *et al.* (2010, p.89) discussed the lack of tools for scientific data management and a concoction of application specific solutions with some built on top of commercial DBMS. Szalay

and Blakeley in Hey *et al.* (2009) discussed the need for template and best practices to deal with these volumes of data.

### **2.7.6 Change (Improvement and Innovation Pillar)**

The database is continually evolving and changing with data being added, modified or deleted and added new features. Watzlawick *et al.* (1974, p.1) argued that persistence and change needed to be considered together. To manage this effectively changes required practices and procedures. Thus change management (Kotonya & Sommerville 1998, p.123) involves procedures and standards necessary to manage changes to system requirements. For management of change to take place there should be events such as market changing or incidents that caused a change catalyst, Pettinger (2000, p.225). Child (1983) suggested that change was ubiquitous in modern industrial society with the structure requiring to be maintained during these periods. Whilst Capra and Luisi (2014, p.315) argued technology was a fundamental part of being human and the shape of human nature was affected by technology.

The Independent Oracle Users Group (IOUG) has recently completed several surveys that advanced understanding of the complexities of database systems. The IOUG report (Mckendrick 2011b) entitled 'Managing the Rapid Rise in Database Growth' identified the importance of database change management practices. Improving the systems through enhancement, the mitigation of risk and updating all require change.

There are two types of technical changes: planned changes and emergency changes. Emergency changes may result in unplanned downtime (Joch 2007) which proactive working may prevent. The change control process should be planned and well documented and should disclose that; all proposed changes were

fully documented; all changes were tested to ensure that: the functionality was not affected; the likely impact of the change on the production system identified whether the level of impact on the production system was low, medium or high risk; that the management team have approved the change; the change should be scheduled so changes do not conflict on the same servers and to enable ease of identification if problems occur; and there must be a rollback plan.

To manage changes effectively there are various key factors which might give a higher level of accuracy. Items such as risk assessment of the rest of system, detailed deployment plans, rollback plans and documentation needed to be considered. Kotoyna & Sommerville (1998, p.128) called this traceability information. A side effect from the organizations commitment, Pettinger (2000, p.233) argued, that the process could generate a life of its own and highlight problems and opportunities.

Mullins (2012, p.243) argued the need to meet customer expectations in the light of constant changes which might have business impact. Similarly, Schein discussed increasing organizational effectiveness through controlling desired outcomes whilst noting people's resistance to change (Schein 1980, p.65). This technical management of changes to the database systems was important to protect the database and data used within production systems.

### **2.7.7 Forecasting (Improvement and Innovation Pillar)**

The final layer, forecasting, in the improvement and innovation pillar, reports on trends in capacity and performance workloads. Forecasting database systems future requirements, becomes entwined with future developments of the database server, application software, the user requirements, and the organizational requirements. Data manipulation predicating and reporting on business trends

whether they were financial for stock levels for health, astronomy or managing workload levels could lead to a rethink on the requirements. Forecasting utilised the data by transformation, sharing, reporting and analysing to show trends and patterns to ensure systems were proactively maintained and improved. The Beckman Report (Abadi *et al.* 2016) highlighted that visual analytics was essential to cope with large volumes of data in the database. There was a higher rate of data change in some systems, such as financial markets and scientific research. Where more data existed it could be used in prediction models (Johnson 2009, p.113). Although Johnson was referring to financial markets and their complicated dynamic systems, database systems also exhibited this behaviour. This was not only because they could store the financial data but due to the environmental factors that were associated with the storage of this type of data, the hardware physical attributes and the organizational requirements.

Capacity management of a database system should identify whether a database could deal with the current workload on the systems and to continue to keep dealing with the workload. Capacity management looked at historical data and identified future trends which enabled proactive management forecasting for the short and longer term.

It is one thing for a DBA to calculate the requirements using automated tools and scripts but it is another presenting this to the business to make informed decisions. Business Intelligence is about collecting data, then transforming data, analysing data for reporting or data mining and presenting that data to the end user to support the organization's needs (Mundy *et al.* 2011, p.xxxvi). Davenport (2014, p.18) argued that discovery, agility and speed of reaction gave businesses the upper hand.



Providing reports which could be quickly digested through visualisation adds increased value to the business. These visualizations could share common ground (Heer *et al.* 2010, p.67). Dilla *et al.* (2010) also discussed visual reporting.

Databases Systems were increasing the number of standard reports that were available to the user in addition to custom written reports.

The research aim is to provide improvement and innovation for the management of database systems and this is discussed further in Section 2.9.

## 2.8 Database Management Lifecycle Frameworks

The management of database systems is sometimes carried out or linked to existing frameworks, techniques and models not specifically designed for database management but for IT as a whole. This section looks at some of these frameworks, techniques and models in connection with the database system.

This section aims to provide details of the current examples and patterns of overarching frameworks and methodologies that play a significant part in the day to day life of the database. Frameworks are a way of helping manage certain components within the database system. The section includes architectural frameworks, agile management, management of the database service and data management frameworks.

### 2.8.1 Architectural Frameworks

A few frameworks exist that have an impact on how databases are designed at the enterprise level. Process management, information management and storage management are considered. One such category was enterprise architecture. This covered aspects such as goals, strategies, vision, operations and technical capabilities.

The Open Group Architecture Framework (TOGAF) (1996) was a framework for enterprise architecture considered the de facto standard that enabled designing, planning, implementing, and governance for information technology. It covered areas such as business, application, data and technology.

The Zachman Enterprise Framework (1987) was another widely known architectural framework which was now called framework2. The framework was a blueprint for an organization information infrastructure. Some enterprise data architects use this framework taxonomy as a tool to help design. Sowa & Zachman (1992, p.590) stated the taxonomy helped describe the information system clearly which raised sight of issues often overlooked.

The Capability Maturity Model (CMM) (Paulk *et al.* 1993) was used to help with business architectures and to improve IT processes related to development. The model had five levels of maturity, initial (chaotic), repeatable, defined, managed and optimizing. The capability maturity model aimed to help an organization control the software process, effectiveness and predictability for the business.

The information lifecycle management (ILM) was a framework for understanding information needs (Tallon & Scannell 2007) and was based on the premise that information had a natural lifecycle. Organizations were required to evaluate data and the risk of its unavailability. If data was no longer used or if it was perceived to have less value it should be archived to cheaper media or deleted. ILM was adopted by storage organizations which had evaluated the business value of data and placed the data in the most appropriate infrastructure (Peterson 2004, p.4). Chen (2005) suggested ILM was coined by the IT industry to improve resources and maximise value. Reiner *et al.* (2004) discussed the EMC Corporations viewpoint. Data was classified active, less active, historical or archive. Peterson (2004, p.3)

argued that automating a bad process would not solve the complexity of the problem. Bhagwan *et al.* (2005) discussed improvement with the optimization of data placement based on a time varying manner. The whole process starting with user practices enables ILM to deal with all aspects of data.

This section has described proven architecture frameworks that IT related projects methods follow. Choosing a database based on features could be influenced by many factors. Richardson (2015, pp.54–61) suggested a framework model to evaluate database features rather than classifications of databases.

## **2.8.2 Agile Management of Databases**

A management and development paradigm which has emerged over recent years is Agile. Agile is a quick moving process that can change direction quickly. This method allows responses to unpredictable events in a turbulent environment (Fowler 2005). The Agile software alliance was formed in 2001 and the principles of Agile were born. A manifesto for Agile software development was created by Beck *et al.* (2001) setting out twelve principles.

These principles can be applied to database management as a new way of working. When the database development team subscribe to this method collaboration takes place amongst other teams. This new development method was significantly different from the traditional waterfall method. The waterfall model was linear and sequential with predefined goals and deadline without iteration. Agile is a lightweight framework with a different philosophical outlook to the traditional limitations of the waterfall development. The focus was on rapid delivery, reducing the risk to the business, dealing with rapid change of the database landscape and value for the business. Gregory *et al.* (2015) discussed the challenges of adopting an Agile

approach within organizations. The adaptive process with continual feedback loops increased the visibility of this process to the business.

Agile is more people orientated rather than process driven. It delivers working software in collaboration with customers over comprehensive documentation and responded well to change. Agile dealt with issues such as priorities and vision of management, technology mismatch, and poor documentation that had caused software issues and development issues in the past.

*“The goal of the agile data (AD) method is to define strategies that enable IT professionals to work together effectively on the data aspects of software systems. This isn’t to say that AD is a ‘one size fits all’ methodology. Instead, consider AD as a collection of philosophies that will enable software developers within your organization to work together effectively when it comes to data aspects of software based systems.” (Ambler 2003, p.3)*

The use of Agile started with the development teams but expanded to include database administrators to help improvement of the analysis and feedback. An Agile database administrator is defined as:

*“An Agile DBA (Schuh 2001) is anyone who is actively involved with creation and evolution of the data aspects of one or more applications. The responsibilities of this role include, but are not limited to , the responsibilities typically associated with the traditional roles of database programmers, database administrators (DBA’s), data testers, data modellers, business analysts, project managers, and deployment engineers” (Ambler 2003, p.11)*

Ambler continued to discuss evolutionary development as chaotic although order could come from chaos particularly when the situation was addressed by people who had become familiar with it.

Agile techniques have added a new dimension to database management which needed to be considered when reviewing the paradigms currently in place which affect database management. Armour (2015, p.38) suggested that complex unpredictable situations were being dealt with through Agile project management.

Agile database management techniques (using shorter sprints with a set time limit for repeatable work patterns) could be used to help improve the effectiveness of database management. Agile could be used for database development – VersionOne (2013) found that 37% of respondents were using Agile for “76-100% of projects”.

In infrastructure projects Agile approaches are now being implemented through DevOps (IT Revolution 2015). The other lean management techniques such as the Vanguard method (Seddon 2003) are also used within organizations to help management of database systems.

### **2.8.3 Database Management Service**

The operational state of databases was probably the most well managed area of the database system due to the potential impact to the business if databases became unavailable. Challenges reported in an Independent Oracle Users Group survey (Mckendrick 2011a) for operations included: an increase in the number of databases, databases of larger size, a reduction in the number of older systems being retired as they are kept in operation for longer, as well as more features and functionality being included in newer systems. Stonebraker *et al.* (2013) also discussed the operational challenges of new features which add to the complexity.

Cloud services were such a feature that added complexity and these services had no accountability which added risk for organizations (Mourad & Hussain 2014).

There are various IT service management frameworks: ITIL (The IT Infrastructure Library), MOF (Microsoft Operations Framework), CMM (IT Service Capability Maturity Model), FITS (Framework for ICT Technical Support), SaaS SDLC (Systems Engineering and Software Development Life Cycle Framework) and CobiT (Control Objectives for Information and related Technology).

A paradigm that was often leveraged against databases and their management was the Information Technology Infrastructure Library (ITIL), which was adopted within many organizations worldwide. ITIL<sup>1</sup> was a framework of guidelines which organizations had adopted for service management. The philosophy of ITIL was that it worked for both small and large organizations and was scalable (Macfarlane & Rudd 2001, p.4). It was defined as follows:

*“ITIL is a public framework that describes Best Practice in IT service management. It provides a framework for the governance of IT, the ‘service wrap’, and focuses on the continual measurement and improvement of the quality of IT service delivered, from both a business and a customer perspective” (Cartlidge et al. 2007, p.8).*

Pollard & Cater-Steel (2009, pp.165–166) discussed the use of ITIL within government as a means of providing cost effective IT within the computing centres. ITIL was prescriptive with no specific details on how to implement it. Cannon *et al.* (2007, p.3) argued that properly conducted, controlled and managed processes

---

<sup>1</sup> ITIL was derived by the UK Government, being published between 1989-1995

were required. There were no formal checks which were used to validate whether an organization was ITIL aligned or not. ITIL was continually evolving and in 2007 ITIL 3 was introduced adopting a new service model to provide an end to end approach. There were five sections of ITIL core which were service strategy, service design, service transition, service operation and continual service improvement. To enable the continual service improvement (CSI) model Cartlidge *et al.* (2007, p.35) gave a seven step improvement model, a service measurement model and service reporting process. Potgieter *et al.* (2005) found evidence that ITIL was producing effective best practice results for customer satisfaction with operational performance.

There are general techniques that may be applied by the business to specific areas of problems within database management. Problem management was a core feature within operational management. There have been numerous suggestions of methods of solving problem situations (De Bono 1990; Kepner & Tregoe 1981; Kepner & Tregoe 1965, p.73; Checkland 1999, p.155). Senge identified two types of problem. One type (1990, p.95) which “limits growth” by succeeding in one task caused secondary effects which slowed down success. The other type (1990, p.104) was “shifting the burden” when a problem was difficult to solve an easy fix was undertaken to make the problem better but the underlying problem was not fixed. Ackoff (1981b, pp.20–21) discussed various other types of solutions that could be used for problem solving one of which proposed a course of action that offered an outcome that was good enough and another to dissolve the problem by changing its nature or the environment of the system surrounding the problem.

#### **2.8.4 Data Management Framework**

Data has become critical to everyday business as an asset and there was pressure to report a single version of the truth (Khatri & Brown 2010, p.148). Thus data

governance has emerged. Khatri *et al.* has identified five interrelated decision domain data principles, data quality, metadata which is data that describes data, data access and data lifecycle. Mosley *et al.* (2009, p.3) proposed a data lifecycle, in which data flows in and out of data stores and was delivered as information which could be used now or could be used in the future.

The Method for an Integrated Knowledge Environment, MIKE2.0, (Rindler & Hillard 2013) is an open source standard for information management. Ubiquitous information is a challenge for every organization and MIKE2.0 is a complete framework for information management, best practices, for business issues, and technology solutions.

The data management framework was called the data management body of knowledge DAMA-DMBOK guide Mosley *et al.* (2009). The DAMA-DMBOK guide set out some guiding principles and common organizational and cultural issues relating to data. The scope of the data management function described in the framework was a brief guide of concepts, goals, functions and activities and provided a holistic view of the data system (Mosley *et al.* 2009, p.12). The framework identified seven environmental elements which affected the system: goals and principals; organization and culture; activities; deliverables; roles and responsibilities; practices and techniques; and technology. The DAMA-DMBOK guide sections covered: concepts, descriptions of data management, principles and process provided the essential understanding of the area. These were all areas of database systems but the approach described here provided further insight into the area

A key factor of data governance was that the quality of data in the system was maintained, improved and accountable. DAMA (Data Management Association)



depicted data governance as 'the management roof', the controlling position over other data management functions (Mosley *et al.* 2009, p.39), with a data steward in control. The governance of master data management, a way of linking all management information to one file, followed a maturity model (Shankar & Menon 2010, p.20) .

The data ecosphere (Ambler 2007) encompassed data architecture, data management, data culture and data governance. Ambler suggested some key areas for improving data architecture such as data quality techniques, promoting teamwork over politics, adopting a lean governance approach and becoming as agile as possible.

## 2.9 Improvement and Innovation

This section looks at analysis, improvement and innovation and the effect of chaos and taxonomies on the holistic system to ameliorate the management of database systems. A reflection of the system boundaries will hopefully provide epistemological and ontological viewpoints on the system in hand and lead towards a unified theory for managing database systems which might lead towards database improvement and innovation of database management. A goal of the research is to improve the management of database systems and to potentially provide an innovative solution. It is necessary to understand what is improvement and innovation and how that can be achieved.

Improvement is adding value or making something better whereas innovation is the act of innovating, producing a new method or idea. Improvement of technology requires evolution of either, the technology, the processes or quality of tasks

completed. Basalla (1988, p.25), set out his theory of evolution with the concepts of diversity, continuity, novelty, and selection.

Improvements in database management and technology are shared amongst organizations transforming their culture. Basalla referred to this as a source of novelty. Technological knowledge is proliferated through the vendors and database community who had a rich diversity of ideas and techniques to improve the system. Basalla (1988, p.216) argued that technological progress had two requirements one of narrowly specified technological goals bounded by culture and time and the second that social, economic or cultural progress be separated for technological advancement.

Various improvement models already existed. The Kepner Tregoe model previously discussed was utilised within many organizations. The Deming<sup>2</sup> or Shewhart cycle often called Plan Do Check Act (PDCA) cycle cited in Smith (1997, p.132) was another such model for continuous improvement with a continuous feedback loop.

An evolutionary, cultural, technical or organizational change could result in innovation. There were many researchers and theories of innovation. Schumpeter (2010), in the 1940's, looked at long term benefits of innovation rather than short term, the role of monopolistic enterprises in promoting innovation and a process called creative destruction. Others such as Henderson and Clark (1990, p.11) created a distinction between component knowledge of the core design concepts and architectural knowledge explaining the ways components were linked together into a coherent whole. Christensen (1997) reviewed sustaining versus disruptive

---

<sup>2</sup> Deming was the father of modern quality control.

technologies emergence from innovation. There were many concepts and theories of innovation as well as various definitions of what innovation is.

Drucker (1998, p.3) saw innovation as entrepreneurship of a new or existing business. Twiss (1992, p.8) on the other hand discussed innovation in the market place. He defined the conception of an idea as invention and the transformation of that into ideas of economic benefit as innovation. Dodgson *et al.* (2008, p.2) argued that innovation was an exploitation of new ideas that were commercially successful which was more than invention. Innovative organizations as Peters and Waterman (1982, pp.12–13) had a whole culture which drove towards innovation. They adapted, transformed, changed and reviewed existing systems and processes. Peters and Waterman found that excellent innovative companies kept things simple, provided top quality work, engaged with employees, flattered their customers and allowed some chaos.

Dodgson *et al.* identified new paradigms for organizations and how they learn. These new paradigms affect innovation. A few new paradigms mentioned are knowledge as a source of competitiveness, organizational learning, and business process structures. Dodgson *et al.* (2008, p.62) also raised the move towards a flexible, inclusive and process driven approach which encouraged innovation. Kanter (2006) discussed the classic traps that organizations make and argued innovation could flourish by reviewing strategy, structure, process and skills.

Utterback (1996, pp.18–19) classified developments in innovation as new technology, changing leadership, a wave of technological change, new innovation from old capabilities, document design and shifting ecology of the firm. Utterback (1996, p.133) was of the view that the evolution of innovation often came in waves with incremental improvements in between. This clash of technologies bringing

technological discontinuity may mean splitting the approach to bridge the old and new simultaneously although this could lead to a diluted success (Utterback 1996, pp.190–191).

Beacham (2006, p.9) listed six approaches to innovation which could be due to new principles of management rather than for product and processes. For market place innovation this could be breakthroughs in applying technology or increasing functionality. For technological push innovation approaches these may be a new design, new business model or incremental. Beecham argued that innovation success was measured by the market place. Critical thinking of how the database management system's interaction with complex systems was created could be stimulated by the heuristic framework for corporate innovation created by Callahan & Ishmael (2005).

Drucker stated there were various conditions to innovation

*“1) Purposeful, systematic innovation begins with the analysis of the opportunities [...] 2) Innovation is both conceptual and perceptual [...] 3) An innovation, to be effective, has to be simple and it has to be focused [...] 4) Effective innovation starts small [...] 5) A successful innovation aims at leadership”* (2007, pp.207–208)

Drucker continued that the way innovation should not take place was by diversification or by being too clever but should in fact look to innovate for the current time.

Kauffman (1995, p.191) discussed technological evolution and the plethora of innovative possibilities through time that had various scales and complexity. The parallels for database management and organizational management could be seen

and the complexity of outcomes equally prevalent. The unrelenting progression of technology and change as the arrow of time moves further into the future, results in increasing entropy. Improvement and innovation in systems would not always produce a stable outcome but would continue to be chaotic and complex but progress in terms of the breadth of information and outputs available was being achieved.

The roles of humans in the data life cycle needs to be considered, Abadi *et al.* (2016, p.98)

*“classify people’s roles into four general categories: producers, curators, consumers, and community members.”*

Agrawal *et al.* (2009, p.61) stated this was an opportunity to reform data management. Agrawal *et al.* (2009, p.63) argued there were many database issues yet to be addressed in the cloud such as sharing of physical resources, data security and privacy.

Armbrust *et al.* (2009, p.7) argued that the database industry, which was dominated originally by technological trails such as transaction processing, now looked at customers’ chains, buying habits, ranking, and so on to understand the new market. This was a shift towards more business analytics, identified the database data as a key to the decision making toolset for improvement to both, the database usage, requirement management tools and organization.

Databases are gradually becoming more autonomous with science and industry. Feedback from the technology, science and industry user base improved the usability and usefulness of the database. The DBMS was a complex system

including a varied set of technologies. To allow for improvement Korth & Silberschatz stated

*“Increasingly, databases will need to deal with inherently imperfect and incomplete data. Therefore, database systems must emerge from their artificially simple closed world and join the broader world of human information (Korth & Silberschatz 1997, p.141)*

Database administrators continually strive to improve the manageability, quality and performance of the database. Improved database management techniques were important in all fields of usage. The analogous view only worked if the whole database system was considered. The technical tools within the DBMS now included many of these features but management of the overall process to achieve analysis, improvement and innovation was required.

It was possible that the improvement of the database management system could cause the Jevons paradox. Polimeni and Polimeni (2006, p.344) explained the Jevons paradox as an increase in efficiency was achieved as technology progressed there was also an increase in the consumption of natural resources. This paradox was caused due to complex systems utilising the new technological advances, whilst at the same time consuming resources. The increase in the demand for databases and data continued to explode due to the continuing increasing ease with which data could be retrieved and utilised. Improvements might only be part of the change process as this could lead to innovation.

Database systems are rapidly evolving. The new practices and procedures developed may allow progress of the scientific and business revolution.

*“most innovation, especially the successful ones, results from a conscious, purposeful search for innovation opportunities [...] Three additional sources of opportunities exist outside a company in its social and intellectual environment: demographic changes; changes in perception and new knowledge” (Drucker 1998, p.4)*

Pantula (2011) talked about the great statistical adventure and that statistics were key to innovation in a data centric world. Databases had been key building blocks that allowed statisticians to gain insight at speed from vast volumes of data.

There is a move towards creating self-managing DBMS for technical components such as indexes and memory (Holze & Ritter 2011). A current example of this is the new Microsoft SQL Server 2016 query store functionality (Varga *et al.* 2016, p.51). Holze and Ritter argued that knowledge of the sensors and effectors and systems behaviour within the DBMS requires documented manuals or experience of the DBA. Holze & Ritter purport that a systems wide model of self-management logic is required although the reconfiguration of the DBMS is a complex task.

In conclusion the complexity of the database system is continuing to increase and evolve. Research is required to gain insight into these changes and to try to improve and innovate the management of the database systems. The various aspects mentioned in the previous sections should be brought together to provide a holistic view on the management of database systems and best practices that are in use.

## 2.10 Summary

The discussion of the literature presented here depicted the wide area of literature connected to the management of database systems. It initially discussed the founding concepts of the DBMS, a way of management through best practice and

how best practice was a part of the information system. It then looked at how systems thinking could provide a holistic approach to managing database systems rather than the analytical approach of much former research.

From there organizational management areas were discussed followed by a discussion of the technical DBMS and its anatomy. Many frameworks were highlighted that database management use in practice. The final section shared various improvement and innovation strategies.

The literature dealing with views on variety of practices and procedures suggested that this is a field where no one management system could be adopted, thus further information and understanding of the complex issues involved in the gap is required. Best practice was discussed as a defined practice rather than a continually moving target.

As database systems developed to deal with a wider variety and type of data and the requirements increased so the complexity of the systems has led to issues that have not been satisfactorily resolved for all. Research into current operations and how the experiences of the people involved in the running of these databases are dealing with issues arising was sought. These diverse fields contribute to parts of the management of database systems and underpin this research. The literature discussed the components which are connected to the management of database systems although all have not been explicitly brought together before. Has progress been made dealing with the complexity merging these diverse fields together to successfully manage database systems in today's world - could data be controlled to fulfil requirements expediently?



The next chapter introduces the method used in this research. The method used was mixed method research starting with quantitative data collection followed by qualitative data collection.

## Chapter 3: Research Design

### 3.1 Introduction

This chapter describes the research design. The problem situation was set out in Chapter 1, identifying the management of database systems as complex. Many areas needed to be considered together when reviewing the whole system as one. Management tasks, technology, the changing environment, diverse and increasing data volumes and culture were a few of the areas that the research design needed to be able to analyse. A better understanding of how tasks were carried out and by whom was sought. The experiences of those involved in database operations and the business drivers could help with the management of database systems in the future.

The three types of research design considered were quantitative, qualitative and mixed methods. The use of comparative case studies was considered but this method addresses a different type of research question (Yin 2009, p.9). Case study research examines the 'how and why' type of research questions and is good for longitudinal research. (It was not practical to carry out a longitudinal study for the current research.) The research for this study required the use of frequencies in order to address research questions of the type 'who, what, where, how many and how much'. These are characterised by using a survey method. The 'how and why' research questions could be addressed using interviews. On this basis, a mixed methods approach was adopted. This research was not a longitudinal study.

Ivankova and Stick (2007, p.97) found that using quantitative or qualitative research methods individually did not capture both trends and details of complex situations, but together they provided a holistic picture. In Information Systems (IS) research

Venkatesh *et al.* (2013, pp.22, 35) argued that diversity in method was a major strength of mixed methods, but also that the choice of research questions, the purpose and the context governed whether or not mixed methods should be chosen. Lund (2012) argued mixed methods research could result in a more complete picture from the complementary quantitative and qualitative methods, and help with certain types of complex research questions. For the research presented in this thesis, mixed methods seemed an appropriate approach to enable the current usage of best practices globally to be assessed and a more in-depth understanding of the complexities gained from the fieldwork.

## 3.2 The Research Strategy

The type of design that this research followed was mixed methods, which Creswell defines as:

*“An approach to inquiry that combines or associates both qualitative and quantitative forms. It involves philosophical assumptions, the use of qualitative and quantitative approaches, and the mixing of both approaches in a study. Thus, it is more than simply collecting and analyzing both kinds of data; it also involves the use of both approaches in tandem so that the overall strength of the study is greater than either qualitative or quantitative research”* Creswell (2009, p.4)

This approach allowed both open and closed questions to be asked, drawing on multiple forms of data, together with statistical and textual analysis. Mixed methods were complementary due to the collection of various types of in-depth information, and this enabled the researcher to enhance and clarify the results. The mixed methods design can provide researchers with a better understanding through the

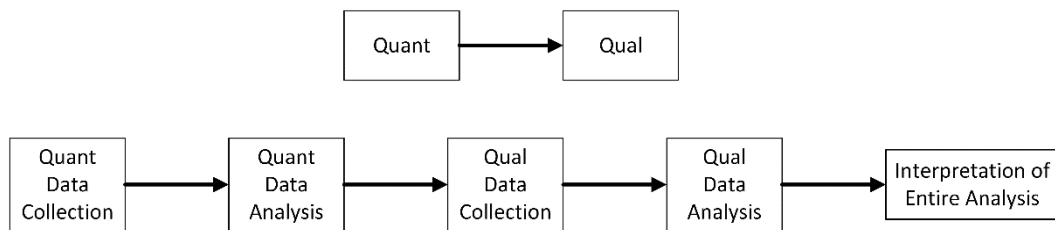
combination of statistical trends and stories. Venkatesh *et al.* (2013, p.24) argued that using mixed method strategies add to theory and practice for exploring rapidly changing environments, and is a powerful tool for IS research.

There are various types of mixed method approaches. Creswell & Plano (2011b) discussed four key aspects to consider regarding which type to choose: the level of interaction, priority, timing, and where and how to mix quantitative and qualitative methods. The major mixed methods strategies are convergent parallel design, exploratory sequential design, embedded design, transformative design and multiphase design (Creswell 2009, pp.209–10).

The approach used for this research was the sequential explanatory design (Figure 3.1). The other designs were not considered further as possible approaches after consideration of the research question, iterative reflection and reviewing the four key research aspects mentioned above. The strength of this method is in its simplicity:

*“A sequential explanatory design is typically used to explain and interpret quantitative results by collecting and analyzing follow up qualitative data”* Creswell (2009, p.211)

The approach (Creswell & Plano Clark 2011b, pp.119–123) is shown in Figure 3.1.



**Figure 3.1 Sequential explanatory design**  
(Creswell 2009, p.209)

**Key: Quant = Quantitative, Qual = Qualitative**

**A '->' B indicates a sequential form of data collection - the qualitative data building on the quantitative data.**

Surveys have been used to assess other database and data management methods in papers such as Aiken *et al* (2011) and Kahn & Garceau (1985). Other database surveys have also previously been undertaken (McCririck & Goldstein 1980; Gillenson 1982; Gillenson 1985; Gillenson 1991; Mckendrick 2013). Kahn *et al.* broke the database administration function into areas for data management and my research follows similar lines. Blasis (1977) however looked at database administration as a team function and the report creation, resulting from a survey, produced recommendations on how DBA implementations are reflected in reality. Blasis carried out in-depth questionnaires and face to face interviews.

Sequential exploratory design requires the use of multiple methods and the interpretation of two sets of results. An advantage is that sequential explanatory design can answer a range of questions and provide insight and understanding from the two methods. The strengths of using both quantitative and qualitative methods means that they can be used to overcome the weakness of one method and enable mixed narrative and numbers.

In this research, the findings from the quantitative data informed the empirical qualitative study. The objective of the research was to understand the phenomena and contribute to improvement and innovation. A concurrent design was not chosen

because it was not necessary to gain an understanding whilst the phenomena happened or to capture immediate impacts.

Potter (2006, p.84) discussed explanatory research which has two shared goals pertaining from epistemological and ontological aspects of enquiry, which suggested the goals of explanation were to explain the world and deal with complexity through reductionism. Reductionism is a viewpoint from the 'machine age' where elements were understood through breaking them down to their smallest parts.

*"Explanation is at the core of the positivist quest for knowledge. The word 'explain' comes from the Latin explanare, and means, literally, 'to level out' [...] the primary purpose of explanation is to produce a lawful account of cause and effect. [...] Positivism is highly reductionist – it smooths out complexity"* (Potter 2006, p.84)

Explanatory research uses both inductive and deductive methods, an approach which is becoming increasingly important (Guest *et al.* 2012, p.37). Induction and deduction have a positivist viewpoint in common. An inductive approach draws inferences from observations to make generalizations - an early advocate of this was Francis Bacon. In this research, applying this method allowed the researcher's own viewpoints and experiences to be partially removed. This should allow fresh insight into the problem situation.

Validity and reliability are important. Reliability is a measurement quality showing that the research produced repeatable results. Validity was defined as:

*" mixed method research involves employing strategies that address potential issues in data collection, data analysis, and the interpretations*

*that might compromise the merging or connecting of the quantitative and qualitative strands of the study” (Creswell & Plano Clark 2011a, p.417)*

Additional aspects that influenced this design were timing, weighting, and theorizing, see Table 3.1. The timing for this approach was sequential, which meant that there were two phases of data collection. The researcher needed an understanding of the problem situation before expanding knowledge of it. The weighting of the two phases of research (the relative importance of each section) was undecided until after stage 1.

There was an increase in usage of explanatory sequential design, which has been used less frequently than exploratory design in the past (Guest *et al.* 2012, p.200). Explanatory sequential design is about explaining the findings using qualitative research of the quantitative results. Guest *et al.* (2012) argued that this explanatory sequential design is complementary in nature allowing the qualitative data to shine and this design combines the strengths of quantitative and qualitative methods to address shortcomings of each method.

**Table 3.1 Aspects considered in planning mixed methods design amended from Creswell (2009, p.207)**

Timing	Weighting	Data Analysis
Sequential Quantitative first	Quantitative	Highlights common and unusual practices
Qualitative second	To be decided after stage 1	Participants selection and explanation both explicit and implicit

A dominant study is generally found where rigorous standards of data collection and analysis take place (Venkatesh *et al.* 2013, p.38) and a non-dominant is less rigorous in IS research though Venkatesh *et al.* recommend that both sets of data are rigorously analysed.

In this research, the scope of analysis was affected by the research questions, data and context. The quantitative phase was completed first and the data analysed. Subsequently after careful consideration, the approach to the qualitative phase was decided. When this had been completed and analysed, it was clear that this latter phase should carry the most weighting because the complexity needed further explanation.

An example of a 'follow up explanation variant' was the work of Igo *et al.*(2006). A 'participant selection variant' has the priority deferred to the qualitative phase – this approach is also called Quantitative preliminary design (Morgan 1998), and an example is May & Etkina (2002).

Creswell and Plano Clark (2011b, p.82) stated that the sequential explanatory design was

*“most useful when the researcher wants to assess trends and relationships with quantitative data but also be able to explain the mechanisms or reason behind the resultant trends.”*

After completion of the sequential explanatory data collection and data analysis the interpretation for the entire analysis stage commenced. In this stage of the research, a systems approach was used to complement analysis with synthesis. The final stage took a holistic approach to look at the database systems to explain the operations of components (Ackoff 1981a, p.17).

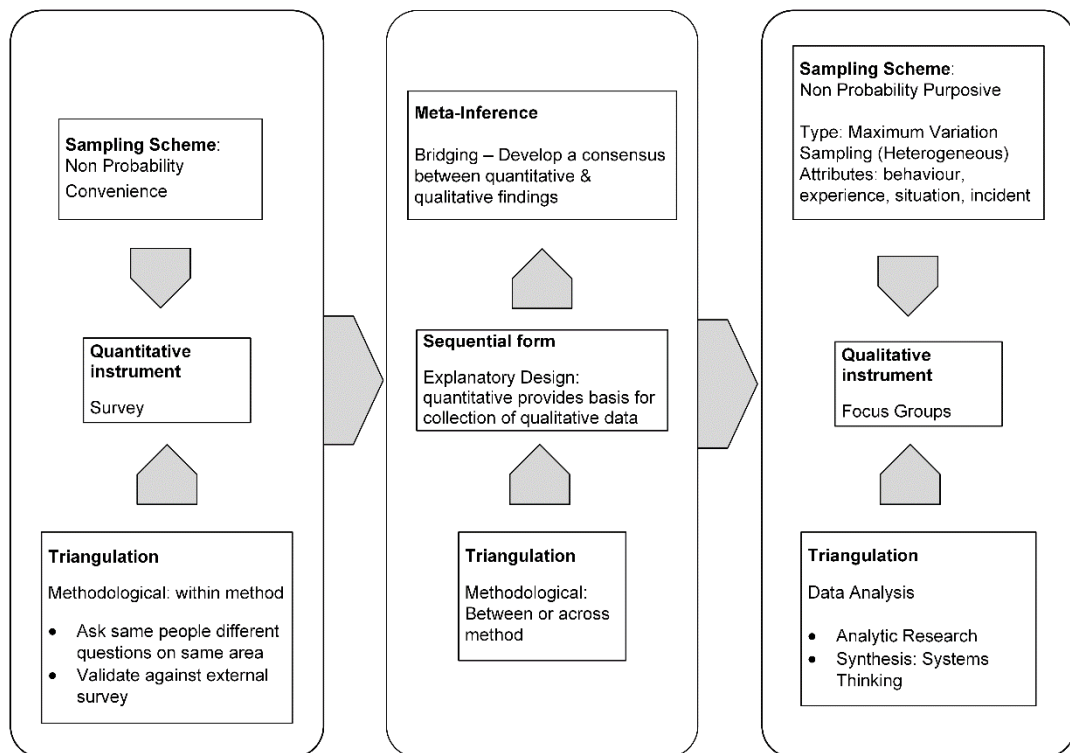
With any research undertaken there is always the risk of potential bias. The researcher's work responsibilities as a Database Architect and Senior DBA meant that industry related issues could have affected this research. Aware of this issue, the researcher relied on self-reflection to keep the research grounded. There are



potential limitations to this research as it is from the perspective of the database administrator. This viewpoint is not the only possible way to look at a database system, for instance it ignores the user perspective. The strength is in internal knowledge of database administration but the weakness is in other areas such as the user perspective. It could also be said that this close relationship between industry and research was advantageous in understanding the current issues. Validity of the research was gained through checking the accuracy of the findings, and reliability was determined through a consistent approach.

### 3.3 Mixed Method Summary

A summary of the method design is shown in Figure 3.2.



**Figure 3.2 Method design based on (Cameron 2009, p.150)**

Figure 3.2 gives a high level view of the research depicting the quantitative and qualitative phases which received inputs from the chosen sampling scheme and triangulation method. The meta-inference bridging was an output of the middle stage which fed into the qualitative phase.

### 3.4 Linking the Research Questions to the Research Methods

This research was governed by a set of research questions to identify the practices and procedures used by the database community as a whole, and then to find features and trends which existed in this complex area. Further in-depth data collection was required to provide both breadth and depth of understanding, which might contribute to improvement and/or innovation. Each question is connected to a research method and the retrospective connection in this research shown in Table 3.2.

**Table 3.2 Research questions and research method**

<b>Questions</b>	<b>Research Method</b>	<b>Method used</b>
To what extent are best practices and procedures utilised by the database community?	Quantitative In the majority	Survey
What are the complex interactions that are an integral part of the management of database systems?	Qualitative	Focus Groups
Is the adoption of best practices and procedures affected by the complex interactions that are an integral part of the management of database systems?	Qualitative	Focus Groups
How can a better understanding of the complex interactions contribute to improvement and innovation?	Quantitative and Qualitative	Survey and Focus Groups

### 3.5 Quantitative Design

The quantitative research is the empirical investigation of the database community usage of best practices using statistical techniques. It aimed to answer the first research question (see Table 3.2).

The quantitative research phase was carried out through the use of a web based questionnaire which produced pre-coded numerical output. Quantitative analysis has interrelated variables and characteristics or attributes that can be measured. The survey allowed data to be collected from the population to discover facts and additionally gain evidence about some behaviours and attitudes.

A web based survey was used in the quantitative method to gather practitioner and organizational demographics within the boundary of study, identifying their knowledge and current use of management systems in database management. The survey method was chosen as a way of collecting a baseline of information from a large population across the world. This method allowed the researcher to be partially external to the research. The survey provided details about current practices and procedures adopted and provided numerical data. This allowed the database population characteristics to be identified and provided detailed evidence of the problem for further investigation.

Other quantitative methods such as visual observation and telephone surveys were considered. Visual observations would have put the researcher within the field and this could have caused bias in the research study as the researcher was already submerged in the field. Visual observation was thought to be an unsuitable method for bulk rapid collection of data. Telephone surveys were not practical due to the volume of data required to be collected at this stage.

Detailed thought was given to determine the best method with which to collect data from the respondents. An appropriate method within this cultural area, where respondents were likely to respond and which would work best for worldwide distribution, was a web based survey. The limited resources available, such as time and costs, also meant a web based survey was a suitable instrument and the most feasible choice. The survey needed to be unobtrusive to the participants to encourage responses. The data collection needed to be objective and verifiable, and thus a survey was trialled to determine if there were any issues with the questionnaire design.

### **3.5.1 Quantitative Sampling**

This research was based on non-probability convenience sampling (Bradley 1999; Buckingham & Saunders 2009; Denscombe 2008). Convenience sampling is also known as opportunity, accidental or haphazard sampling as defined by the National Audit Office statistical and technical team (1998, p.11). It defines the method as

*“using those who are willing to volunteer, or cases which are presented to you as a sample”.*

The overall picture gained from non-probability convenience sampling benefits from a larger sample. It is commonly used during preliminary research to gain a summary of interesting information. It allows the data to be collected quickly and inexpensively.

This was chosen because the entire database community was the target population and the sampling frame. It included all the database management and data professionals who manage databases worldwide, potentially including any type of establishment using a database. It was therefore impossible to know the size and dispersion of the population. In addition the practical limits such as cost and

geographical dispersion of the population mean that probability sampling, which may produce a truly representative result, would be extremely difficult. Practical limitations are such that although the results cannot be extrapolated to give statistical population results, they can form a valid and defensible methodology for this research.

The sample was obtained through advertising the survey via social media such as Twitter, LinkedIn and Facebook groups, email newsletters and blog posts. A strategic decision was taken to gather data as widely as possible across the database population, so as to include a range of geographic locations, job roles and database software. For convenience sampling, larger sample sizes are better, in order to get an overall picture. The primary disadvantage of this sampling method is that there is no guarantee of a representative sample; hence it was not possible to make generalizations of the entire population from the results.

### **3.5.2 Triangulation for the Quantitative Phase**

Triangulation is a technique used in social sciences to verify two or more sources and to enhance the confidence in research results. The literature of Jick (1979, p.609), Denzin (1978, p.291) and Hammersley (2008) on triangulation was studied and taken into account.

There are various types of triangulation: data, investigator, theoretical, methodological, analysis and environment (Thurmond 2001; Guion 2002; Jick 1979). This research was based on a mixed methods approach which requires the use of complementary quantitative and qualitative methods. Methodological triangulation combines 'within method triangulation' and 'between or across method triangulation'; this approach is widely used in social sciences (Hussein 2009).

Methodological 'between or across method' triangulation is discussed later in the connecting stages.

The other type of methodological triangulation was known as 'within method triangulation'. The 'within method' triangulation was used in this quantitative research to check the internal consistency and reliability of the survey (Denzin 1978). Patton (1999) argued that triangulation tests for consistency where inconsistencies were found. The results offer deeper insights and should not be viewed as making the results less credible.

The survey focused on the same construct to ask control questions to validate the questionnaire questions where the same people were asked different questions on the same area. This provided limited validation.

Analysis or data analysis triangulation uses two or more methods of analysing the same data for validation (Hussein 2009). The 'within method' approach generally implies two procedures (Thurmond 2001; Hussein 2009; Bekhet & Zauszniewski 2012), for example a survey questionnaire and an existing database. This research reviewed another set of other surveys undertaken alongside this survey with a few questions from their various survey findings (Mckendrick 2013; Mckendrick 2011b; Mckendrick 2011a) to check consistency.

The other types of triangulation, theoretical, investigator and environmental were considered but seemed inappropriate to this research. Theoretical triangulation uses multiple theories to analyse the data to assess viewpoints. It was not used as it may not increase the validity and credibility of the findings (Thurmond 2001). As the research has only one researcher, investigator triangulation was ruled.

Environmental triangulation (Guion 2002) uses different locations, times or setting to see if environmental factors influence the information. Environmental triangulation is

only to be used if there is a likelihood that the results are influenced by environment factors.

### **3.5.3 Quantitative Data Collection Method: Survey**

Morris (2008, pp.458–459) defined various stages of the quantitative approach. The initial stages of that approach have been mapped onto this research including the survey design method developed by Buckingham & Saunders (2009). The survey instrument used was Survey Monkey, and the purpose of the survey research was to sample the database population to discover facts about the population (descriptive research), highlight characteristics, attitudes, behaviour (analytical, explanatory research). A pilot study was run to verify that the questionnaire produced the intended results, to make sure the questions were clear and checked that the time taken to complete it was acceptable.

The survey was cross sectional, with the survey data collected at one point in time with the survey being open for data collection for a set period. The form of the quantitative data collection involved an open web-based internet (Bradley 1999), and was administered online. The length of the questionnaire needed to be carefully considered. The use of a survey to provide reliable results was a consistent way to collect the observations. The design of the questions was carefully worded and specific to improve the reliability of the results. Denscombe (2008) discussed good practice for questionnaires and this was followed. The questionnaire must be easy to complete, the design must incorporate key factors relating to the research, only relevant questions asked and there should be no duplication of questions. The questions should not lead the respondent to the answer. Any technical jargon should be avoided and explained if had to be used.

In the closed ended questions a pre-coding of data for themes and variables helped in the analysis of the data. As Buckingham & Saunders (2009) suggested, having several variables to measure the same concept helped increase reliability of the measurements.

The ethical stance of this research was to protect the confidentiality and integrity of the participants. The research proposal was assessed by the OU research committee and considered not to raise any ethical issues. The raw data will not be disclosed to unconnected parties and will be protected under the Data Protection Act. A data management plan was formulated.

The goal of the research was to gain insight into how databases were actually administered and to identify what practices and procedures were utilized throughout the database lifecycle. Further aims were to understand the demographics of people who manage database systems, to investigate how they learned about best practice and whether any IT frameworks were used. As the database community was dispersed globally, the survey sought to reflect this global nature.

The survey design method included the following stages:

1. Defined the goals of the research project
2. Determined the sampling strategy and survey marketing
3. Chose the survey instrument, Survey Monkey ([www.surveymonkey.com](http://www.surveymonkey.com))
4. Created a questionnaire and applied for Ethics approval with a data management plan
5. Conducted a pilot test of the questionnaire to test the questions
6. Reflected and revised questionnaire questions
7. Opened the survey for data collection and started marketing
8. Analysed the data collected



9. Produced the report

The questionnaire comprised 83 questions of different types, including dichotomous, nominal (multichotomous), interval level (Likert scale) and multi-option responses, see below. Several free text questions were also included to allow the respondents to explain items in more depth.

The content of the questions address the first research question: “To what extent are best practices and procedures utilised by the database community?” They were chosen in order to take the respondents through all the different stages of database management, and to consider frameworks, supplementary technology areas and organizational culture. Also some questions took a different viewpoint from that of database management, moving to an application-centric viewpoint, to try to elucidate more information on best practice and thoughts about the future in this context. The sections of the survey were:

- Demographics
- Server Demographics
- Database architecture, design and development
- Database technical practices
- Data and database security
- Change management
- Data management
- Frameworks
- Storage
- Cloud
- Organizational culture
- Application centric

- Best practice
- Future: Cloud and Business vision

The different types of question used:

- Dichotomous questions for Yes/No answers
- Multichotomous where there was a choice of several answers
- Nominal questions that could be used to calculate central tendency mode and dispersion.
- Check-box or multi option variable
- Likert Scale questions (Table 3.3) requiring fixed choice responses to questions. These questions were for measuring agreement, frequency or attitudes using ordinal scales.

**Table 3.3 Likert Scale**

Degree of agreement	1	2	3	4	5
Five-point scale	Always	Often	Sometimes	Rarely	Never
Three-point scale	Always		Sometimes		Never

The survey included questions to collect demographics of the sample, for example industry sector, job role, organization size and training opportunities. There were three different types of survey questions used in the survey to obtain different results, see Table 3.4.

**Table 3.4 Survey question types**

Types	Question
Facts, Demographics	What is? Are you? Where do you?
Behaviour	Do you use?
Opinions, views	What do you think?

A pilot survey was undertaken to review the developed questions to get timings for completion and ensure no ambiguity was contained within the questions. After the pilot a few questions were removed due to the questions being interpreted differently than anticipated and the text of a few questions was changed for clarity. The secondary aim was to test validity, reliability and acceptability to analyse the pilot data results to make sure a range of clear data was being produced. Following revision of the pilot survey ethics of data protection and storage management of data were considered.

Consideration was given to the problem that inconsistency of data might occur within the survey. This was addressed by adding a few questions to the survey which would enable triangulation of some of the data. Two examples of questions that expressed similar intentions were: Q7 and one of the sub-questions of Q66 (about vendor certifications); and Q38 and Q40 (about managing change).

The survey marketing plan was followed once the survey was launched and the survey was open for data collection between 13 December 2012 and 6 February 2013. A total of 453 respondents (n = 453) participated from within the global community of database and data professionals.

#### **3.5.4 Quantitative Data Analysis**

The method for quantitative analysis in general was defined in Creswell (2009, pp.151–152) with more in-depth steps identified for the rigorous quantitative data analysis in Creswell & Plano Clark (2011b, p.205). These included preparing the data for analysis, exploring the data, analysing the data, representing the data analysis, interpreting the results and validating the data and results.

Statistical tests are a method of hypothesis testing, but the purpose of the quantitative survey has been to reveal the areas likely to benefit from further

knowledge of the practices and procedures used in operating databases and whether the complexity of database systems was such that the quantitative survey data alone was insufficient to draw conclusions without further investigation.

Statistical tests on the quantitative data results could produce a false positive or Type 1 error, in which a relationship existed when in fact it did not. In a similar way, a false negative, or Type 2 error, could be produced in which no relationship existed when in fact it did.

The use of descriptive statistics provided a simple summary to describe patterns and trends in the data set. The analysis consisted of finding the central tendency using calculations for the total number of people using certain methods. In some cases the distribution frequency or dispersion of the spread of values was considered. The raw data was organised and the frequency of category choices selected by a respondent was determined.

Cross tabulation, sometimes called contingency table analysis, is a quantitative method that looks at relationships between more than one variable. For survey questions this can show how the questions are interrelated. Cross tabulation provided a way of delving into the data to gain further understanding. The results are often displayed as counts showing the frequency and are the joint distribution of two or more variables. This was useful for analyzing nominal data. Cross tabulations were used against some of the survey data which were likely to reveal useful information.

Analysis by descriptive statistics and some cross tabulated data was useful for showing a side by side comparison of two or more survey questions. The tools used in the analysis were SQL Server, Excel, Power Query and SPSS.

### 3.6 Connecting Quantitative and Qualitative Phases

The explanatory research design was sequential, with the qualitative strand following on from the quantitative phase. The interface between the two phases built upon the quantitative results in which specific results were highlighted that would benefit from further explanation.

Qualitative questions were then developed and refined. The questions asked for opinions, thoughts, ideas, knowledge and insight. The questions aimed to guide the participants through the research interviews. These questions were open ended and semi structured. The semi structured nature of the questions could lead to the participants raising issues not previously thought about. The questions began with an introductory question, then other questions in sequence with clustered topics leading to a concluding question.

Creswell & Plano Clark (2011b, p.181) discussed various decisions that were needed when connecting the two parts of the research together. These decisions included:

- Trying to ensure the individuals who took part in the first quantitative data collection participate in the qualitative phase
- Noting that the sample size will be smaller in the qualitative phase
- Carefully considering the design of the qualitative questions that follow
- Considering what questions should be asked
- Follow up participants could be selected based on initial quantitative results.
- Making sure an addendum is submitted to the ethics review board based on the second data collection phase.

Venkatesh *et al.* (2013) argued that quantitative research in information systems recognised reliability and validity importance.

Both quantitative and qualitative research have their own methods to deal with reliability and validity. Teddlie & Tashakkori (2003; 2009) argued validity was not clear and used the term 'inference quality' and reliability as 'data quality'.

Venkatesh *et al.* (2013, p.35) used the terms inference quality and data quality to help clarify mixed method validation:

- Inference quality is the validity of accuracy of derived conclusion or interpretation made in the study
- Data quality is the reliability in mixed methods, quality of the measures or observations in the data collection validity (trustworthiness) and reliability (repeatability)

These validation principles and those of Creswell & Plano Clark (2011b, p.181) formed part of the rigorous strategy in this thesis. The framework in Venkatesh *et al.* (2013, p.44) covered aspects of inference quality, design and explanation quality. The findings from the quantitative and qualitative analysis should be effectively integrated.

The data analysis strategy for sequential mixed methods may employ meta-inferences. Venkatesh *et al.* (2013, p.38) defined meta-inferences as:

*“Theoretical statements, narratives, or story inferred from an integration of findings from quantitative and qualitative strands of mixed methods research.”*

Venkatesh *et al.* (2013, p.35) argued that drawing meta-inferences is critical in mixed methods research. Two approaches have been suggested to develop meta inferences: bracketing and bridging (Lewis & Grimes 1999). The aim of bracketing is to theorize between contradictions and oppositions captured by the findings to help identify the nature of their sources, incorporating diverse views; this approach is suited to concurrent mixed methods. The alternative approach is bridging, the development of consensus between quantitative and qualitative results, which is suited to sequential mixed method research. Venkatesh *et al.* (2013, p.39) stated that using the use of bridging can help develop a “theoretically plausible integrative understanding”.

Because the research methodology is one of sequential mixed methods, bridging was used rather than bracketing in this research. The research in this thesis built on this technique to draw together both quantitative and qualitative stages to provide a holistic explanation. The output of the quantitative analysis was used directly to inform the design of the questions for the qualitative data collection, as can be seen in section 4.3 below. This created an expanded view of the management of database systems as Chapter 5 shows.

### **3.6.1 Triangulation**

The second stage of triangulation used was ‘between or across method’.

Methodological triangulation was the use of both quantitative and qualitative data collection in the same study (Thurmond 2001). A survey questionnaire with follow up participant focus groups was combined to elucidate the same research problem. This is expanded upon later in this chapter. Inconsistencies and areas of agreements could be compared with both the survey data and focus groups.

## 3.7 Qualitative Design

The qualitative phase identified the complex system involved, reporting multiple perspectives and a multitude of connected factors. This approach enabled the stories that combined with the quantitative data ensured a better understanding of the problem. It was possible to find patterns in qualitative data. The data could demonstrate real world issues and show complexity and “mess” (Braun & Clarke 2013, p.10). The qualitative research aimed to answer the second and third research questions. It enabled an in-depth understanding of the participants’ views and their perspectives.

To deal with the practicalities of when, where and how to collect qualitative data, focus groups were proposed to collect information at user group meetings or database conferences. This allowed interesting results from the initial survey to be explored. The reason for the qualitative research phase was to seek an in-depth understanding of the viewpoints of the practitioners, to further explain the initial findings. The focus groups were semi-structured with question topics, but the following discussion aimed to be free flowing.

The research was limited by the experiences of the respondents to the initial survey, and the participants of the focus groups. The participants’ and respondents’ anonymity was maintained so their identity could only be revealed with their approval, which otherwise might have limited some disclosures.

### 3.7.1 Qualitative Sampling

The sampling type used for the qualitative data collection was purposive sampling, a nonprobability method explained in Table 3.5.

**Table 3.5 Qualitative sampling selection used**



Method	Sampling Type	Sample Method	Collection Method
Qualitative	Non-probability	Purposive. Samples chosen using Maximum Variation Sampling (n=29).	Focus groups types Face to face Asynchronous email Asynchronous forum

Table 3.5 indicates the type of purposive sampling used, showing details of the number of participants and the collection methods.

Purposive sampling, also known as judgmental, selective or subjective sampling, is the collection of information with a purpose in mind, the sample to include the people of interest. The sample to be studied focused on a small number of cases. Simon (2008) stated the sample is deliberately selected to achieve a goal and was non-random. Participants with a range of knowledge and experience (e.g. years of experience, 'junior' and 'senior' and a range of specialisms) in the area were selected to help gain a through understanding of the field.

Patton (1990b, p.169) stated that selecting information rich cases could provide illumination of issues of central importance. The benefit of purposive sampling was that there were many sampling techniques to choose from. The maximum variation sampling method selected aimed to select participants that were spread across the database system to capture and describe themes or principal outcomes

(Patton 1990a, p.172). Patton stated that the analysis has two findings, detailed descriptions for documenting uniqueness and shared patterns across the cases that emerged from their heterogeneity.

Participants in this research were selected purposefully for the qualitative sample. The rationale used for selection was based on the strategies of Simon (2008) and

Patton (1990a .p 172, 1990b p.169). The selection was by job role, the field, the length of time worked in the field and the size of organization. This was similar to the same variation as the quantitative survey. The strategy for selecting participants was to include people with different experiences *yet also* experienced in the field. This was to look for patterns within the variation and to capture the central themes across the variations. To maximise the difference, the factors used were diversity amongst database roles, small and large organizations and country of operation. This provided a complex picture of the phenomenon.

### **3.7.2 Qualitative Triangulation**

This research has looked at using triangulation against an existing data source and between quantitative and qualitative 'between-across methods'. Braun & Clarke (2013, p.286) argued triangulation could be problematic in qualitative research and there may not be a single truth, and Guest *et al.* (2012, p.202) argued the term was over used and so broad it was virtually meaningless especially as in many cases only two data points were used. Triangulation is a mathematical process that requires three data points. The qualitative stage was the third data control point in this research.

A third type of triangulation was sought to add diversity against the data checks. This check was that of data analysis triangulation. That was the combination of two methods to analyse the same data set (Thurmond 2001; Hussein 2009). This added validation and completeness to the research process. This is the use of analytical research methods and a synthesis systems thinking approach explained later.

### **3.7.3 Qualitative Data Collection Method**

The qualitative data collection method required that the results could be obtained with little cost and required a method that would "*encourage a range of responses*"

(Liamputtong 2011, p.3) . It was also advantageous to be able to collect the data in the minimum time due to the fact that many database professionals have very little free time and if database conferences were to be used as a place to advertise and hold the meetings, these conferences have a limited duration.

Database personnel were always under pressure from work to keep database services available for users, often 24/7, every day, and they were required to attend to unpredictable incidents that could cause downtime. Contacting people from abroad was otherwise impracticable, so to reduce geographic barriers and reduce cost, a variety of structured focus groups was necessary.

A focus group is made up of a group of people who come together to discuss a particular topic

*“The primary aim of a focus group is to describe and understand meanings and interpretations of a select group of people to gain an understanding of a specific issue from the perspective of the participants of the group” (Liamputtong 2009)*

Focus groups are an increasingly popular data collection method, and can collect data from multiple participants at the same time. Also the strengths of the focus group was that it was fast, less expensive and provided social interaction with group dynamics and was able to elicit wide ranging views and perspectives (Braun & Clarke 2013, pp.107–110). It provided insights and not rules; it was flexible and not standardised. Focus group interviews also provided insight into the social structure of organizations. Workshops have been used successfully for over 20 years to discuss and debate database systems problems and the future of database research (Abadi *et al.* 2016; Agrawal *et al.* 2009; Abiteboul *et al.* 2005; Bernstein *et al.* 1989; Bernstein *et al.* 1998; Silberschatz *et al.* 1995). Due to these practicalities

the focus group was chosen as the quantitative data collection method for the research.

There were some weaknesses with focus groups such as less control, no generalisation, and that they could be affected by the most dominant person. Also, asynchronous groups might lack spontaneity with no real-time group interaction. Asynchronous groups enable participants to contribute in their own time, a result of which is that responses may be more considered.

The focus group method used was a single category design which dictated that several focus groups should be held until theoretical saturation was reached. Saturation occurs when the range of ideas was not providing any additional insights. It was suggested (Krueger & Casey 2009, p.26; Morgan & Krueger 1997, p.44) that three to four focus groups would be sufficient. Braun & Clarke (2013, p.115) have found small focus groups (3-8) work best for generating rich discussion.

A mixture of groups were used, some face to face, some using online forum groups which resulted in asynchronous responses and some conducted remotely by e-mail. Sufficient responses for saturation were obtained using this collection method. The remote email group involved the use of a group email being sent to all participants who were asked to reply to all.

The recruitment of participants was through members of database and information management groups, on Twitter and invitations sent to people of similar job roles in the field of database management both in the local area, UK & Ireland and to international attendees of database conferences.

How to conduct on-line forum groups remotely by e-mail was discussed in Denscombe (2008, p.186). Participants were not required all to be available at the

same time. This also allowed for more reflective responses (Oringderff 2004, p.152). The ethical consideration where ideas were shared in written form were set out in the additional guidance rules.

Face to face focus group full transcripts were made from recordings and notes. Where a problem occurred in the session recording, the respondents were asked to clarify their discussion comments via email. The asynchronous forum and email group transcripts were used as typed by the participants.

There was an initial focus group to test the suitability of the questions through an online e-mail, virtual asynchronous group. The check involved receiving feedback from the participants about the comprehension and whether the order of the questions could be improved. The trial did not identify any issues with the ten questions. The only issue was the delayed responses of some of the participants which resulted in there being minimal interaction. This was avoided in subsequent groups by personally speaking to the participants in advance to request minimum time gaps in replying and thus momentum was maintained. The questions are listed in Appendix B.

The demographics of the respondents in each group were collected: the job area, length of time in the field, organization type and country.

The focus group demographics recorded were as shown in Table 3.6.

**Table 3.6 Focus group demographics**

<b>Focus Group Dimensions</b>			
<b>Job Area</b>	<b>Time in Field</b>	<b>Organization Type</b>	<b>Country</b>
What is your job area? <ul style="list-style-type: none"> <li>• DBA</li> <li>• Developer</li> </ul>	How long have you worked in the database field?	What size of organization do you work in?	In which country do you work? <ul style="list-style-type: none"> <li>• UK</li> <li>• Non UK</li> </ul>

<ul style="list-style-type: none"> <li>• Business Intelligence (BI)</li> <li>• Architect</li> <li>• Other (Accidental DBA)</li> </ul>	<ul style="list-style-type: none"> <li>• Less than 5 years</li> <li>• 5 to 10 years</li> <li>• More than 10 years</li> </ul>	<ul style="list-style-type: none"> <li>• Self Employed Consultant</li> <li>• Small-to-medium enterprise (SME)</li> <li>• Large Enterprise</li> </ul>	
---	--	--	--

Some of the people that attended the focus groups were from a select group of attendees at conferences. The events were a monthly local data group event, a PASS SQL Saturday event <http://sqlsaturday.com/> and SQL Rally <http://www.pass.org/sqrrally/2013/amsterdam/Home.aspx>, which is an annual event. The respondents were practitioners, some of whom were database managers, rather than users or other stakeholders. They were therefore people who were interested in increasing their skills, receiving help and advice with new developments of database software systems, and networking with other data professionals. The implications could be that these people are specifically interested in databases, and may already be quite skilled. The people from these different sources may have had different perspectives from each other because they were different types of practitioners, in different business spheres or because they were from different countries.

Each respondent was issued with a set of ground rules (Eliot & Associates 2005), with each remote participant given additional guidance with written responses. The number of focus groups produced sufficient information for saturation.

### 3.7.4 Qualitative Data Analysis

There are various different strategies that can be used for data analysis and interpretation of qualitative data. These strategies help gain an understanding of the data and use various textual processes for data analysis. The data analysis process recommended by Cresswell (2009, p.185) and Cresswell & Plano Clark (2011b,

p.205) were observed when preparing the data for analysis, interpreting the results at multiple levels and validating the data and results.

The textual process for data analysis chosen was Thematic Analysis for finding patterns within the data. Specific steps were recommended to analyse the data which included the coding of data and themes. Coding of data is a way of analysing qualitative data:

“Coding is heuristic (from the Greek , meaning” to discover”) – an exploratory problem-solving technique without specific formulas or algorithms to follow” (Saldana 2013, p.8).

Considerable interpretation goes into developing the codes and rigorous strategies needed to be followed to capture the complex meanings of the data set (Guest *et al.* 2012, pp.10–12). Reflection on the entire process was key, to ensure focus was maintained on the question in hand, and to ensure the researcher’s working experience did not cloud the researcher’s judgement of the data received.

### 3.8 Thematic Analysis

Thematic Analysis is a common method used for descriptive analysis and reports patterns (themes) within the data. Thematic analysis is

*“a form of analysis which had the theme as its unit of analysis and which looks across data from many different sources to identify themes.”*

(Braun & Clarke 2013, p.337).

An advantage of thematic analysis is its flexible approach (Braun & Clarke 2006, p.76,81). It is an analytical method (Clarke & Braun 2013, p.120). It can cover an array of research interests and questions, analyse different data types, such as

focus groups, work with small or large data sets and produce data driven analysis.

The disadvantages are around continuity with data from different respondents, how the analysis meshes together and how to ensure flexibility could be maintained.

The steps followed in this method were as shown in Table 3.7.

**Table 3.7 Phases of thematic analysis  
from Braun & Clarke (2006)**

<b>Phases</b>	<b>Description of the process</b>
1. Familiarising yourself with your data:	Transcribing data (if necessary), reading and re-reading the data, noting down initial ideas.
2. Generating initial codes:	Coding interesting features of the data in a systematic fashion across the entire data set, collating data relevant to each code.
3. Searching for themes:	Collating codes into potential themes, gathering all data relevant to each potential theme.
4. Reviewing themes:	Checking the themes work in relation to the coded extracts (Level 1) and the entire data set (Level 2), generating a thematic 'map' of the analysis.
5. Defining and naming themes:	Ongoing analysis to refine the specifics of each theme, and the overall story the analysis tells; generating clear definitions and names for each theme.
6. Producing the report:	The final opportunity for analysis. Selection of vivid, compelling extract examples, final analysis of selected extracts, relating back of the analysis to the research question and literature, producing a scholarly report of the analysis.

The steps undertaken in this research relating to each section are described in detail in Section 3.9.1.

Thematic Analysis also could straddle various research approaches and perspectives (Braun & Clarke 2012, p.58). The analysis used a combination of approaches. Whereas the quantitative research used a deductive approach to determine causality with a top down approach the qualitative research explored the



phenomena, an inductive approach, from the bottom up where the themes and codes were derived from the actual data (Braun & Clarke 2012, p.58).

Thematic analysis requires an understanding of codes and themes. These definitions are explained in the following two sections.

### 3.8.1 Codes Defined

Codes were described by authors in a variety of ways

*“Codes identify a feature of the data (semantic content or latent) that appears interesting to the analyst, and refer to ‘the most basic segment, or element, of the raw data or information that can be assessed in a meaningful way regarding the phenomenon’”* (Braun & Clarke 2006, p.88)

Tucket (2015, p.82) referred to code assignment as a ‘tag’ or ‘label’.

A code is a short word or phrase to describe part of the data. Saldana (2013, p.262) defines it as

*“most often a researcher generated word or short phrase that symbolically assigns a summative, salient, essence- capturing and /or evocative attribute for a portion of language- based or visual data.”*

Carpenter and Suto (2008, p.116) describe codes as

*"Shorthand labels - usually a word, short phrase, or metaphor - often derived from the participants' accounts, which are assigned to data fragments defined as having some common meaning or relationship."*

This research used the word code to mean a label made up of a short word or pair of words to describe part of the data of interest for the research.

An example of the coding style used in this research is shown in Table 3.8.

**Table 3.8 Descriptive code summarizing the primary topic  
(Saldana 2013, p.4)**

Data Excerpt	Descriptive Codes
I notice that the grand majority of homes have chain link fences in front of them. There are many dogs (mostly German shepherds) with signs on fences that say "Beware of the Dog"	SECURITY

### 3.8.2 Themes Defined

Themes have been defined in various literature sources. Themes are the units of analysis (Braun & Clarke 2006, p.88) which are often broader than a code with many facets (Braun & Clarke 2013, p.224).

*"A theme captures something important about the data in relation to the research question and represents some level of patterned response or meaning within the data set".* (Braun & Clarke 2006, p.82)

A theme is a phrase or sentence describing more subtle and tacit processes (Saldana 2013, p.14) and is an outcome of coding, categorization and analytic reflection (Saldana 2013, p.175). A theme functions as a way to categorise a set of data into

*"an implicit topic that organizes a group of repeating ideas"* (Auerbach & Silverstein 2003, p.38)

In summary an example of a code and theme

*“SECURITY can be a code, but DENIAL MEANS A FALSE SENSE OF SECURITY can be a theme” (Saldana 2013, p.14)*

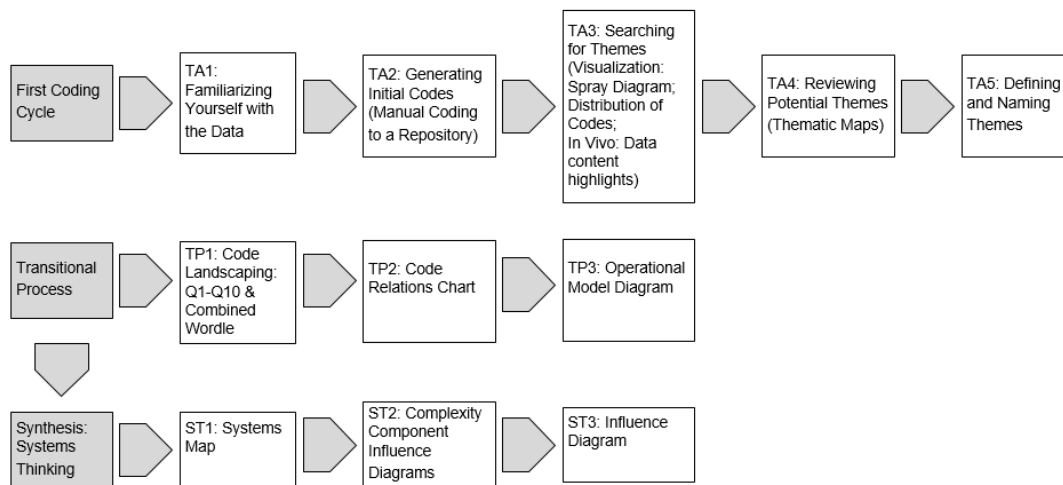
### 3.9 Defined Data Analysis Process and Method used

Although thematic analysis helps identify themes across the data, the research problem implied that there are complexities that exist that would benefit from further investigation. Saldana (2013, p.59) presented two cycles of coding - a first cycle of coding, post coding transaction and second cycle of coding method. The methods used in this research in each part of the cycle of coding are listed in Table 3.9.

**Table 3.9 Adapted coding strategies and methods based on Saldana (2013, p.59)**

<b>First Cycle Coding Methods</b>	<b>Post Coding Transition</b>	<b>Second Cycle of Coding Methods</b>
Theming the Data Thematic Analysis Thematic Maps	Code Landscaping Wordle	Divergence to Synthesis Systems Thinking
Utilising additional tools:	Code Relations Chart	
Spray Diagrams	Operational Model	
Distribution of Codes		
Elemental Methods In Vivo Coding		

The method in Table 3.9 was incorporated into a formal process. The qualitative research process used is shown in Figure 3.3.



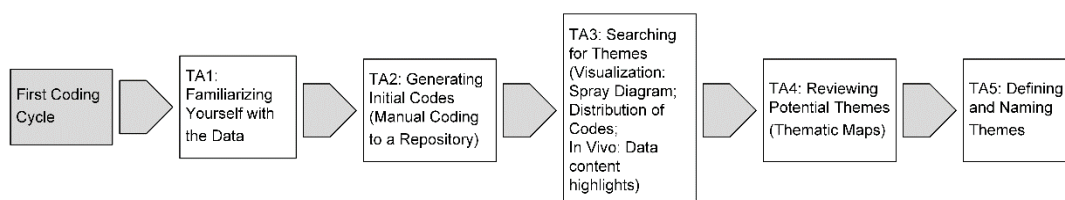
**Figure 3.3 Qualitative research process**

**Key: TA = Thematic Analysis, TP = Transitional Process, ST = Systems Thinking**

The three stages of the research are described in Figure 3.4, starting with the first coding cycle (5 stages), then the transitional process (3 stages) finishing with synthesis based on systems thinking (3 stages).

### 3.9.1 First Coding Cycle

This first cycle of coding, Figure 3.4, incorporated all the initial methods used in the initial coding of the data (Saldana 2013, p.58) . Thematic analysis (Braun & Clarke 2006) was the primary method used in this cycle although two elemental methods, Structural and In Vivo were also explored to understand the data.



**Figure 3.4 First coding cycle**

This is an inductive way of coding and theme development which were directed by the content of the data;

### **TA1 Familiarizing Yourself with the Data**

Initial engagement with the data corpus started with working with the verbal transcripts and transcribing them into written texts ensuring a verbatim account was documented. It was important that the transcripts of the audio recordings were accurately transcribed. During this phase some general notes were made about ideas for coding and general groups constructed.

### **TA2 Generating Initial Codes**

This phase started with creating initial codes from the data. All the data corpus was systematically coded. This included creating codes for data which seemed interesting. The coding might have been data driven or theory driven. In this stage initial themes began to emerge. Potential patterns were highlighted by the repeating of certain themes. Braun & Clarke raised some key points for this stage:

*“a) code for as many potential themes/patterns as possible (time permitting) – you never know what might be interesting later; b) code extracts of data inclusively – i.e., keep a little of the surrounding data if relevant, a common criticism of coding is that the context is lost (Bryman, 2001); and c) remember that you can code individual extracts of data in as many different “themes” as they fit into - so an extract may be uncoded, coded once, or coded many times, as relevant.” (2006, p.89)*

Braun & Clarke (2006) stated that all datasets have contradictory data. Developing a thematic map to conceptualise the relationships and data patterns could still show

data item inconsistency. These inconsistencies were still important in the codes. A code book, was created to start documenting the codes, themes and data mappings.

### **TA3 Searching for Themes**

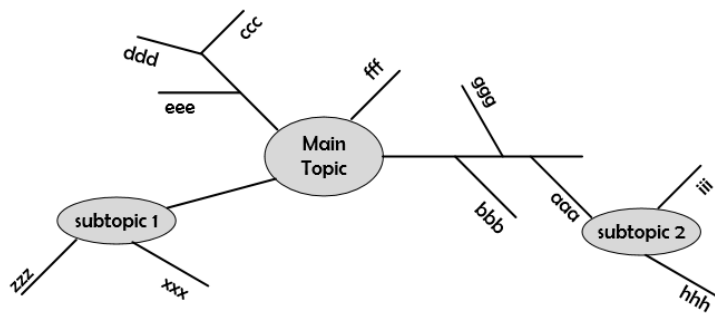
This stage within the analysis looked at potential themes. In the process of thinking about the codes, relationships and hierarchy between the codes a few different research tools were used. Visual representation of the data codes and themes such as thematic maps and mind-maps were used to help with the creation of the overarching themes and understanding their relationships. The start of creating thematic maps began in this stage. Three other processes were used.

#### **Spray diagrams**

Spray diagrams were used to help understand the relationships between the codes and to provide a conceptual structural map. These were used in preference to mind maps which were generally used for brainstorming ideas in an unstructured way.

Spray diagrams were developed by Tony Buzan (1974):

*“Spray diagrams show the connections between related elements or concepts associated with a particular issue. They do not show the nature of the relationship between the elements. A spray diagram can be thought of as a conceptual map of a situation or issue.” (Reynolds et al. 2014a)*



**Figure 3.5 Format for a spray diagram  
(The Open University 2012)**

Spray diagrams (Figure 3.5) were used to help gain an initial understanding during the codification of the data. The letters 'aaa' or 'zzz' depict labels at the ends of line which are elements / components or concepts connected to the main topic. Spray diagram are often used in the early stages of system thinking analysis.

### **Distribution of codes**

The next process used was distribution of codes. Looking at code frequencies was useful for analysis purpose and could help as a data reduction technique. The codes were effectively metadata, removed from the text. Consultation of the actual data was required for interpretation – as (Guest *et al.* 2012, pp.138–139) argued, this was a useful technique for validity to ensure maximum reliability, and added rigour to the research process.

It was intended that the data corpus for each of the ten qualitative questions be included in the code frequency count. The scope of the code count was to compare the codes in each of the 10 questions. The objectives of the analysis was to find if any patterns existed across the data set through the use of a frequency report (Namey *et al.* 2007, pp.141–144; Ryan & Bernard 2000, p.776). This tool helped with rapid scanning for topics, to enable exploration of the patterns in the data.

These were consolidated into the code book where codes appear multiple times. The frequency report also helped with the redefinition of codes that rarely occurred.

### **In Vivo Coding**

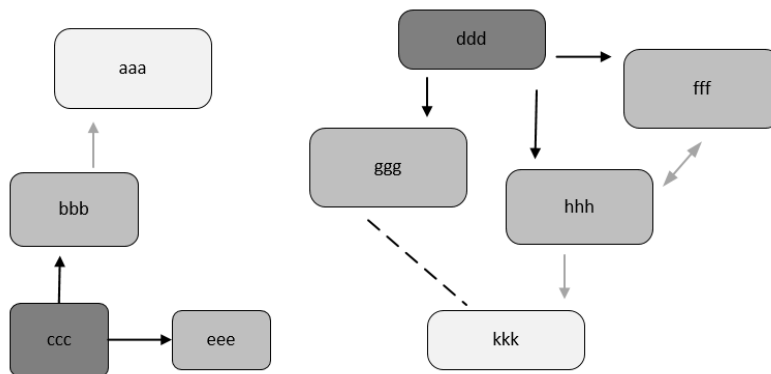
The third process within this stage was a high level review using In Vivo coding. This is also known as “literal coding or inductive coding”. In vivo is a word or short phrase determined by the participant’s content. Strauss referred to “the terms used by participants themselves” (1987, p.33).

This was a process driven from the data whereby specific broad data quotes from each of the 10 questions were collected. It was based on using the participants’ own language to produce data driven codes. These codes were considered remarkable comments and these key text passages were used in the creation of the codes. Some words and phrases might be considered significant (Charmaz 2006), at this early stage. This was an alternative way of coding and was used to give a high level view of the participants’ own words. The codes created were the unit of analysis in this section and helped to provide further findings in the data.

### **TA4 Reviewing Potential Themes**

During the process of reviewing potential themes, Thematic Maps were constructed. These are tools used by researchers to visualise the identifiable themes, subthemes and provide a rough idea of the relationships of the themes. The creation of Thematic Maps was not a formal way of presenting an overview of an analysis and the process was not prescriptive. The thematic map provided a conceptual view of the data patterns and relationships. The relationships could be hierarchical based on the codes. Using the visual thematic map tool was essential to the analysis, to have an alternative view of the potential themes and sub themes.





**Figure 3.6 Candidate overarching themes, themes and subthemes from (Braun & Clarke 2013, p.233)**

*“Key: single directional solid arrows demonstrate hierarchical relations between overarching themes, themes and subthemes; a bi-directional solid arrow signals a close lateral relationship between themes; a dotted line indicates a tentative relationships between a theme and a subtheme of a different theme.” (Braun & Clarke 2013, p.233) “The colour simply corresponds to the ‘level’ – so the black arrows are relationships from overarching themes to themes, and the light grey are from themes to subthemes.” – Email clarification 27-06-2016.*

Reviewing candidate themes required iteration and continuous reference back to the coded data, data corpus and data set. As a result some themes were changed, some merged and some removed entirely.

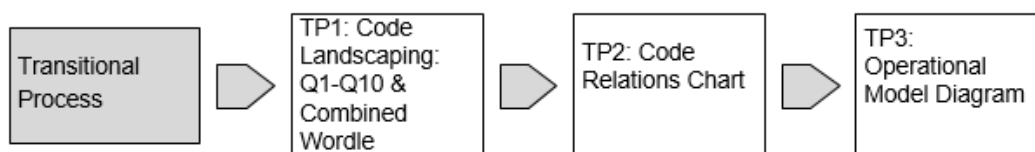
### **TA5 Defining and Naming Themes**

In this stage the themes were defined and refined to understand their essence. This followed on from the completion of the thematic maps. The codes and themes might be renamed in the process to ensure the narrative was clear. Also another revision of the thematic maps might be required. It was useful to write theme definitions to focus the boundaries. Each theme was discrete, rich, and coherent, addressing the

research questions and naming the themes should capture the essence of the theme and analytic 'take' on the data (Braun & Clarke 2013, pp.249 & 258).

### 3.9.2 Transitional Process

This section looks at the post coding transition, Figure 3.7, by further reanalysis of the combined data set to provide a clearer focus for the direction of study through reflection (Saldana 2013, p.187).



**Figure 3.7 Transitional process**

The final stage thematic analysis in Braun and Clarke's approach is the writing of a report. In this research the writing of the report was undertaken at the end of the synthesis section. This analysis used a few tools in the seamless transition from analytic research to synthesis, applying a systems thinking approach to the situation. Saldana stated

*"The goal is not to "take you to the next level," but to cycle back to your first coding efforts so you can strategically cycle forward to additional coding and qualitative data analytics methods." (2013, p.187)*

This section was an iteration through the data corpus to affirm classification of codes, create preliminary models, to move forward and reassemble the data, and to transform the focus of the study (Saldana 2013, p.187). The tools used in this stage were adapted to the particular research undertaken.

### **TP1: Code Landscaping**

Code landscaping was a data management technique defined as

*“a method of manually organizing codes, subcodes and sub-subcodes into categories based on frequency” (Saldana 2013, p.80)*

Saldana (2013) suggested code landscaping as a useful preliminary analytic technique to organise and assemble codes (2013, p.194). This was a simple innovative way to examine text and thus codes using a visual method of “tags”. The word cloud immediately allows visualisation of the words that have been used the most in the data corpus. Wordle (wordle.net) was an instrument that could be used for this. A list of all the word counts from the data corpus was obtained. This provided a high level visual view but no data analysis or description. This was a comparable method to Tag Cloud and Cluster Analysis in other Qualitative tools (Saldana 2013, p.199). Word clouds were used to allow quick visualization of general patterns for preliminary analysis and to help validate and interpret findings (McNaught & Lam 2010). The transparency of using Wordle is in the actual word counts. The only things which are removed are the “stop word” (a frequently-used, but unimportant word, such as “the”, “and”, or “but”) and by default, Wordle strips numbers from the text before drawing the picture e.g. “1 apple” would display as “apple”. Wordle does not provide stemming. Stemming means understanding different words as variations of some root or stem, e.g., "walking", "walked", and "walks" are understood as variations on "walk", but Wordle would include all of these as separate words.

To provide an audit and more specific output a second instrument was used, Power Query, to count the words ignoring extremely common words prepositions, postposition, conjunctive adverbs, transition and linking words (e.g. stop words).

This allowed the outlining of text frequencies to be reviewed for the main semantic content, the top x number of words being examined depending on the size of the text. Note that this was not an indicator of significance. Braun and Clarke (2013) argued against using frequency counts when reporting patterns in the data. These could be useful to report certain practices. Braun and Clarke discussed the insightfulness or importance and mentioned that the counts of codes or themes could show a prevalence of them across the data corpus.

*“frequency does not determine value [...] It is important to note that these terms are not in any way attempting to “count” the instances of a theme’s occurrence (as per content analysis), but rather to provide some indication of the strength or consistency of a theme.” (Braun & Clarke 2013, pp.261–262)*

## **TP2: Code relations chart**

Code landscaping provided a check on the prevalence of key textual words. Code relations are reconsidered by iterating through the data set to search for overlaps in coded passages, sequences and proximity (Saldana 2013, p.32). This review process enabled the mapping of patterns in the data set, relationships and processes by visualizing the data (Lewins & Silver 2014, p.10). The original output from this helped with order and cognitive grasp of the data items. The Code Relations spreadsheet allowed the determination of the weight of the different codes.

The interconnections showed the prevalence within the data set, an example was as follows:

List of interconnections, connecting codes

- A -> B A connected to B

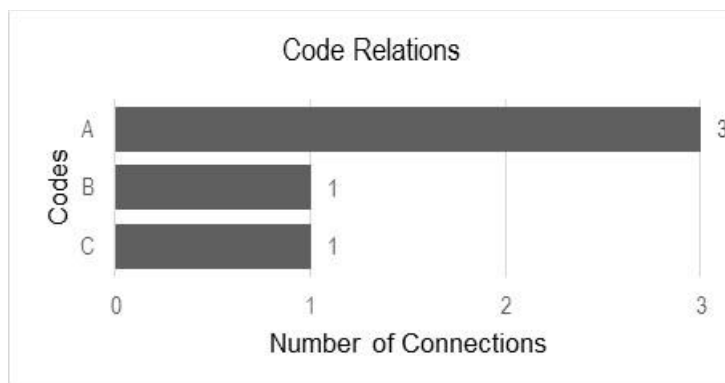
- A -> C A connected to C
- B -> C B connected to C
- A -> B A connected to B
- C -> B C connected to B

This list was then converted into Table 3.10.

**Table 3.10 Interconnections and prevalence**

	<b>A</b>	<b>B</b>	<b>C</b>	<b>Total</b>
<b>A</b>		2	1	3
<b>B</b>			1	1
<b>C</b>		1		1

The total number of components A that have some connection, interconnections, between them was as Figure 3.8.



**Figure 3.8 Code relations**

This visual representation displayed the codes with the highest mix of connections.

**TP3: Operational Model Diagram**

The operational model is a blueprint for how the components of a system work, together using an abstract visual representation. It can often include people, process and technology. In this research, the operational model took a holistic view of the code relations, grouping the codes into logical groups and named systems. This set of groupings were visually displayed to show the emergent relations of the codes. The concepts of systems mapped in this manner started to demonstrate the transition from analysis to synthesis.

Operational model diagrams can show various shapes, connecting lines or links (solid and dashed) and arrows (one way or bidirectional to show space, flow, stream of convergence, action, reaction, interaction and sense of quality and magnitude (Saldana 2013, p.202). The operational model diagram focused on a particular area to ensure clarity was maintained.

An output of this operational model design was the representation of the logical groups to which the codes belonged and illustrated the group relationships. This model demonstrated how the high level groupings were connected within the strategy and design within the holistic model. This stage was used to move from the real world entities to the conceptual world of systems (Holwell & Reynolds 2010).

### **3.9.3 Synthesis: Systems Thinking**

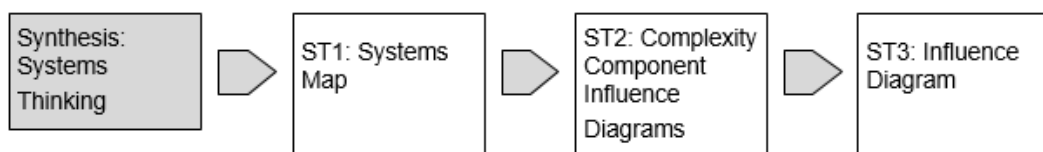
Within the thematic analysis the codes and themes were created from the data set. The transitional process began the transition from the reductionist approach to a holistic approach. This synthesis systems thinking stage presents the results of the data in a holistic way that identifies interconnectedness of the components and complexity. This research diverges and expands its scope from traditional thematic analysis to move from an analytical research, where the data was considered as a

whole to be broken down, to synthesis thinking where the components were a part of the whole (Ackoff 1981a, pp.16–17).

Systems thinking is about taking a holistic approach, thinking in wholes, rather than a reductionist approach, to investigate the interconnectedness of the components in the complex system. Systems thinking can be used for looking at the unpredictable behaviour of complex systems, can have feedback loops, varying people's behaviour and unintended consequences. Complexity can be created from a combination of factors, different perspectives, conflicting decisions and uncertainty. Systems thinking is defined in chapter one and two (Checkland 1999, p.318; Ackoff 1981a, p.15).

Systems thinking is used in many diagrams to help explain the situation (Reynolds *et al.* 2014a; Reynolds *et al.* 2014b). Diagrams describe structure or process. The Open University (Lane *et al.* 2012) classify diagram types, four of which are used in this research. This research uses structured diagrams: Spray Diagrams, a Rich Picture, Systems Maps and Influence Diagrams.

The systems thinking stages of analysis are shown in Figure 3.9.



**Figure 3.9 Synthesis: systems thinking**

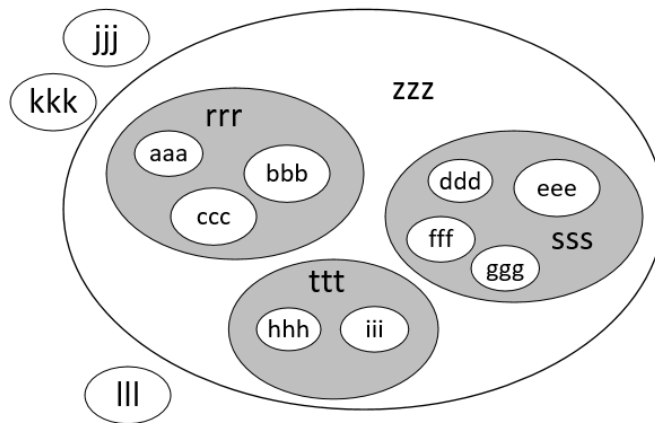
### **ST1: Systems Map**

This stage consolidated the work started in TP3 with the operational model and the groupings. Within this stage the terminology changed and the mapping is below:

- Groups are relabelled Systems and Subsystems.
- Codes are relabelled Components.

The purpose of a systems map is to generate insight and to understand the systems at the point in time the data was collected. It was possible that several iterations might be required to represent the situation clearly.

In this stage it was important to define the system, its components and subsystems and its components.



**Figure 3.10 Format for a systems map  
(The Open University 2012)**

A systems map (Figure 3.10) provides a snapshot in time of the situation or issue being explored and its environment. The system of interest is the situation being investigated. The system map is from the perspective of the researcher constructing it and can have a boundary separating the components of the system and subsystems from the environment. The components can be grouped together as subsystems. Words (e.g. aaa, bbb, ccc, ddd) name each system or component.

*“Use of systems maps help to identify the themes and elements that you see as being relevant to an issue” (Reynolds et al. 2014a)*



### ST2: Complexity Component Influence Diagrams

Data items from the data set were drawn together to provide a visualisation of the system of influences. Individual influences between the components (i1 – i6) in Figure 3.11, were described and could be in more than one direction as shown by the arrows. Each of the influences was explained and the data item quoted. The elements of the research to study in depth should be a result of the Transitional process, TP1-3, and ST1.

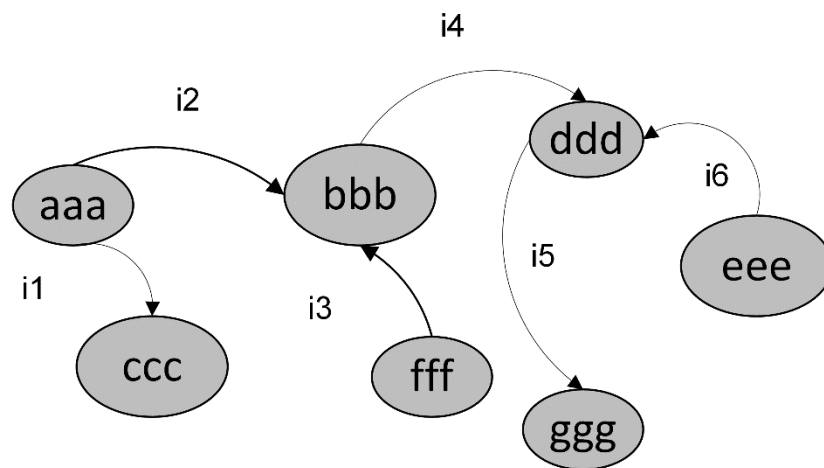
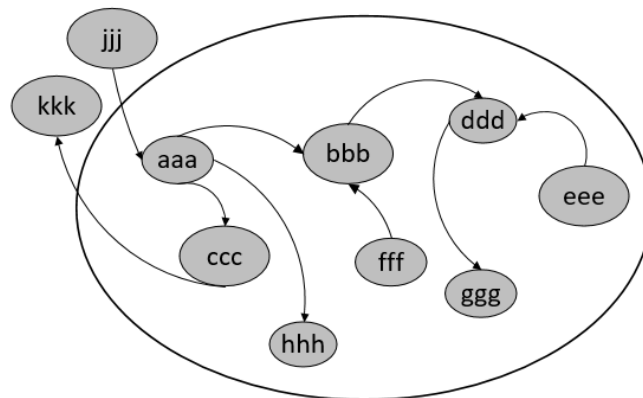


Figure 3.11 Component influence diagram

### ST3: Influence Diagram

The final stage built on the sample influence diagrams created in ST2 to create one combined diagram. The influence diagram contains the combined elements, codes in the situation, and shows the interconnectedness of the components and the complexity within the system. A graph rendering tool provided a way of easily visualizing the influences.



**Figure 3.12 Format for an influence diagram  
(The Open University 2012)**

An influence diagram (Figure 3.12) is a snapshot of a situation including organization features, people and relevant elements in the system. The influence diagram can be used to explore broad interrelationships with strong or weak influences. The arrows indicate the direction of the influence. The arrow joining component aaa to component bbb or ccc shows that aaa can or does influence bbb or ccc. Words (e.g. aaa, bbb, ccc, ddd, etc.) label components

*“Influence diagrams identify the factors (structural features such as people and events) that have direct and indirect influence on a system and its environment.” (Reynolds et al. 2014b)*

Influence diagrams can be developed from systems maps and sometimes the thickness of the lines can show different strengths of influence.

At the end of the Synthesis stage the data was analysed, themes created, the components found and complexity identified within the situation.

### 3.10 Summary

This chapter has discussed the research strategy which used a mixed methods approach. The specific design chosen was sequential explanatory design. This is quantitative data collection and analysis followed by qualitative data collection and analysis. This captured both trends and in-depth details of the complex situations. The quantitative and qualitative analysis stages were discussed and a process methodology presented to provide a robust examination of the data collected.

The qualitative data used thematic analysis for the analytic process to highlight codes and themes within the data. The final stage was that of synthesis, using systems thinking, considering the components as a part of the whole system.

The subsequent two chapters present the results of the quantitative survey and the qualitative focus groups.

## Chapter 4: Quantitative Survey Findings on the Utilization of Best Practices

### 4.1 Introduction

The first stage of data collection was to gain an understanding of best practices and procedures and their usage within the database community. Best practices are those practices that have demonstrated successful and effective outcomes when managing database systems. This stage of data collection was to help answer the first research question: “to what extent are best practices and procedures utilised by the database community?”

At the same time the survey looked at the handling of data throughout the database lifecycle. Together these quantitative survey questions have led to the investigations in the second phase, the qualitative survey. The survey findings reported in this chapter refer to the quantitative data survey and share the experiences of the respondents in the operation of databases and supply background information related to the management of the databases. Appendix A contains the full list of quantitative questions asked in the survey.

The survey data collection using Survey Monkey ([www.surveymonkey.com](http://www.surveymonkey.com)) was open between 13 December 2012 and 6 February 2013. The responses to the survey over this time had an initial surge, plateaued over the Christmas period then with another marketing campaign at the beginning of the New Year the number of respondents increased again. The total who started the survey (Partial responses) were 453 respondents. Partial responses are where the Next button is clicked and the Done button is not. The total who finished the survey (Completed responses)

were 226 respondents (49.89%). Completed responses were where a respondent answered at least one question on every page, clicked the Next button and the Done button.

The responses were checked for completeness to monitor the sampling, and to see whether all the questions were being answered. This helped manage the marketing of the survey to encourage more responses.

## 4.2 Survey Findings

The survey investigated the current practices and procedures that were used, together with respondents' perspectives on 'best practice'. The main areas of the survey comprised: demographics, respondents' organizations, understanding of best practice, control of best practice, database demographics, database servers, training, database architecture design and development, database technical practices, database operations, cloud databases, data management, application-centric, change management, organizational culture, improvement methods, and future vision. Findings around each of these areas are discussed in turn in this section. The full questions with question numbers (Q) that are listed in Appendix A are included in the figure or table description for clarity and ease of reference. As well as the 2D charts there are some charts that show cross tabulation when appropriate. The percentages are based on the actual number of respondents (N) for a particular question. The value of N for each question is added. The cross tab questions reported were the ones which gave the most insight. After an initial cross tabulation analysis, I analyzed 58 cross tabulations in depth to try to find relationships between more than one variable. The resulting five cross tabulations included are:

- Figure 4.2 The number of database administrators in different sizes of respondents' organizations
- Figure 4.21 Type of database training received for the length of time working in the field
- Figure 4.44 The types of cloud database services used for the type of environment
- Figure 4.51 Database product selection constrained by employee skillset for data requirements, driving database management
- Figure 4.52 Legal procedures followed for data with historical data policies

Figure 4.2 shows that, in organizations of any workforce size, 2-5 administrators is the most common number of administrators. Figure 4.21 shows that, database training for people who have worked in the field over 10 years mostly comes from reading articles and more training is provided to people with over 10 years' experience.

Figure 4.44 shows that, different environments (private cloud, public cloud and hybrid cloud) may be chosen for different types of work. Figure 4.51 shows that, data management was driving database management procedures and database software product selection was constrained by the skill set of in-house employees. Figure 4.52 shows that, set procedures were followed for legal reasons for the great majority of cases of historical data.

Other options such as organisational size (Q4) and experience of participants (Q6) provided no further insight than could be gained from the individual charts. The charts that were included added to the story; there was not space to include every combination.

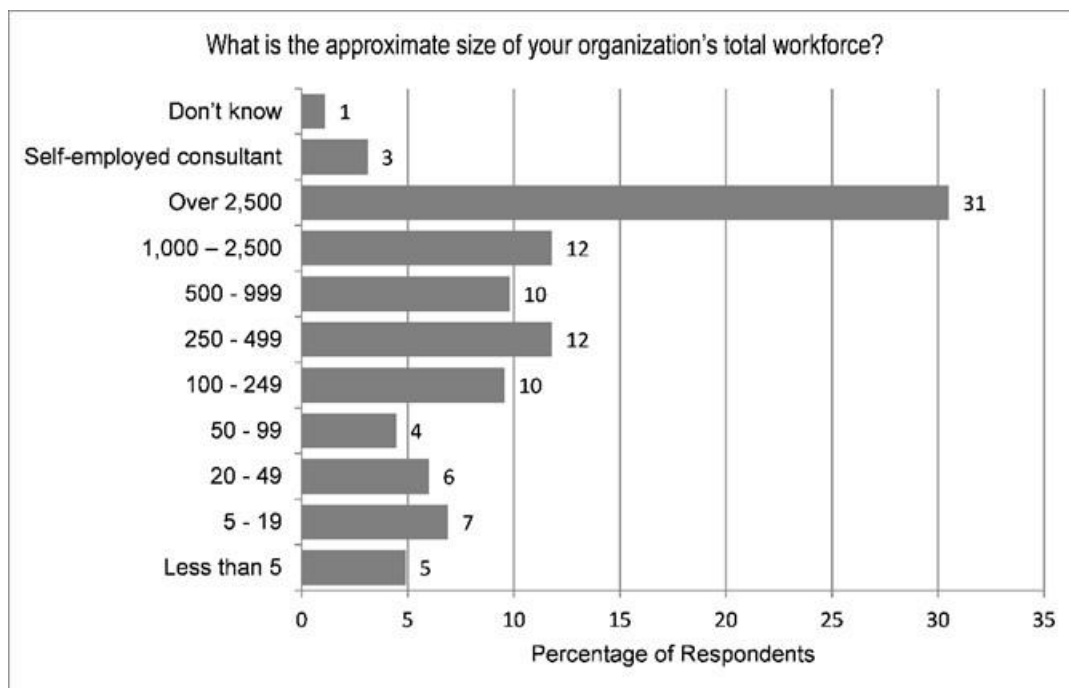
### 4.2.1 Demographics

The demographics section sets the scene for presenting the data showing the diversity of roles, participating countries, industry, workforce size and length of time in the field. The respondents came from a diverse spread of job roles. The three largest groups were database administrators (43%); database developers (15%); and business intelligence (BI) roles, including BI analysts, data analysts, data scientists and BI architects (13%). In all the population was split into 28 different job roles covering all levels of management.

Respondents worked in 40 countries, with the majority based in the USA (40%) and the UK (33%). The remaining 27% were divided amongst 38 countries with no single country greater than 10%.

The respondents worked in 34 different industry sectors, of which the technology sector accounted for 25%; banking, insurance or financial services 21%; healthcare 12%; professional services 12%; and education 10%.

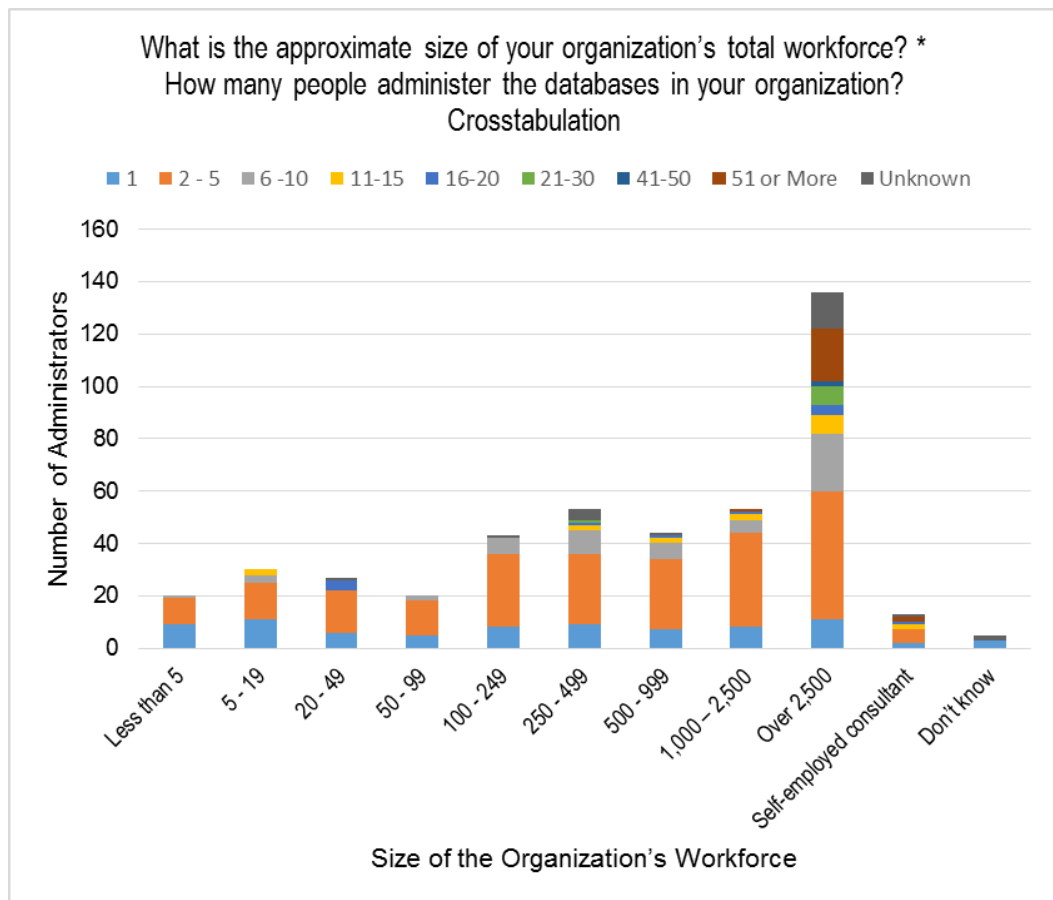
As Figure 4.1 shows, 30% of the respondents worked in large organizations with over 2,500 employees. Every category of organizational size was represented among the survey respondents.



**Figure 4.1 Size of organization’s workforce**  
**Q4 (the question number in the survey Number) N=449 (Number of Respondents)**

Figure 4.2 cross tabulates the size of the organization with the number of database administrators in the respondents’ organizations. In a particular size of organization the number of administrators can vary. This is especially the case for organizations over 2,500 staff, with 11 organizations saying they have 1 administrator, and 49 organizations having 2-5 administrators. In organizations of any workforce size, 2-5 administrators is the most common number of administrators. The number of administrators for the size of the company could indicate how diverse the management of database systems is, and could indicate the need to communicate with more people.





**Figure 4.2 The number of database administrators in different sizes of respondents' organizations**  
**Q4 \* Q10 Cross tabulation N=449\*N=449**

Slightly more than half of the respondents (53%) had worked in the database field for over 10 years, with 29% working in the field between 5 and 10 years (Figure 4.3). The length of time working in the field could indicate increased level of experience and mean that the people have probably adopted changes that have come about due to advances in technology.

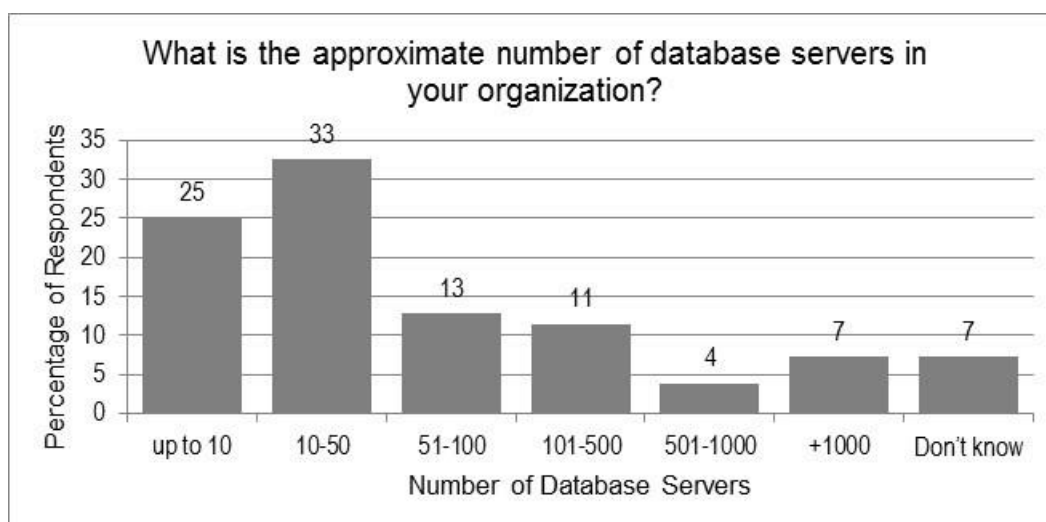


**Figure 4.3 Experience in the field**  
Q6 N=447

## 4.2.2 Respondents' Organizations

Within each organization it was important to understand the respondent's database estates, the number of servers, the number of people administering the servers and whether there were specific roles associated with the tasks. The amount of time spent managing servers may or may not be related to outcomes.

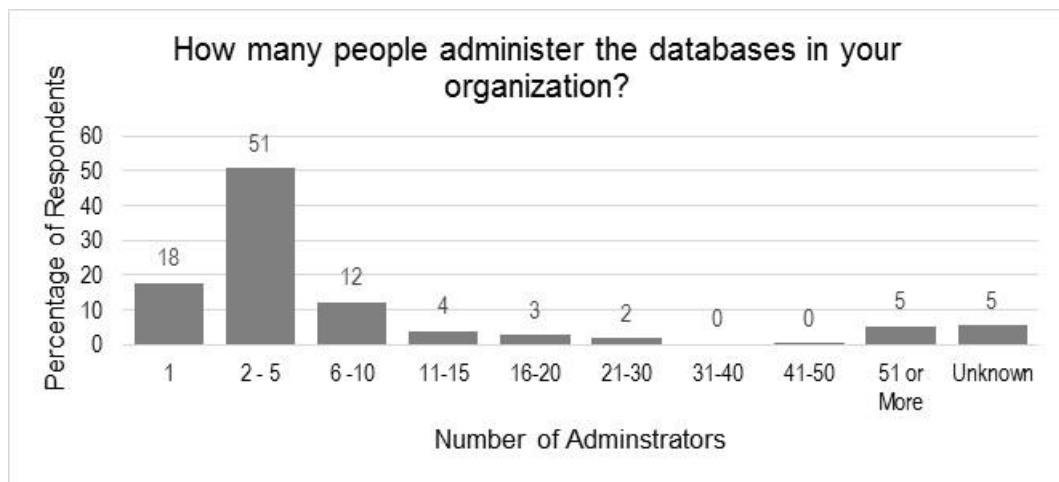
As Figure 4.4 shows, 33% of respondents had 10-50 database servers in their organization, with 25% reporting up to 10 database servers.



**Figure 4.4 Organization's number of database servers**

**Q9 N=446**

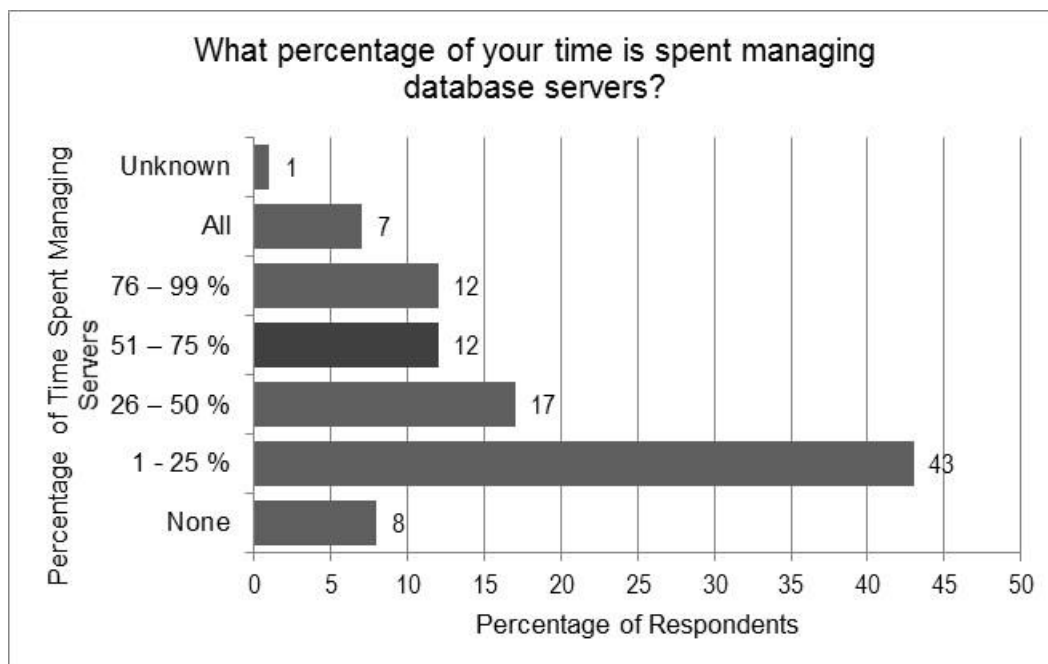
As shown in Figure 4.5, 51% of respondents reported that 2-5 people administered databases in their organization, although 18% of respondents reported that there was just one person. (Note that this role may not have been confined to those with a job title of Database Administrator.) Figure 4.2 looked at people administering databases against the size of the workforce.



**Figure 4.5 People administering the databases**  
**Q10 N=445**

Turning to database-related roles, Question 5 found that 78% of the respondents stated that their organizations had the role of database administrators or database engineers; 58% of organizations had database developers and 44% had staff holding Business Intelligence roles.

Time spent by respondents in managing database servers varied (Figure 4.6), with 7% of respondents spending all their time on these tasks, while 43% spent less than a quarter of their time on this.



**Figure 4.6 Time spent managing database servers**  
**Q17 N=416**

### 4.2.3 Understanding Best Practice

A common question raised when trying to improve the management of database systems is where best practices can be found and what is contained within them. Best practice documentation was provided by vendors to help users achieve the best out of their products and to guide the user when configuring the systems. Identifying whether organizations follow any best practices, identifying their importance and the issues for their adoption, provided insight into the power behind them. Chapter 2 discussed best practice in more detail.

As shown in the literature review, best practice can mean different things to different people. Respondents selected a definition of best practice from a list of definitions (Figure 4.7). The most frequently selected definition was “recommended practice”. This raises the question recommended by whom.

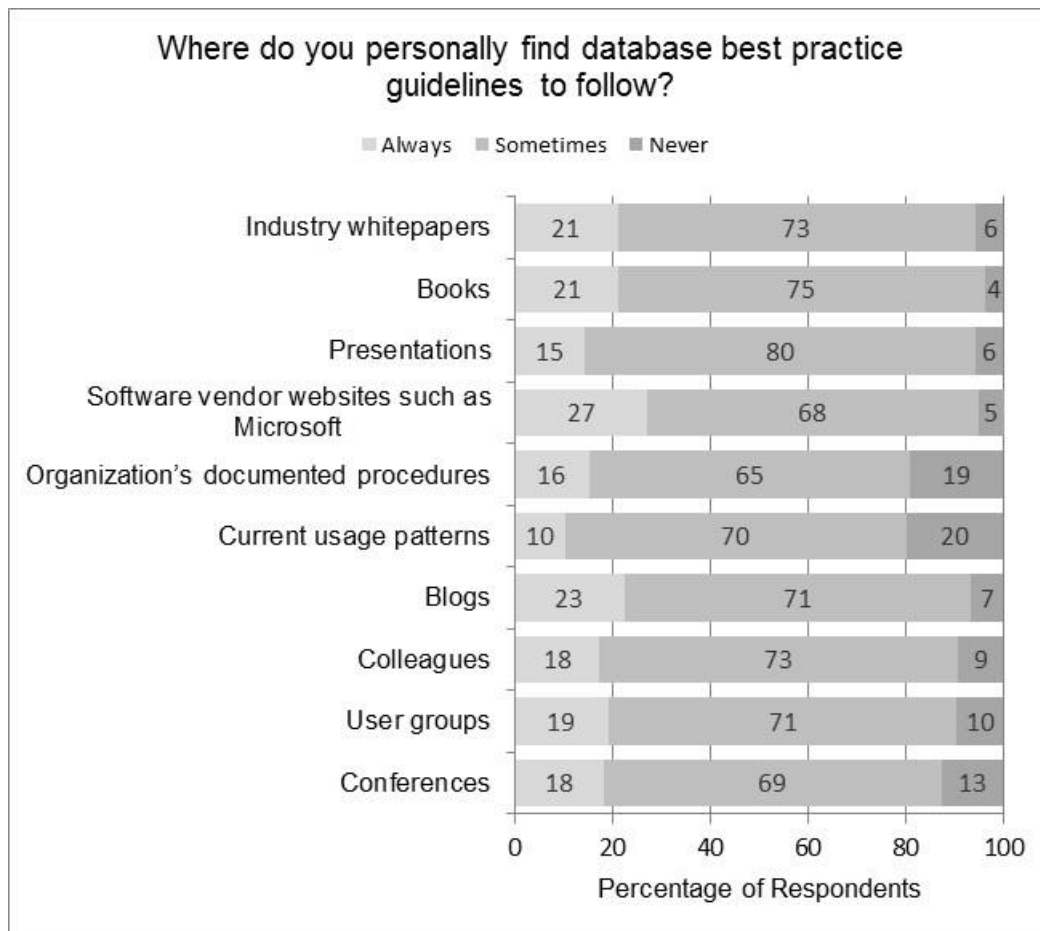


**Figure 4.7** Frequency table of best practice

**Note:** percentages do not total 100% because respondents could check all that apply  
**Q74 N=219**

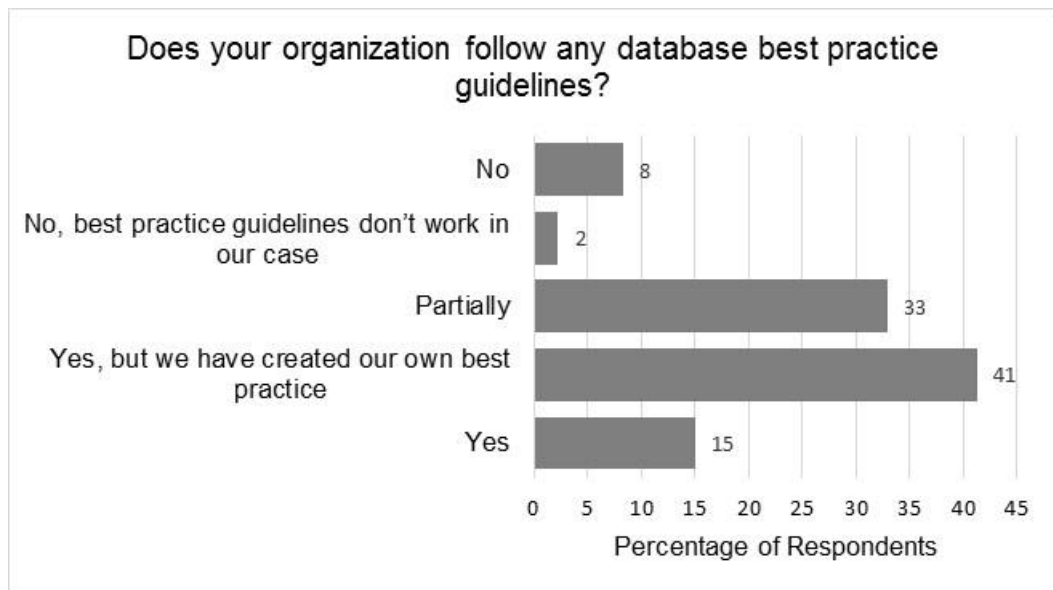
An additional comment, in Q74, from a respondent answering this question highlighted the complexity: *“Best practices are NOT set in stone, simply because project requirements are always differing and technologies that sit on top of the database layer change, which can create new best practices.”*

Best practice guidelines originated from a variety of sources and software vendor websites such as Microsoft were used by 27% of respondents to find best practices (Figure 4.8).



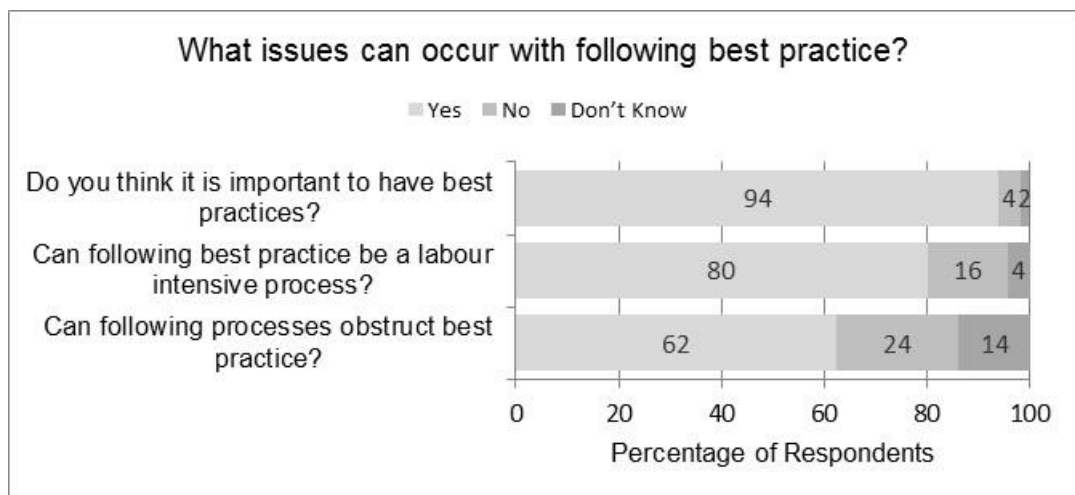
**Figure 4.8 Where do you find best practice Q75 N=219**

41% of respondents reported that their organizations followed best practice guidelines, through creating their own best practices. 10% said that their organizations did not follow best practice guidelines (Figure 4.9). With organisations creating their own best practices, this raises the question whether there are any standard best practices across the industry.



**Figure 4.9 Organization following best practice**  
Q76 N=218

An overwhelming 94% of respondents thought it was important to have best practices (Figure 4.10). 80% of respondents thought that following best practice was a labour intensive process. 62% of respondents thought that following processes could be obstructive to best practice.



**Figure 4.10 Issues with best practices**  
Q77 N=218

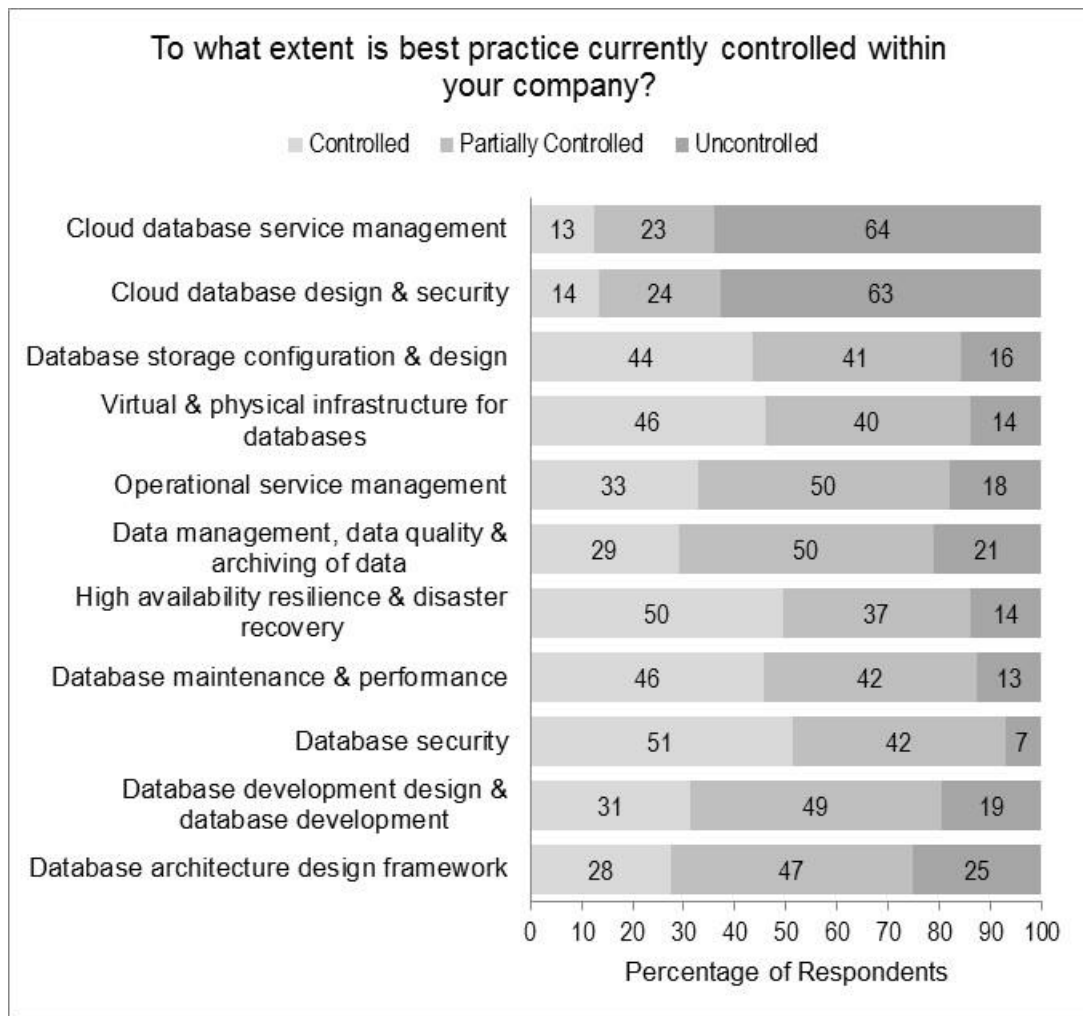
#### **4.2.4 Control of Best Practices**

Once the nature and usage of best practices has been established in an organization (Section 4.2.3), maintaining the quality of best practices within database areas, and improving who controls these choices, enables standards to be maintained.

Figure 4.11 highlights different areas within the database system throughout the management lifecycle. The respondents had to select the level of control they thought could be attributed to each area. The most highly controlled areas were database security, and high availability resilience and disaster recovery.

The areas the respondents reported as most uncontrolled were cloud database design and security, and cloud database service management. Overall, best practices are partially controlled. Cloud services might be marked as uncontrolled as these are probably controlled by the vendors but not the organizations.



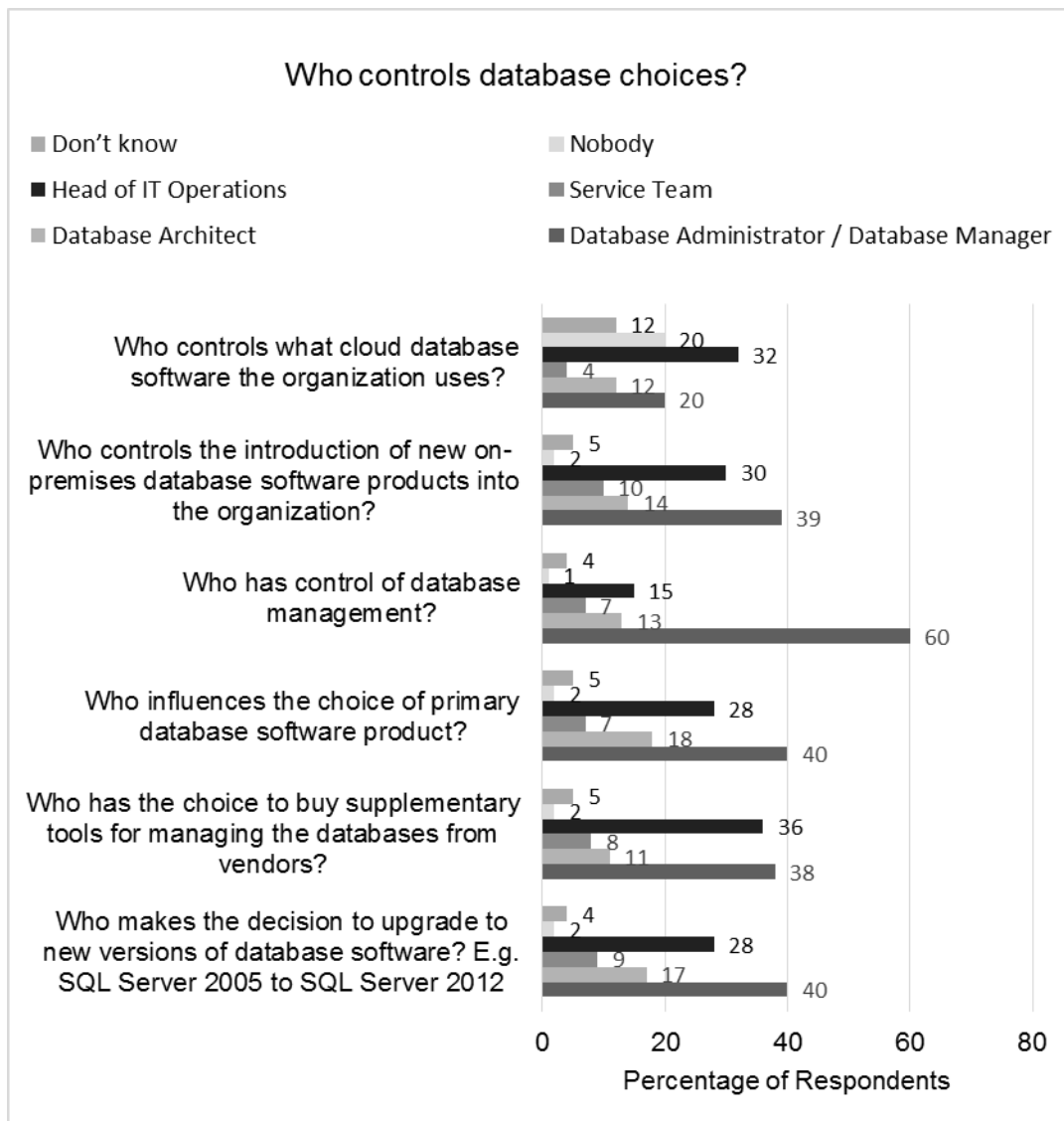


**Figure 4.11 Control of best practices**

**Q79 N=213**

Examining who controls the products, tools, hardware placement and database management was important. The control of these products and tools influences how administration can be carried out. The control of these areas affected the database ecosystem. For each question the respondents had to select who they thought controlled the choices in the various areas.

The “Database Administrator / Database Manager” was reported by the respondents as controlling database choices in most areas, except those organizations which used cloud database software (Figure 4.12). This was controlled by the Head of IT Operations.

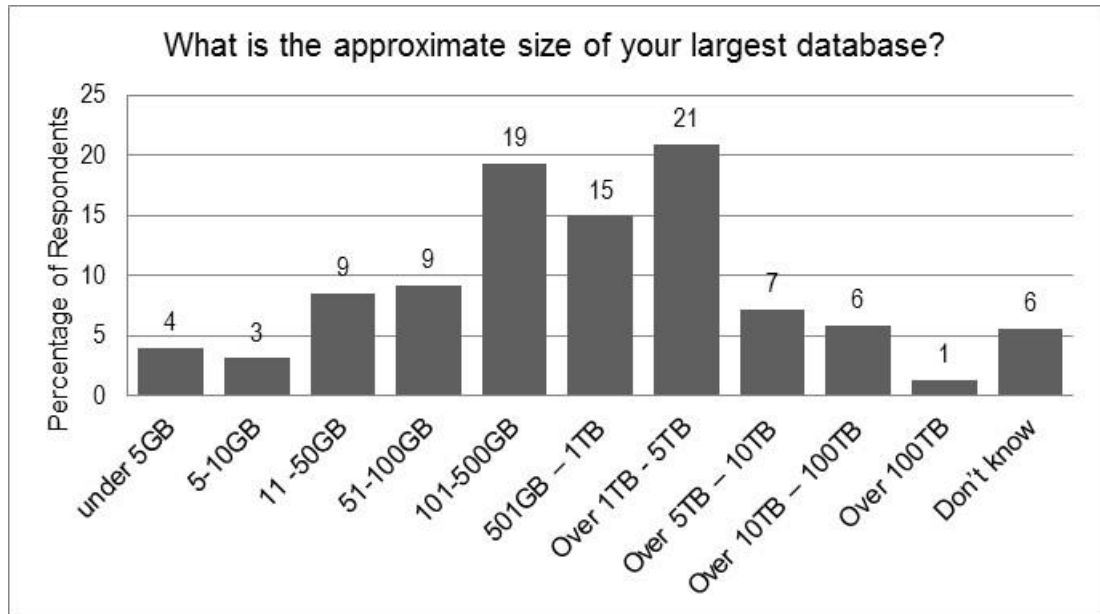


**Figure 4.12 Who controls database choices**  
Q70 N=226

### 4.2.5 Database Demographics

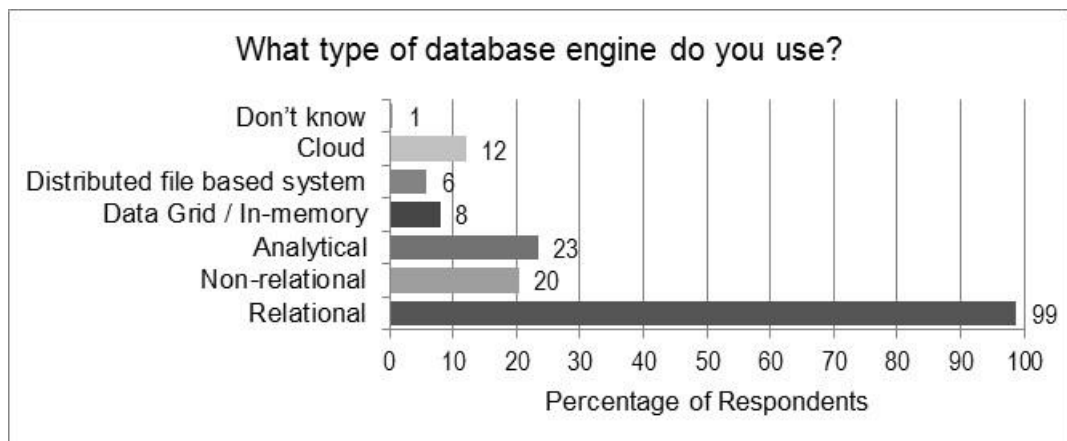
Database demographics provided a high level understanding of the size of the data involved, the types of engines used for storing this data and the specific applications. The service availability for this data is important, as different management techniques can be applied.

The largest database that was managed by the majority (55%) of respondents was between 101 GB and 5TB (Figure 4.13). These database sizes were large when the data was collected.



**Figure 4.13 Size of the single largest database (overall size e.g. including data & logs) Q8 N=446**

The type of database engine used is shown in Figure 4.14. Relational database engines were used by 99% of the respondents. Analytical database engines were used by 23% of respondents (Figure 4.14).

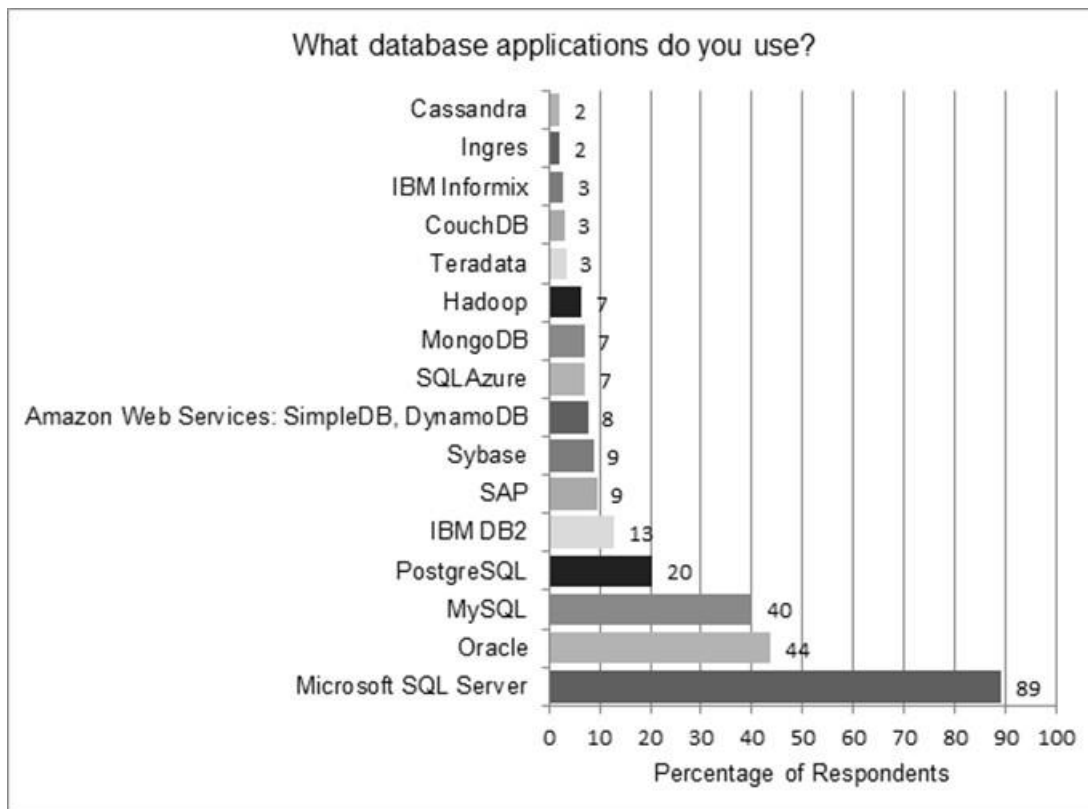


**Figure 4.14 Type of database engine used**

**Note: percentages do not total 100% because respondents could check all that apply**  
**Q12 N=414**

51 different database applications were used (Figure 4.15) including relational, NoSQL, NewSql, InMemory and cloud database applications. Microsoft SQL Server was the most frequently reported database application (89% of respondents), followed by Oracle (44%) and MySQL (40%).

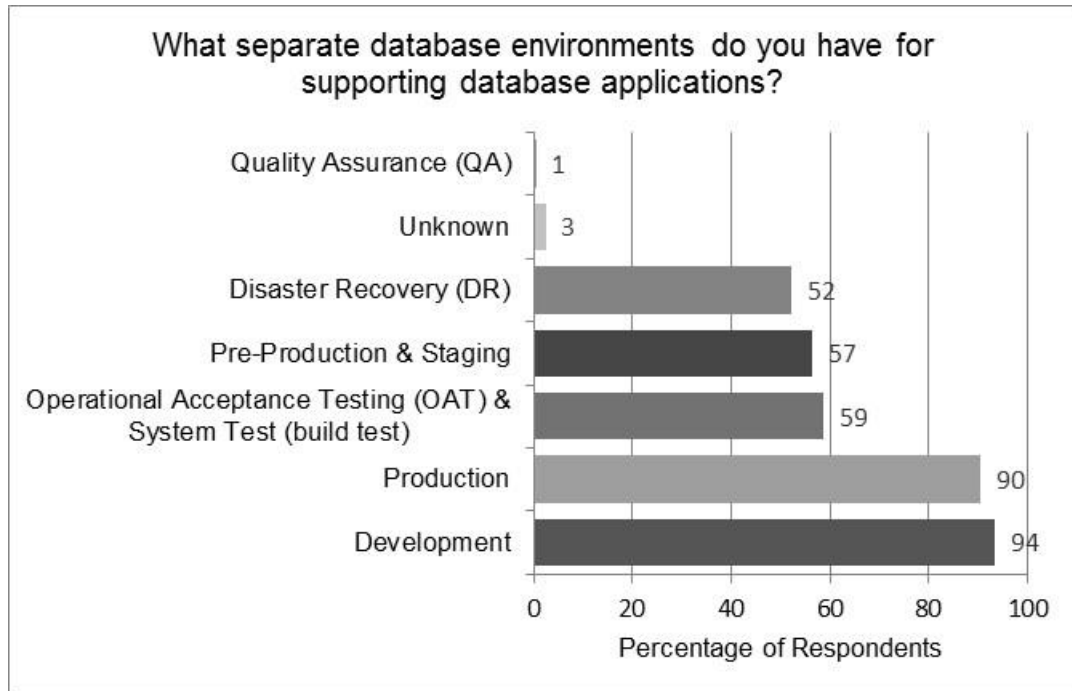
On average a respondent used 2.6 database applications, while 28% of respondents used just one database application. The top 16 applications used by 2% or more of respondents are shown in Figure 4.15.



**Figure 4.15 Database applications used**

**Note: percentages do not total 100% because respondents could check all that apply**  
**Q11 N=414**

Many separate environments are used for supporting database applications (Figure 4.16). Different environments have different use cases and some may need more administration than others.



**Figure 4.16 Database environments**

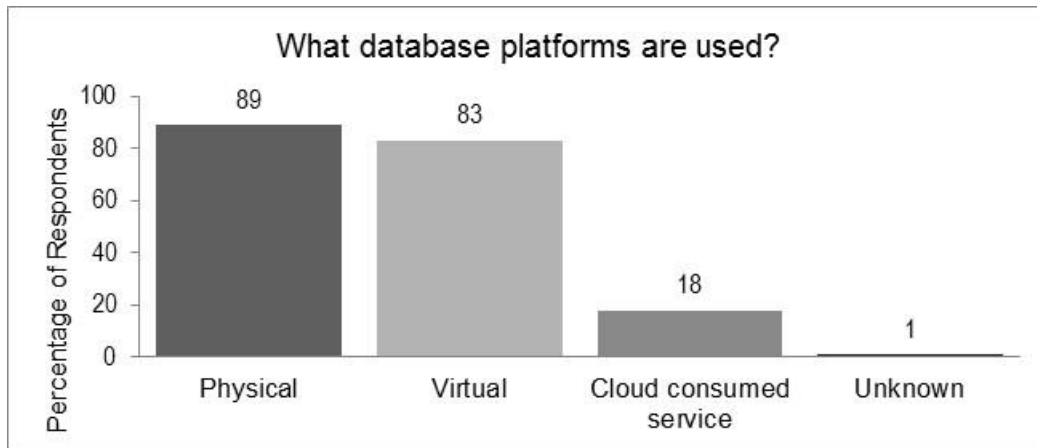
**Note: percentages do not total 100% because respondents could check all that apply; responses under 1% are not included**

**Q25 N=324**

#### 4.2.6 Database Servers

There are various platform types in use, with some on organizations' own premises and some in the cloud. These servers may be managed differently and may have different uses.

89% of respondents used physical database platforms, 83% used virtual database platforms and 18% used cloud consumer services for their database platforms (Figure 4.17). However, 66% of respondents reported that none of their databases used cloud database services (Table 4.2).



**Figure 4.17 Database platforms used**

**Note: percentages do not total 100% because respondents could check all that apply**  
**Q23 N=326**

7% of respondents do not use on-premises database software or outsourced database hosting (Table 4.1). This percentage is probably connected with the fact that cloud database services are in their infancy.

**Table 4.1 Usage on-premises database software**

**Q15 N=408**

Percentage use on-premises database software	Response Percent
None	6.9%
1 - 25 %	3.2%
26 – 50 %	3.2%
51 – 75 %	5.6%
76 – 99 %	23.5%
All	53.2%
Unknown	4.4%

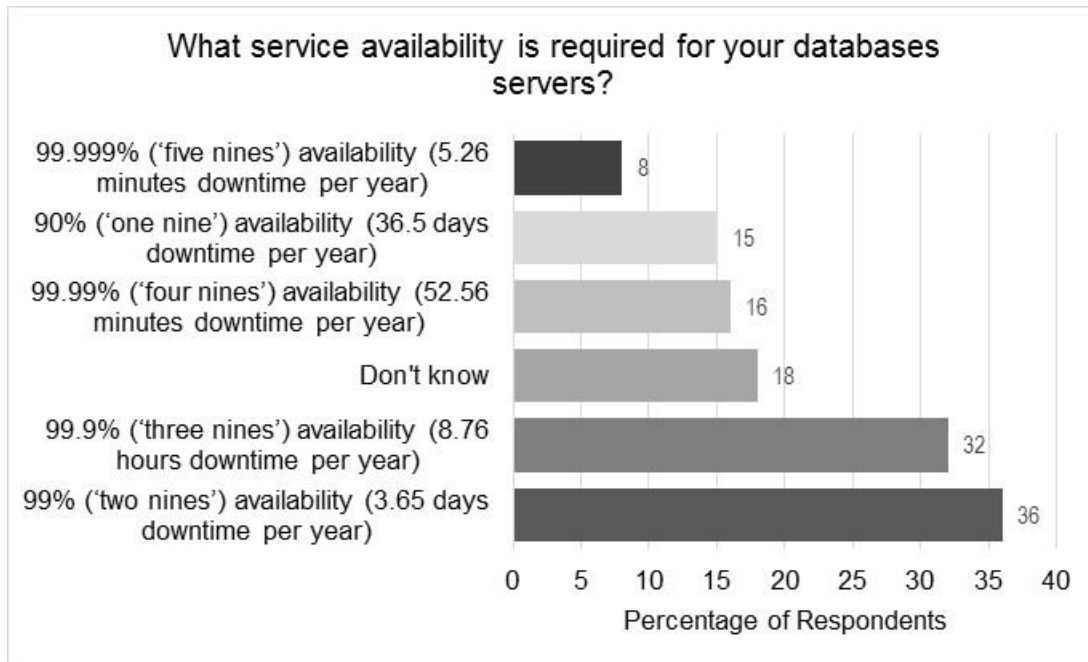
Table 4.1 and Table 4.2 indicate that both on-premises and cloud database services are used. Cloud database services offer a database service which is managed.

They are still fairly unusual: 66% of respondents do not use cloud databases (Table 4.2).

**Table 4.2 Cloud database services used**  
**Q16 N=414**

Percentage of Databases Use Cloud	Response Percent
None	65.9%
1 - 25 %	22.7%
26 – 50 %	4.3%
51 – 75 %	1.2%
76 – 99 %	0.7%
All	1.0%
Unknown	4.1%

There were various service availability targets (Figure 4.18). Service availability is the ability to perform a function for an amount of time that is agreed with a supplier in their service level agreements (SLAs).



**Figure 4.18 Service availability**

**Note: percentages do not total 100% because respondents could check all that apply**  
**Q29 N=306**

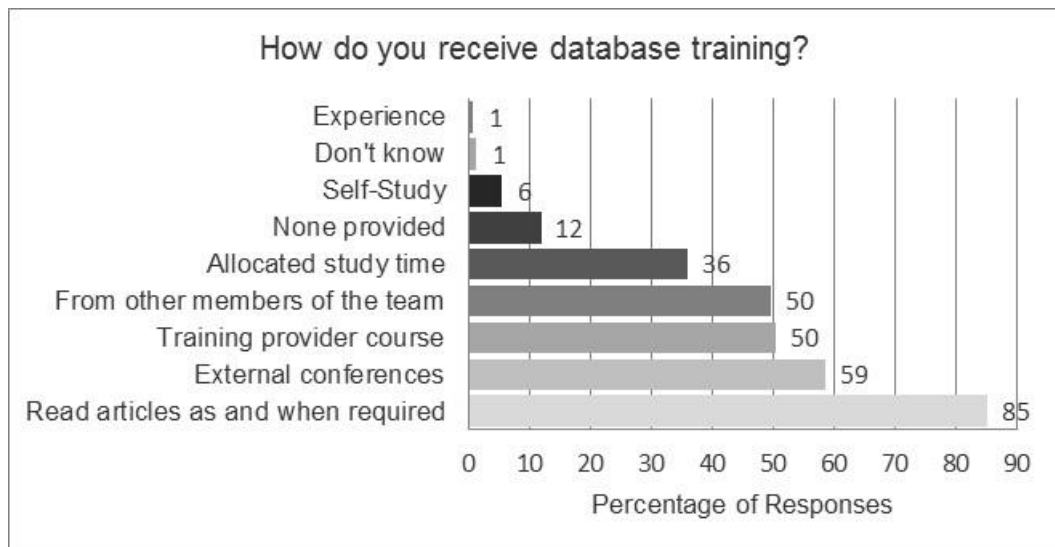
Some comments were that a variety of availability targets were offered. The availability targets of the respondents' databases showed that 8% were critical business services.

## 4.2.7 Training

Technology changes continually and with this change, knowledge of the new technology is required to be able to manage database systems. The dissemination of information about products, enhancements and best practices provided a method for understanding database systems.

Database training was obtained through a variety of methods. 85% of respondents read articles when required, and 59% of respondents receiving trained from external conferences. 12% of respondents reported that no training was provided (Figure 4.19).



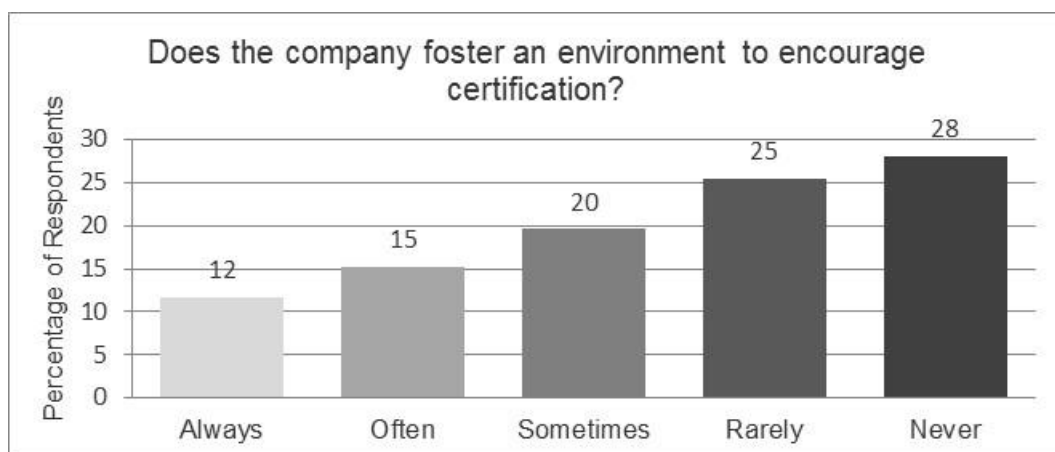


**Figure 4.19 Receipt of database training**

**Note: percentages do not total 100% because respondents could check all that apply; responses under 1% in this question are not included**

**Q60 N=236**

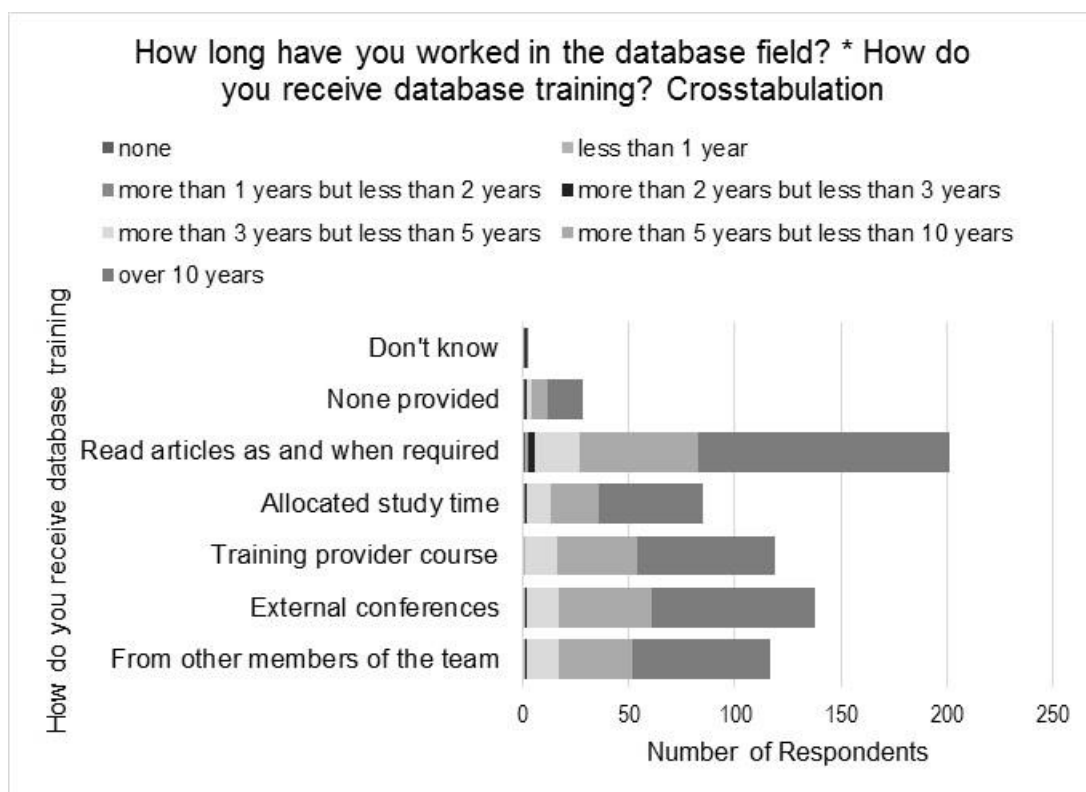
Professional certification covers the technology and new technology advancements with particular products. Although 46% of the respondents had professional certifications, 53% of respondents said their company never or rarely encouraged professional certification (Figure 4.20).



**Figure 4.20 Encouragement for taking certifications**

**Q66 N=228**

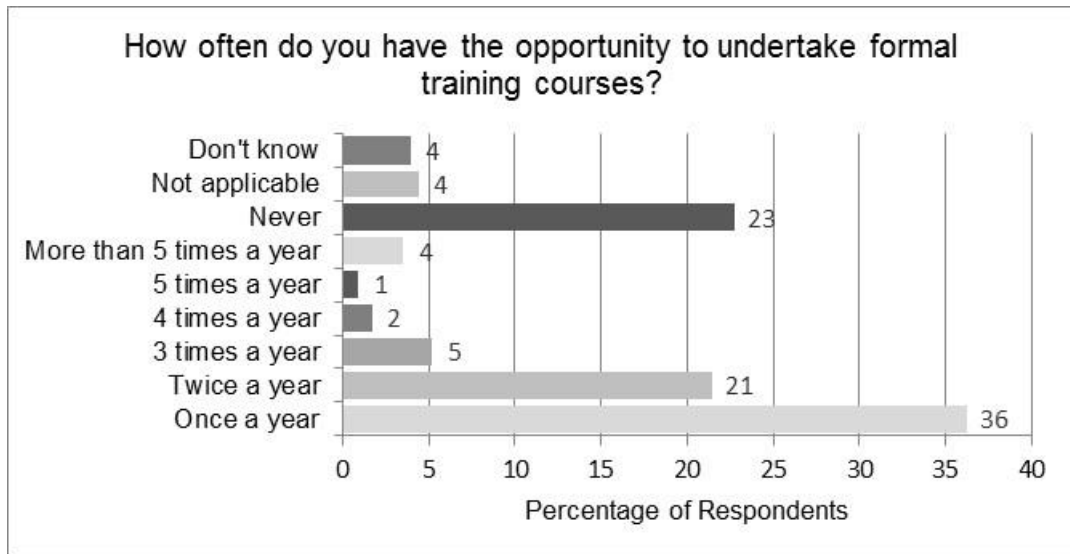
Database training for people who have worked in the field over 10 years (Figure 4.21) mostly comes from reading articles, 118 respondents, as and when required. Reading articles requires individuals effectively to train themselves, which may or may not happen depending on the workload placed on the individual. More training is provided to people with over 10 years' experience. Knowledge gained from training is used to configure and manage best practices.



**Figure 4.21 Type of database training received for the length of time working in the field**

**Q60 \* Q6 Cross tabulation N=236\*N=447**

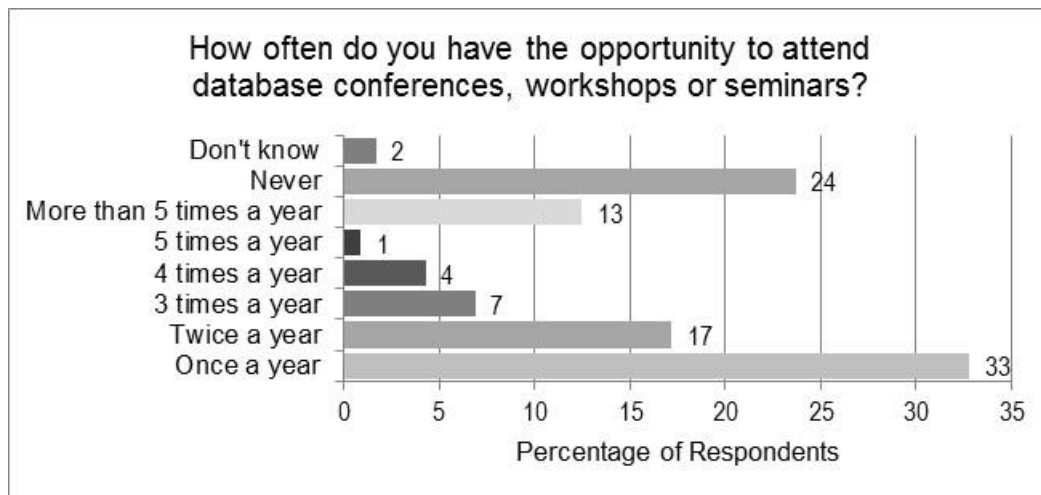
36% of respondents had the opportunity to undertake formal training once a year (Figure 4.22) yet 23% never had this opportunity. This compared with 12% for whom no training was provided (Figure 4.19).



**Figure 4.22 Opportunity for formal training**  
**Q61 N=229**

Many respondents took part in database community associations. These associations help share knowledge in the field through technical tips and training and provide networking connections. 23% of respondents reported that they were not involved in any user groups and did not have the opportunity to undertake formal training courses.

33% of respondents had a chance to attend conferences, workshops or seminars once a year, with others able to attend more often. However 24% were never able to attend conferences (Figure 4.23). Comments in the free-text section relating to this question included: “once every two years”, “not in a long time”, “irregularly” and “at my own expense”.



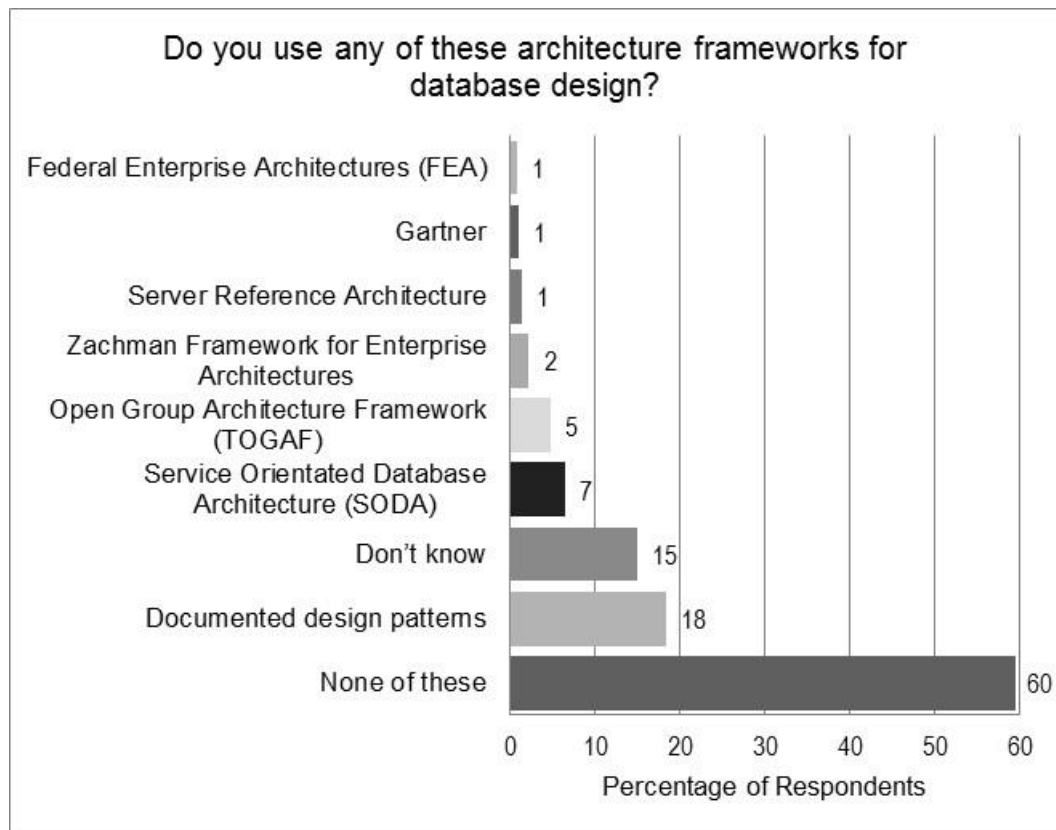
**Figure 4.23 Opportunity to attend conferences**  
**Q63 N=353**

It would seem that the level of training provided varies across organizations and in some cases is not well provided for.

#### 4.2.8 Database Architecture, Design and Development

This section reports findings from the initial stages of architecture, design and development. It firstly identifies the usage of frameworks, which were often followed as a means to show some best practices. The next stage ensured effective database management could be undertaken from the requirements gathered and documented through various means. Understanding whether any method of operation was followed to achieve desired outcomes was useful for repeatable purposes. Finally to understand the business models used for development could indicate the frequency of change and process of development, to understand how the level of accuracy was maintained.

The results in Figure 4.24 indicate that 60% of the respondents do not use common industry standard architecture frameworks for database design, although 18% use documented design patterns.



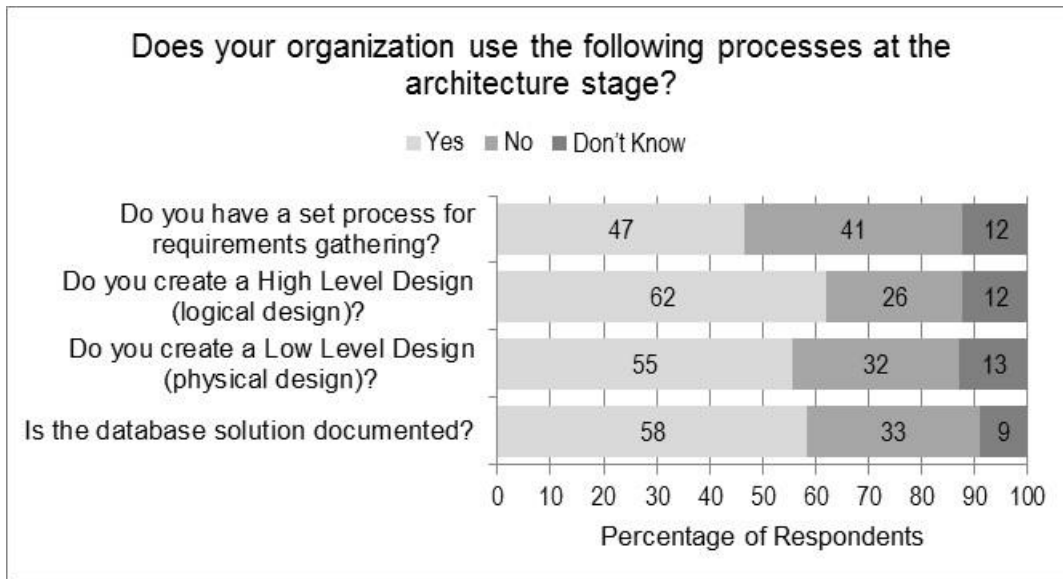
**Figure 4.24 Architecture frameworks**

**Note: percentages do not total 100% because respondents could check all that apply; responses under 1% are not included**

**Q18 N=353**

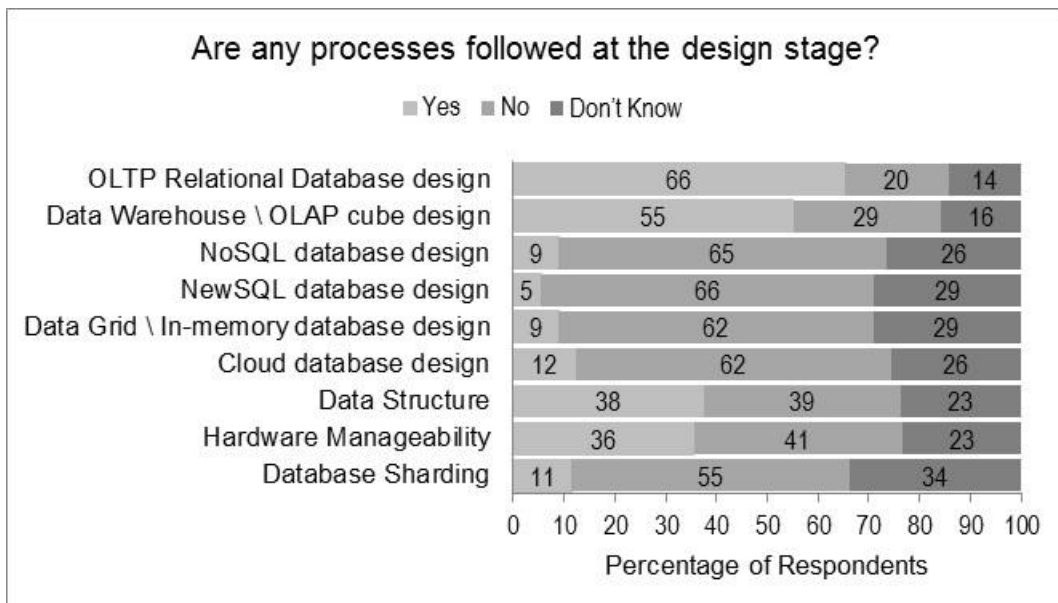
Most respondents reported that, at the architectural stage, high level and low level designs were created and the database solution documented (Figure 4.25).

However, 41% of respondents reported that no set process was employed for requirements gathering.



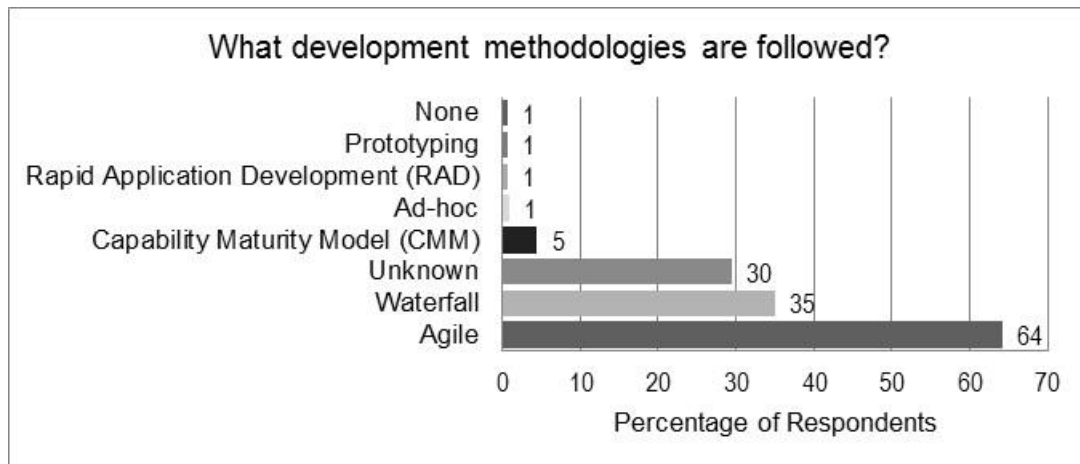
**Figure 4.25 Processes at the architectural stage**  
Q19 N=360

Figure 4.26 shows processes followed at the design stage. For relational and data warehousing processes the majority of respondents reported that some form of process was followed. For all other database engines there was an overwhelming lack of use of design processes. Design process usage was slightly higher for data structure and hardware manageability.



**Figure 4.26 Design processes**  
Q20 N=355

The development methodologies followed by respondents are shown in Figure 4.27. 64% reported use of Agile development methods and 35% reported use of Waterfall development methods.

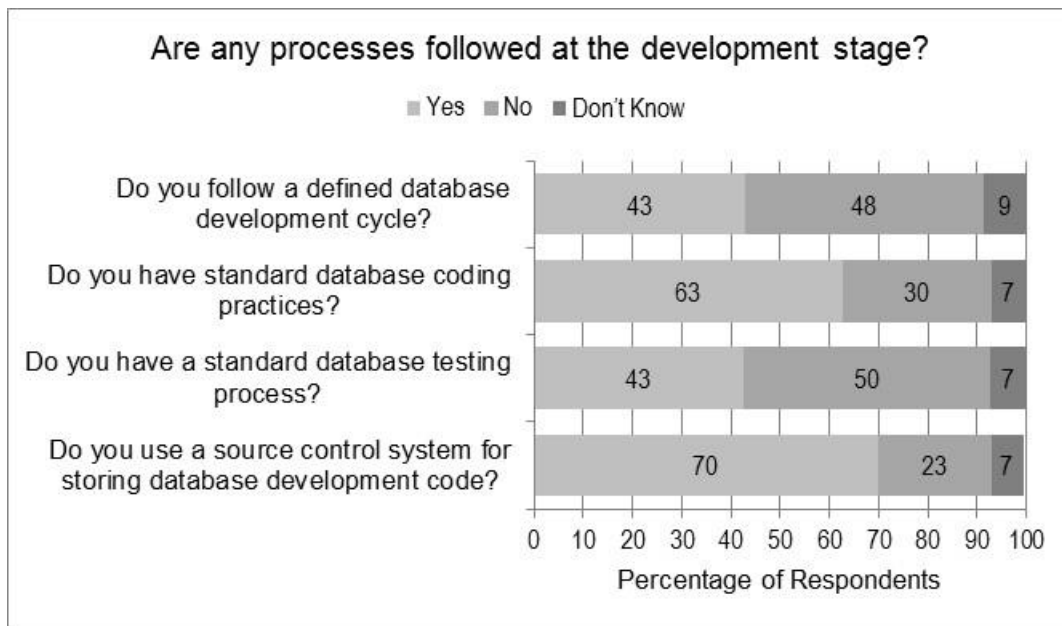


**Figure 4.27 Development methodologies**

**Note: percentages do not total 100% because respondents could check all that apply; responses under 1% are not included**

**Q21 N=336**

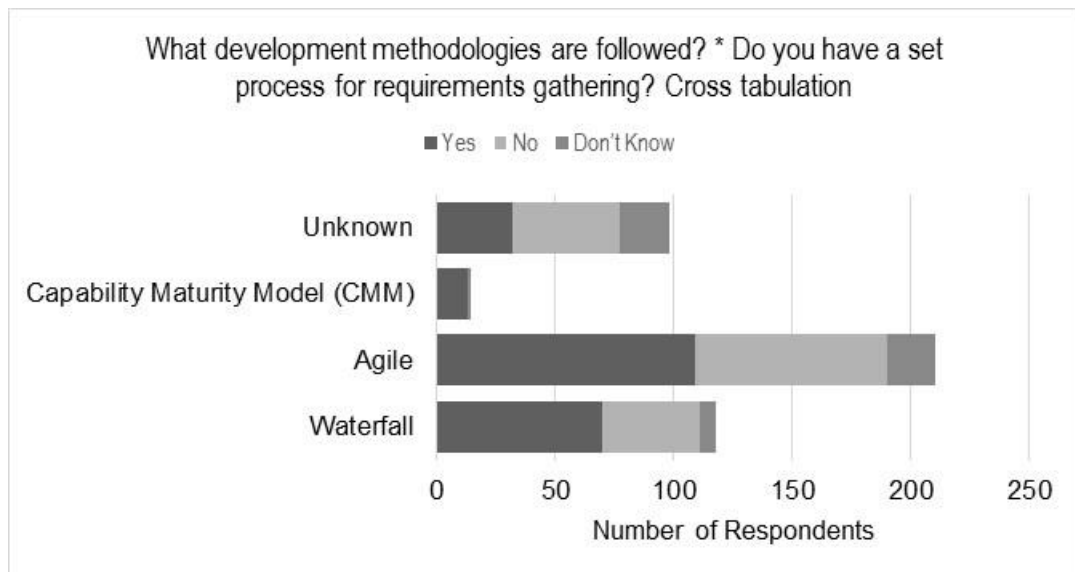
Figure 4.28 shows the extent to which core processes are used at the development stage. 50% of respondents did not have a defined set of standard database testing processes and 48% of respondents did not follow a defined database development lifecycle. By comparison, 70% used a source control system and 63% had standard database coding practices.



**Figure 4.28 Development stage processes**  
**Q22 N=358**

Requirements gathering is key to databases development and important in which ever development methodology is used. Cross tabulation Figure 4.29, showed that requirements gathering at the start and during changes in operation was often, by 109 respondents, dealt with using Agile methods. 70 respondents used Waterfall for some requirements gathering. Using Agile rather than Waterfall enables requirements to change throughout the process.



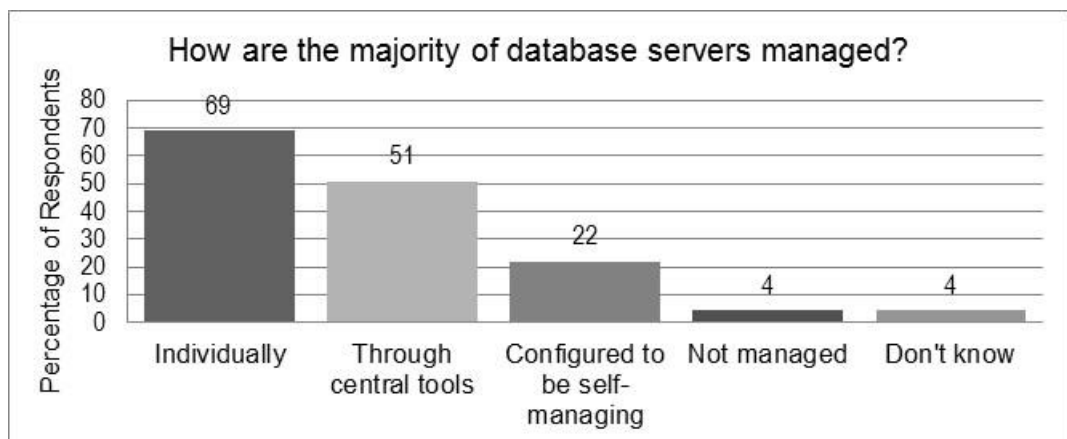


**Figure 4.29 Development methodologies and the process for requirements gathering**  
*Q21 \* Q19 Cross tabulation N=360\*N=336*

### 4.2.9 Database Technical Practices

Databases can be managed in various ways, manually or by automation. These need building, securing and for management tasks to be completed.

The majority of database servers were recorded as being individually managed (Figure 4.30), although 51% of respondents stated they were managed in some way by central tools.



**Figure 4.30 Servers managed**  
*Q28 N=323*

The installation and configuration of the database servers was carried out manually using the GUI (graphical user interface) by 62.5% of respondents (Figure 4.31).

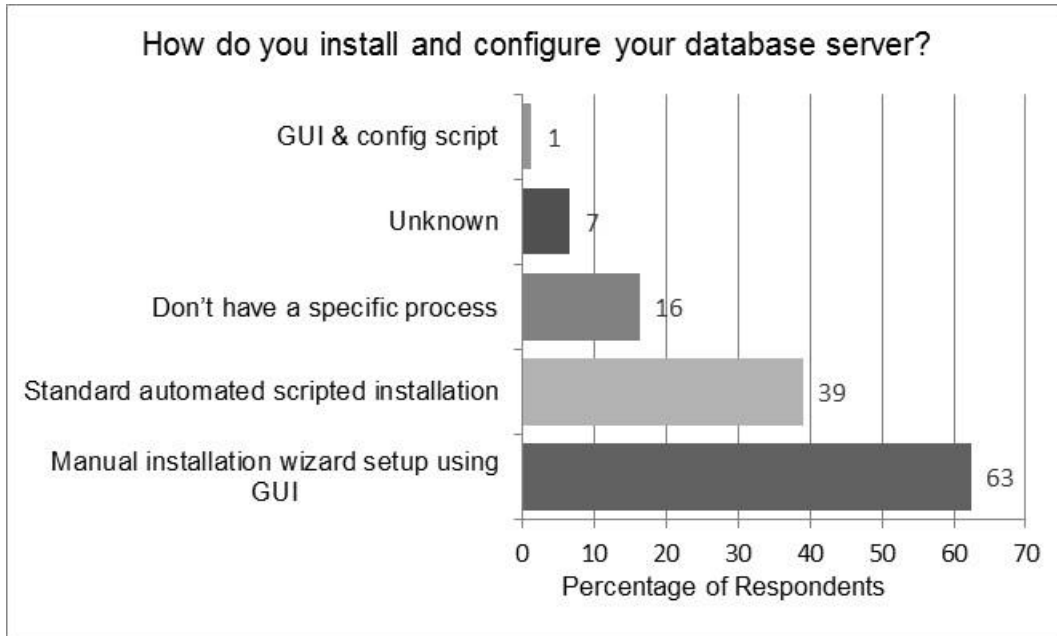


Figure 4.31 Installation and configuration

**Note: percentages do not total 100% because respondents could check all that apply; responses under 1% are not included.**

**Q26 N=320**

Security policies were shared by the respondents as being enforced (Figure 4.32).

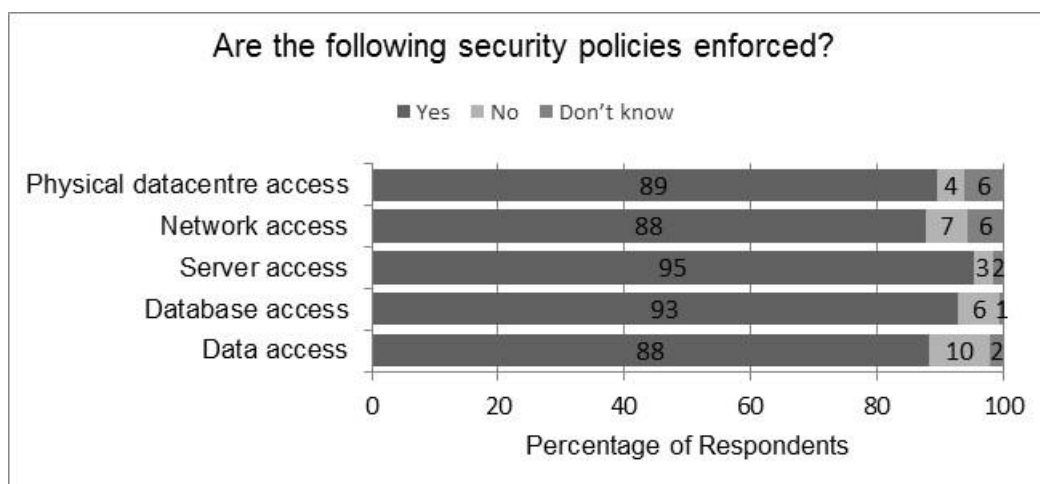
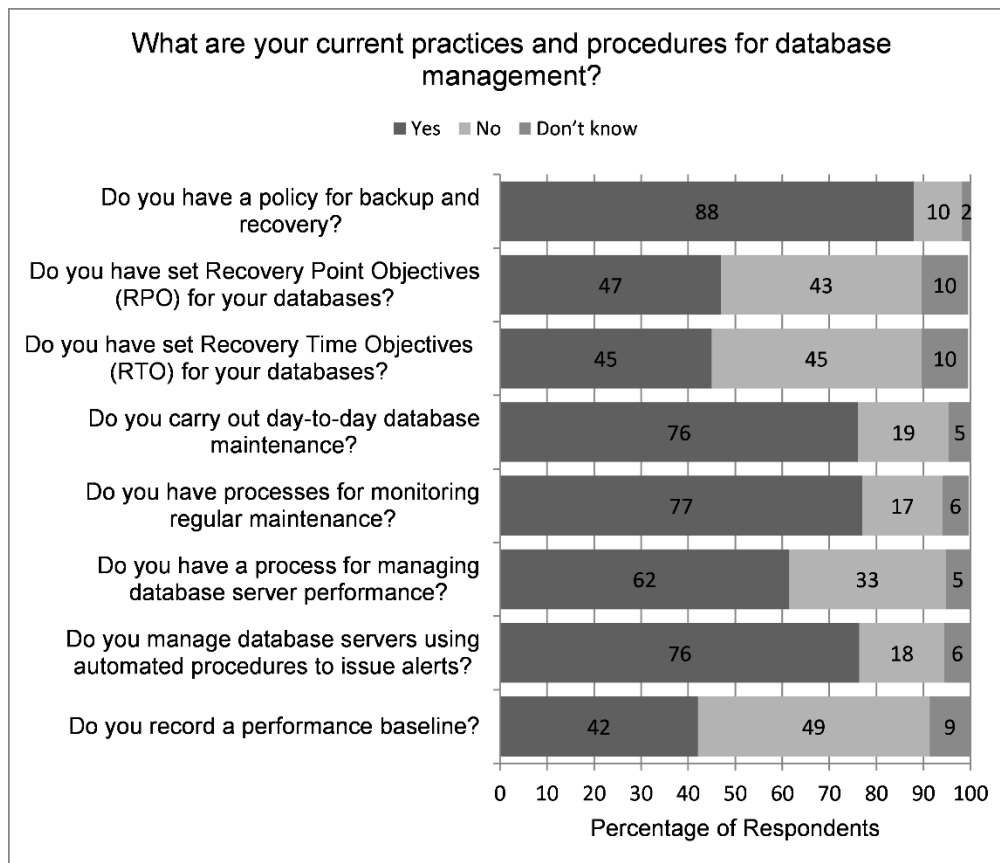


Figure 4.32 Security policies enforced

**Q32 N=295**

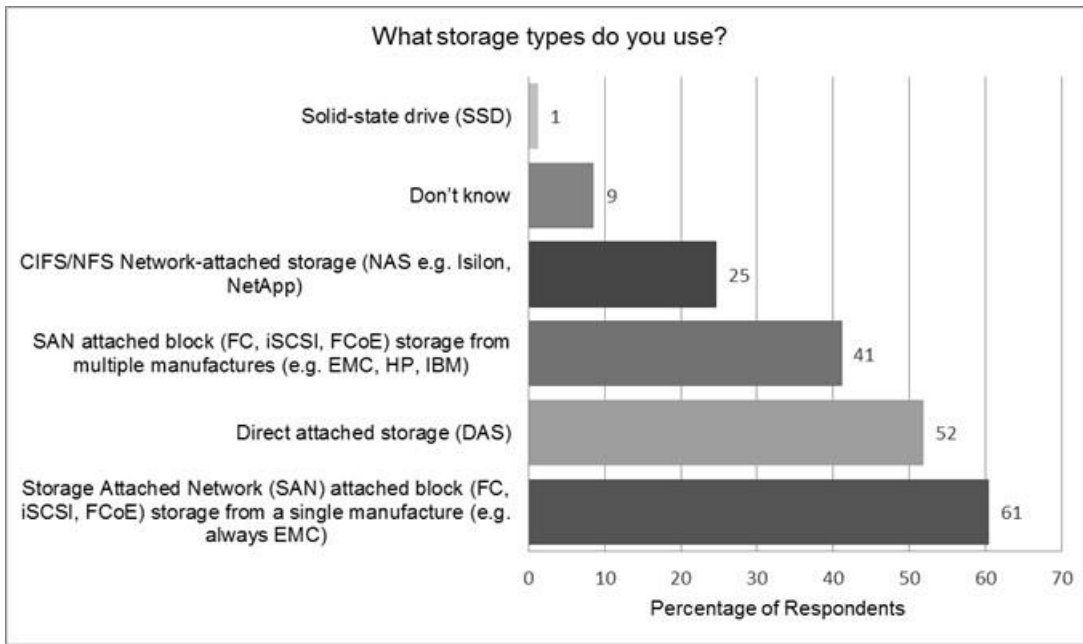
Existing practices and procedures for database management were undertaken for day to day database maintenance, in particular: processes for monitoring regular maintenance, automated procedures to issue alerts, and processes for managing database server performance (Figure 4.33).

The area reported that had the least respondents was the recording of a performance baseline, undertaken by only 42% of respondents. 45% of respondents did not have recovery time objectives (RTO), the amount of data loss allowable, and 43% respondents did not have recovery point objectives (RPO), the time it takes to restore the data.



**Figure 4.33 Practices and procedures for database management**  
**Q27 N=324**

There are many types of storage used (Figure 4.34).

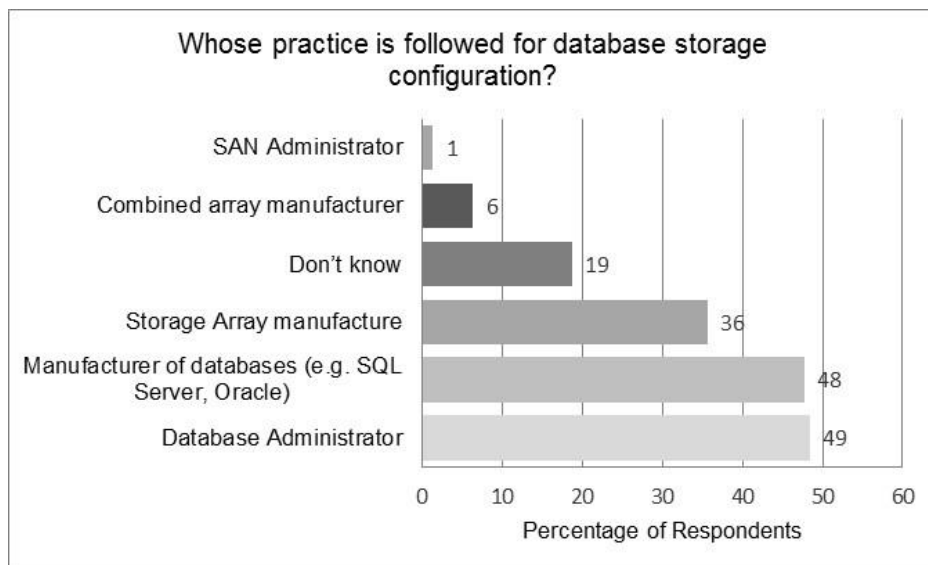


**Figure 4.34 Storage types used**

**Note: percentages do not total 100% because respondents could check all that apply; responses under 1% are not included**

**Q52 N=243**

Database storage configuration practices were followed from three main sources: the storage array manufacturers, database manufacturers, and database administrators (Figure 4.35).

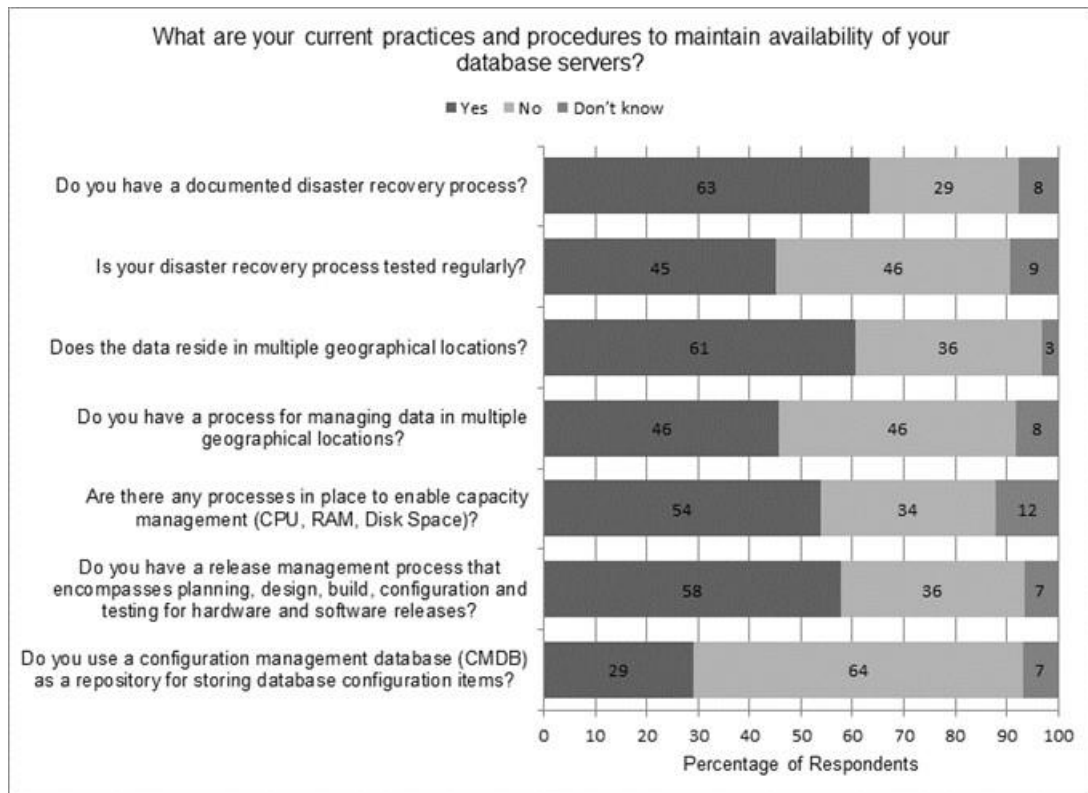


**Figure 4.35 Practice followed for database storage configuration**

**Responses under 1% are not included**

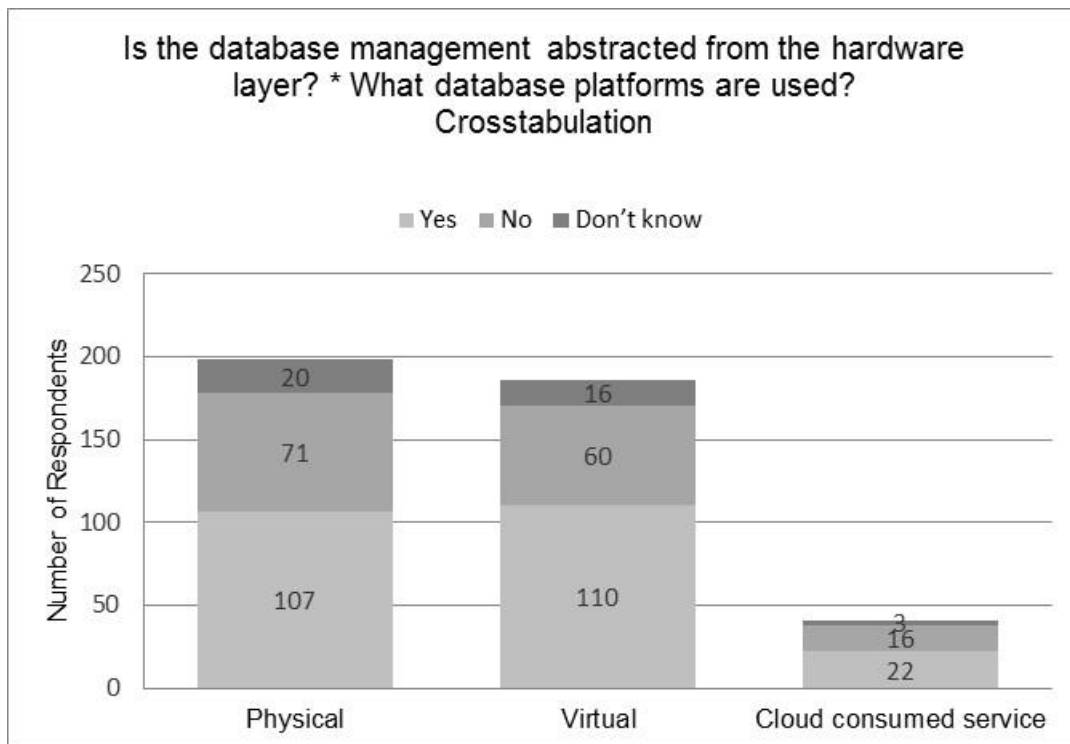
**Q56 N=239**

61% of respondents have data stored on multiple geographical locations, 54% have processes to manage capacity, and release management processes existed for 58% of respondents (Figure 4.36). 29% of respondents use a configuration database which is mostly used to help manage server configuration, often database licensing and change management systems for the server estates.



**Figure 4.36 Practices and procedures for availability**  
**Q31 N=324**

For just of half of respondents, database management is abstracted away from the type of platform (Figure 4.37). Some database management configuration requires specific hardware settings to run optimally.

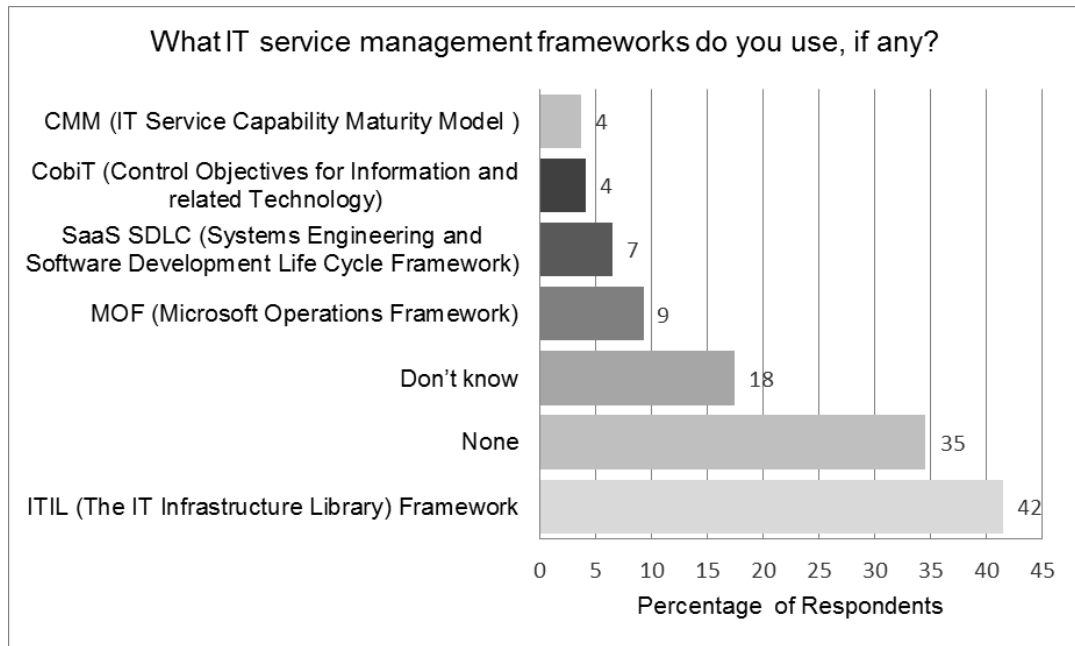


**Figure 4.37 For the platforms used is database management abstracted for the hardware layer?**  
**Q71 \* Q23 Cross tabulation N=326\*N=223**

#### 4.2.10 Database Operations

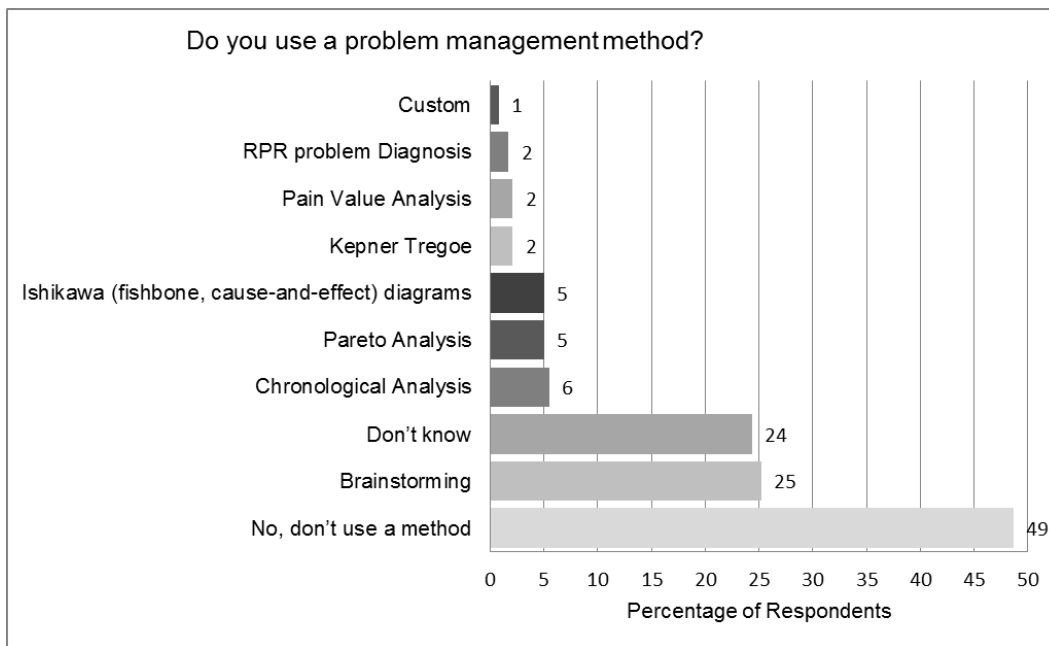
The operational management of database systems is required for production database systems. It is sometimes connected to core IT industry service management frameworks. Working team practices which derive from organizational culture could be reactive or proactive and sometimes problems required resolutions. Service Management frameworks are used to help improve management of IT systems.

In terms of the use of IT service management frameworks: 42% of respondents used the IT Infrastructure Library (ITIL) framework, while 35% of respondents used no framework (Figure 4.38).



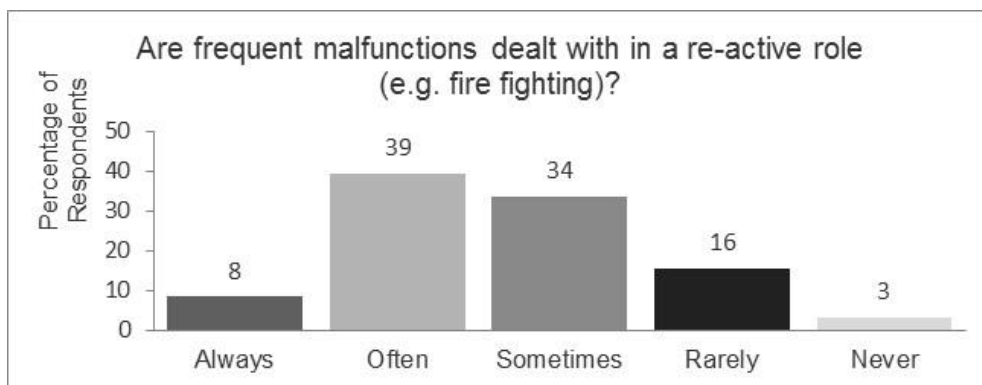
**Figure 4.38 IT Service management framework**  
**Responses under 1% are not included**  
**Q47 N=246**

Problem management methods are not widespread: almost half of respondents do not use one, while a further quarter did not know whether they did (Figure 4.39). Problem management is used to identify recurring issues and to enable the root cause of incidents to be resolved.



**Figure 4.39 Problem management method**  
**Responses under 1% are not included**  
**Q49 N=238**

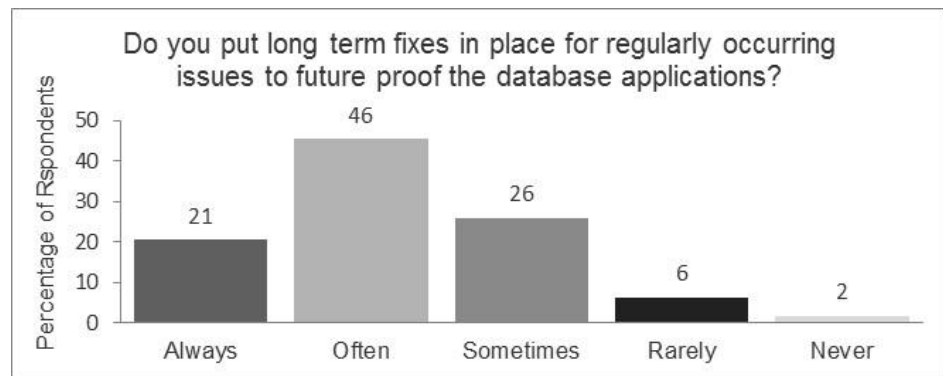
In reporting responses to malfunctions, almost half the respondents reported that they were usually dealt with re-actively (Figure 4.40).



**Figure 4.40 Frequent malfunctions**  
**What are your working team practices? Q66 N=226**

However, 66% of respondents (Figure 4.41) always or often put long-term fixes in place for regularly occurring issues, to future proof the database application.



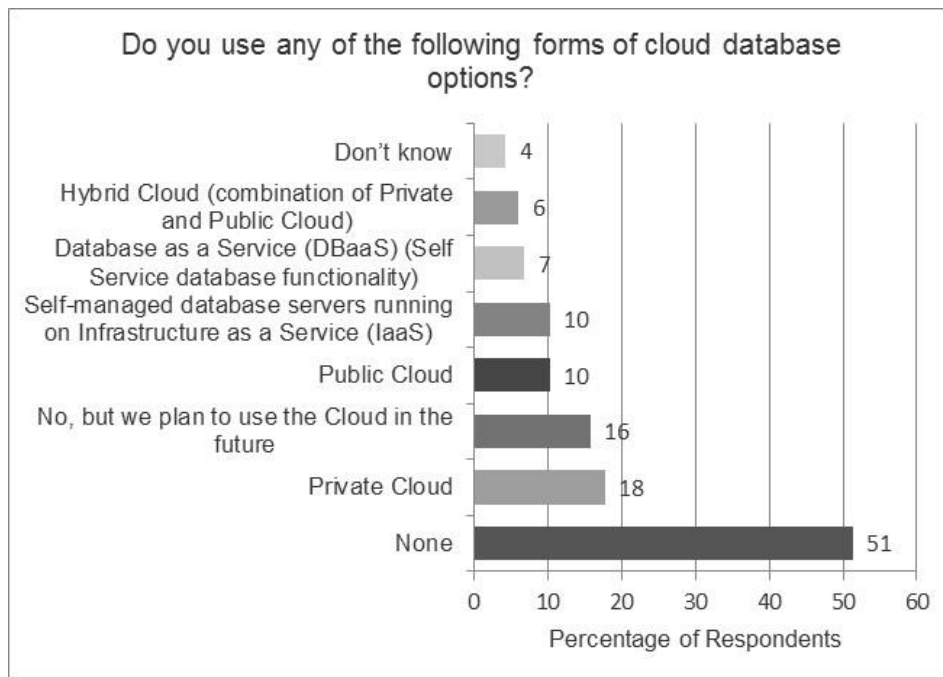


**Figure 4.41 Long term fixes for regular issues**  
*What are your working team practices? Q66 N=224*

### 4.2.11 Cloud Databases

Cloud databases are an alternative to traditional on-premises systems. Cloud databases have different usage patterns, ranging from IaaS (infrastructure as a service) plus self-managed database, to DBaaS (database as a service) where very little management was required. The different platforms might be split between on-premises and cloud which would mean that a particular service may have varying practices and procedures.

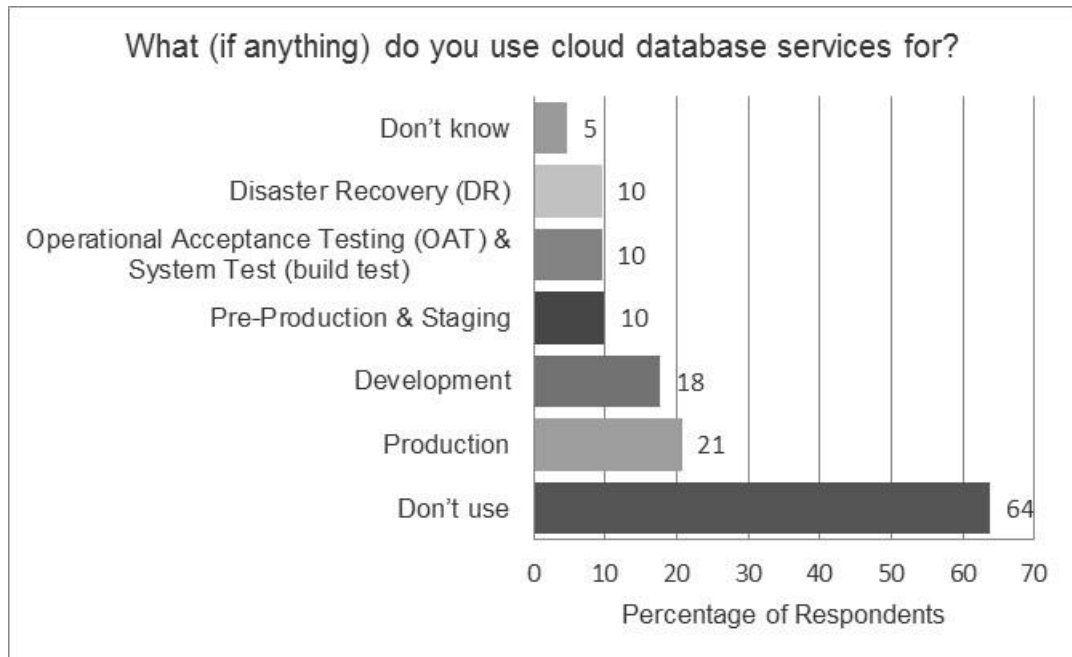
The options used show a proliferation of choices (Figure 4.42).



**Figure 4.42 Forms of cloud database usage**  
**Q57 N=236**

The reasons amongst the respondents against usage of cloud was the lack of trust of cloud vendors, not meeting security policies, lack of methods and lack of appropriateness for their specific uses. When asked if cloud database services were used, and the reason behind this, several respondents regarded cloud service as requiring less in-depth database management. One respondent, in Q82, commented: *“From the point of view of the client, the problem of database management, maintaining in-house skills etc. just goes away”*.

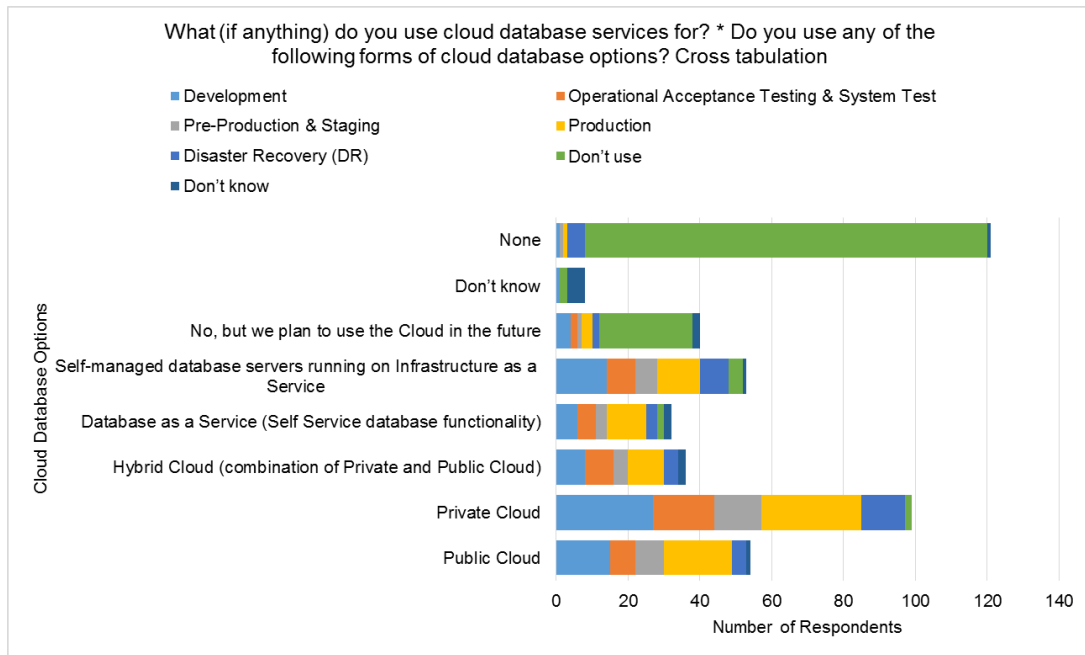
64% of respondents did not use Cloud database services (Figure 4.43). Cloud databases services were used for a variety of purposes.



**Figure 4.43 Cloud database service usage**  
**Q58 N=232**

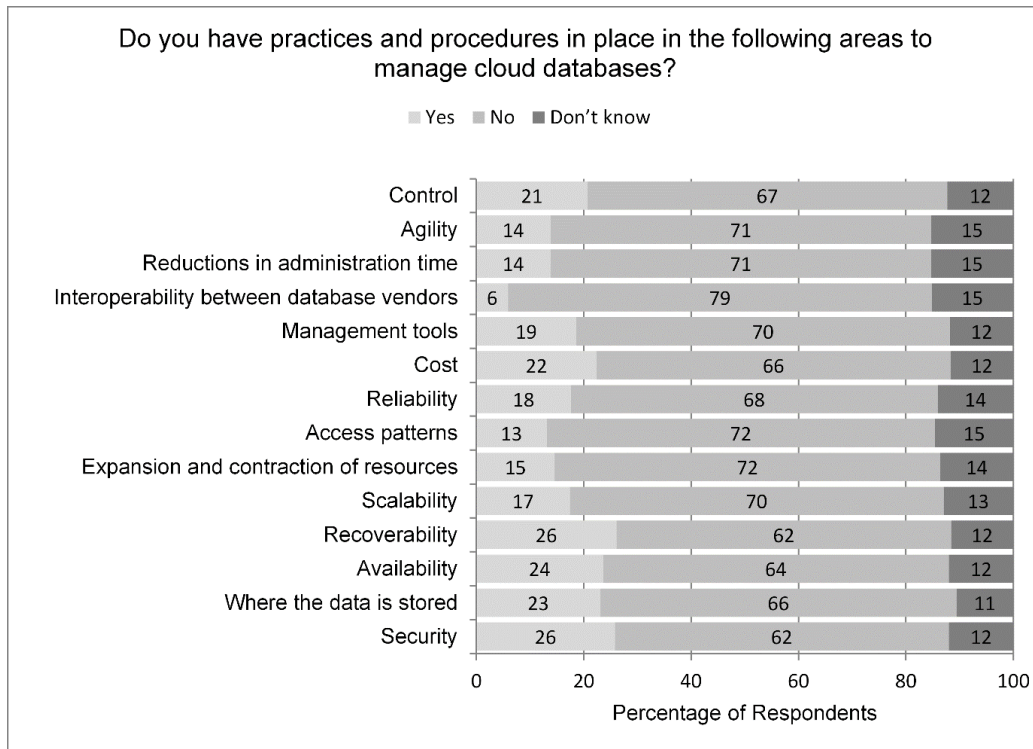
Figure 4.44 shows the cross tabulation (Q58 \*Q57) of the types of cloud database services (DBaaS, Hybrid, private and public cloud) used against the server environments (production, pre-production, test and development). Different environments may be chosen for different types of work and different amount of resources allocated for managing those environments. Private cloud was used for production services by 28 respondents, for public cloud by 19 respondents and for hybrid cloud by 10 respondents.

112 respondents didn't use particular environments, and reported that none of the forms of cloud database options were used. Production and development environments were used approximately the same number of times. Using dissimilar environments for development to production can lead to differing performance outcomes.



**Figure 4.44 The types of cloud database services used for the type of environment Q58 \* Q57 Cross tabulation N=232\*N=236**

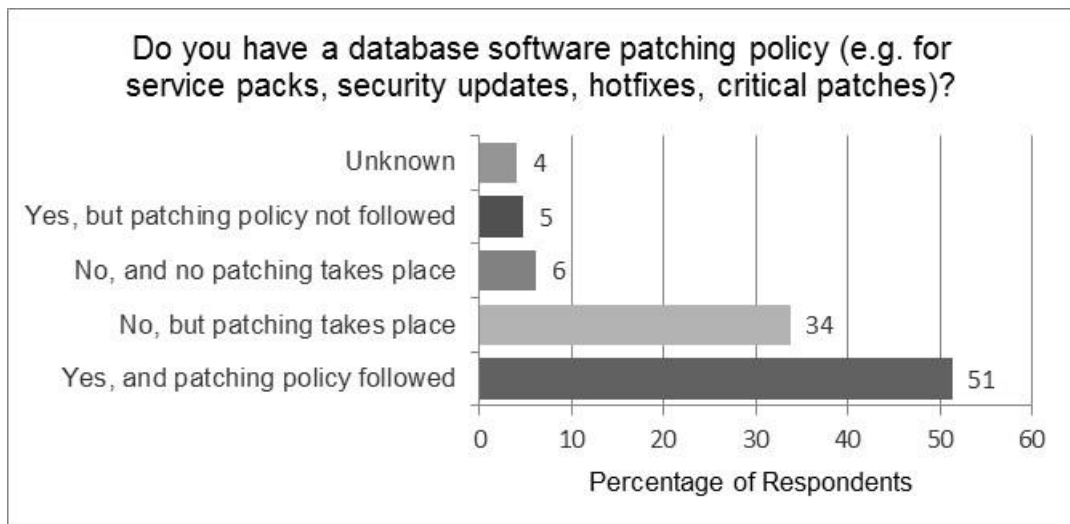
Very few practices and procedures were used by the respondents to manage any area of cloud databases (Figure 4.45).



**Figure 4.45 Practices and procedures to manage cloud**

**Q59 N=208**

A part of the security policy enforcement was through patching the servers for vulnerabilities. Database software patching policies were in place and followed by 51% of respondents (Figure 4.46) although 5% did not follow the patching policy. Some patching (34%) took place without policies being in place.

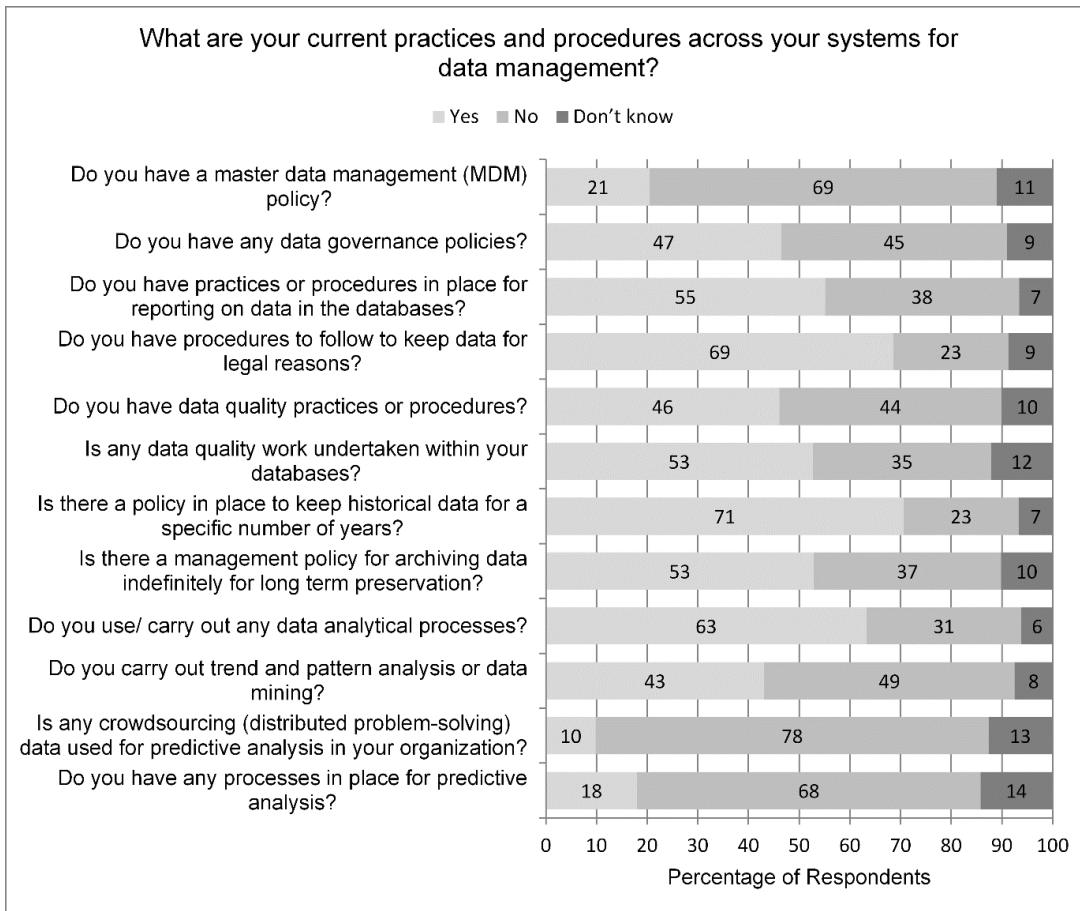


**Figure 4.46 Database software patching policy**  
**Q34 N=290**

#### 4.2.12 Data Management

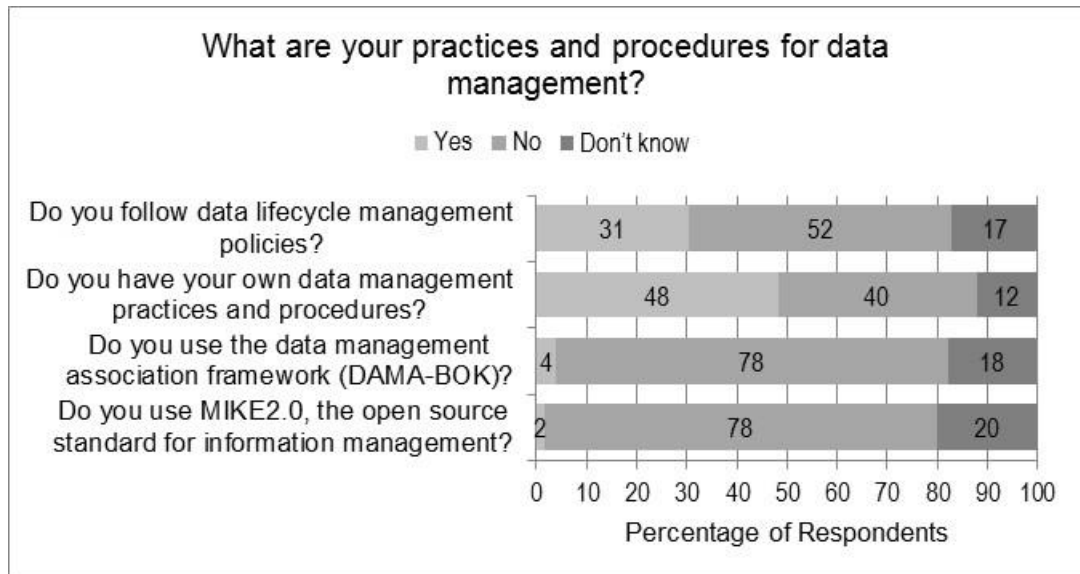
The data is at the heart of the database system. Figure 4.47 asked questions about the current practices connected to the main features of data management. The questions examined the governance, quality, reporting and analytics based on current and historic data.

Policies were in place in the majority of organizations for keeping data for legal reasons, historical data storage and long term preservation (Figure 4.47). For the majority of respondents (78%), crowdsourcing was not used for predictive analysis, 69% of respondents had no master data management policy, 68% of respondents did not have processes in place for predictive analysis and 45% of respondents did not have data governance policies.



**Figure 4.47 Current practices and procedures for data management**  
**Q42 N=257**

Figure 4.48 shows that 52% of respondents did not follow data lifecycle management policies. 78% of respondents followed neither the data management association framework (DAMA-BOK) nor the open source MIKE2.0 standard. However 48% of respondents stated they had their own data management practices and procedures.



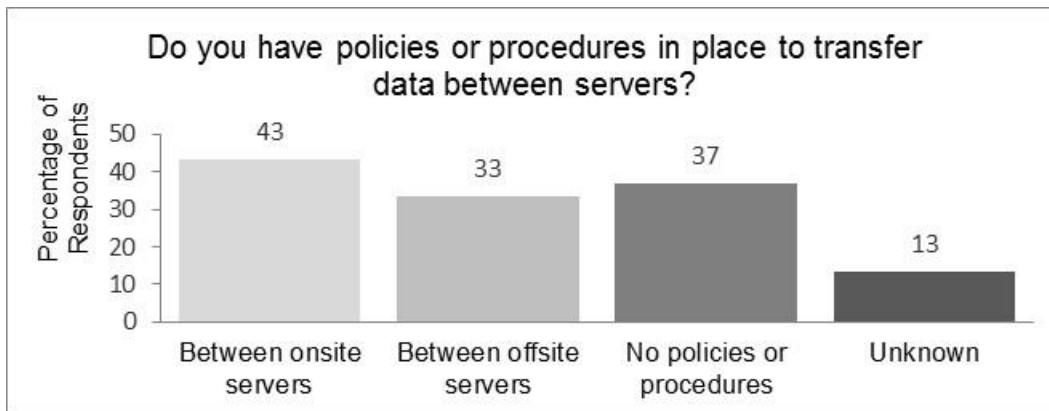
**Figure 4.48 Practice and procedures for data management**  
**Q48 N=246**

Data requirements were thought to be driving database management procedures in all cases by 38% of respondents and sometimes for 33% of respondents (Figure 4.49).



**Figure 4.49 Data requirements driving database management**  
**Q45 N=257**

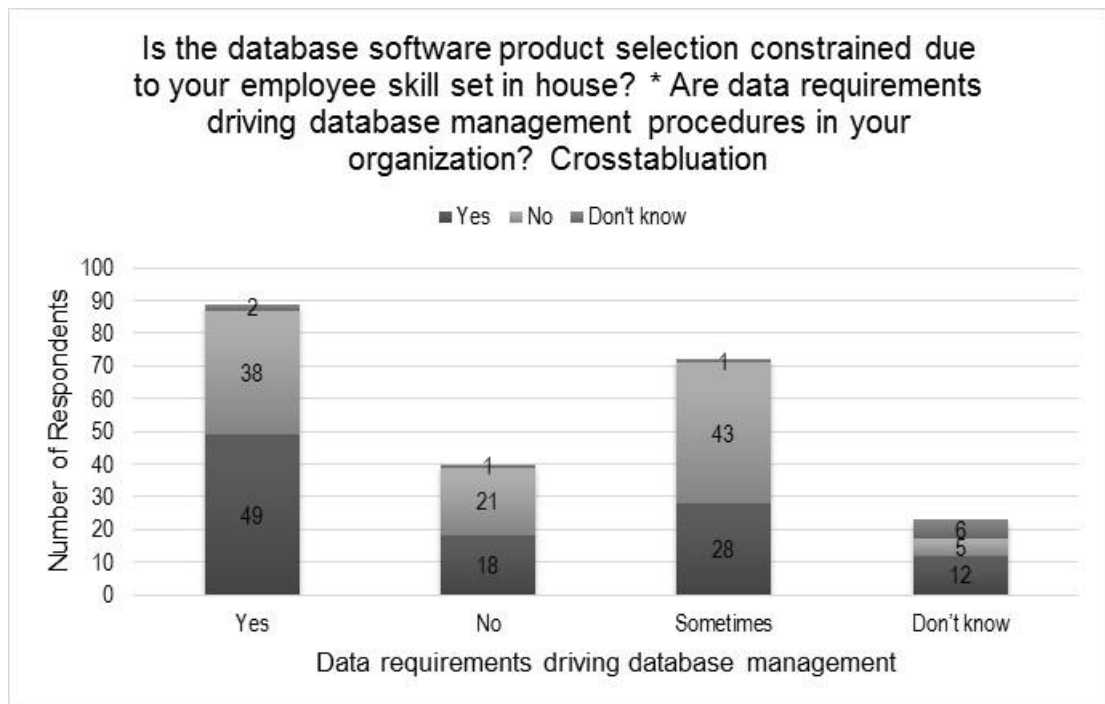
Data transfer policies have not been developed in all cases: 43% of respondents had policies in place to transfer data on site, 33% had policies for offsite data transfer, and 37% had no policies (Figure 4.50). Data transfer policies are important for maintaining security of data. If there are no policies, this could mean that data is no longer secure, and it could be moved or shared anywhere.



**Figure 4.50 Transfer data between servers**  
**Q36 N=293**

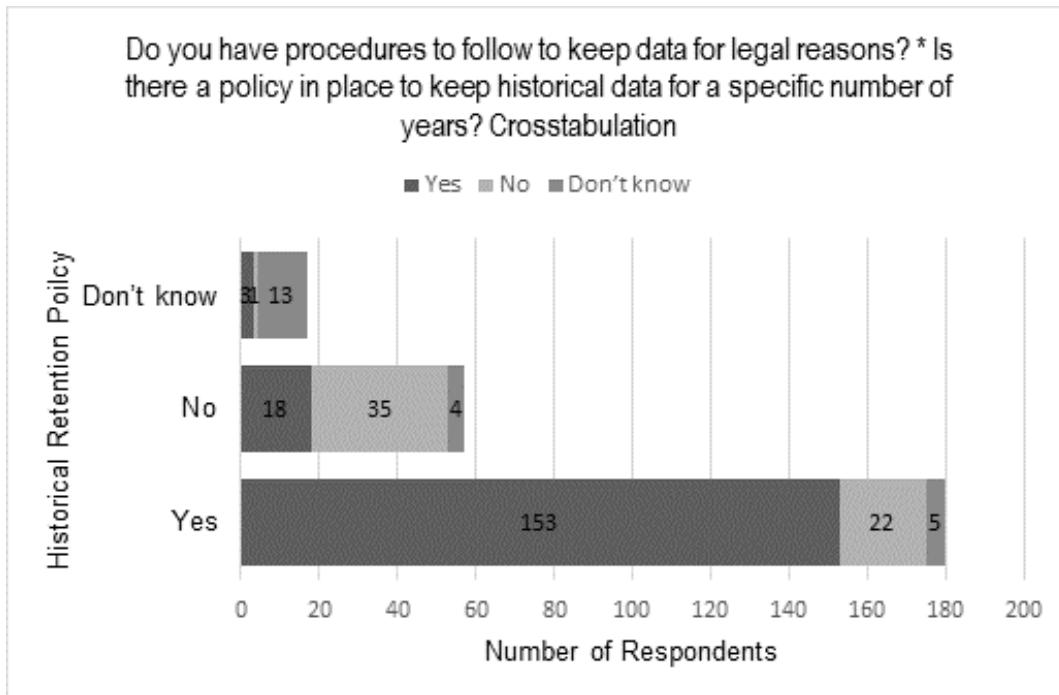
Figure 4.51 cross tabulation identified where the product selection was constrained due to the employee skill set, and data requirements were driving database management. 49 respondents agreed data management was driving database management procedures and database software product selection was constrained by the skill set of in-house employees, although 43 thought sometimes that did not. The production selection determines what functionality is available. Where data management drives the procedures in the organization, the product in use might not be the most suitable for the job, but it is the one in which the employees have skills.





**Figure 4.51 Database product selection constrained by employee skillset for data requirements, driving database management**  
**Q68 \* Q45 Cross tabulation N=224\*N=257**

Cross tabulations showed that set procedures were followed for legal reasons for the great majority of cases of historical data (Figure 4.52). The storage of long term data for legal reasons can impose a considerable administrative burden.



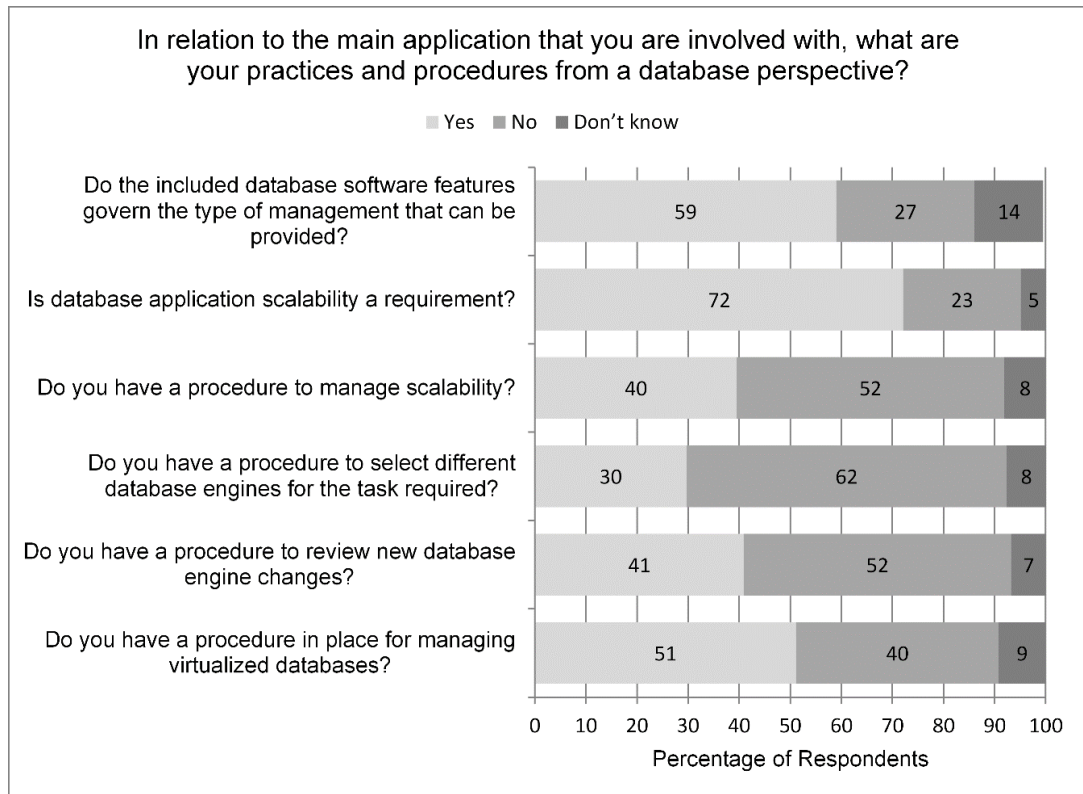
**Figure 4.52 Legal procedures followed for data with historical data policies**  
*Q42 part 4 \* Q42 part 7 Cross tabulation N=255\*N=255*

### 4.2.13 Application Centric

Application centric was a term used for looking from the inside to the outside of the database system, with the database system driving the configuration and management. Database applications provided a plethora of features to meet the demands of business. Tools were required to manage these features and to control and manage fundamental parts of the database systems.

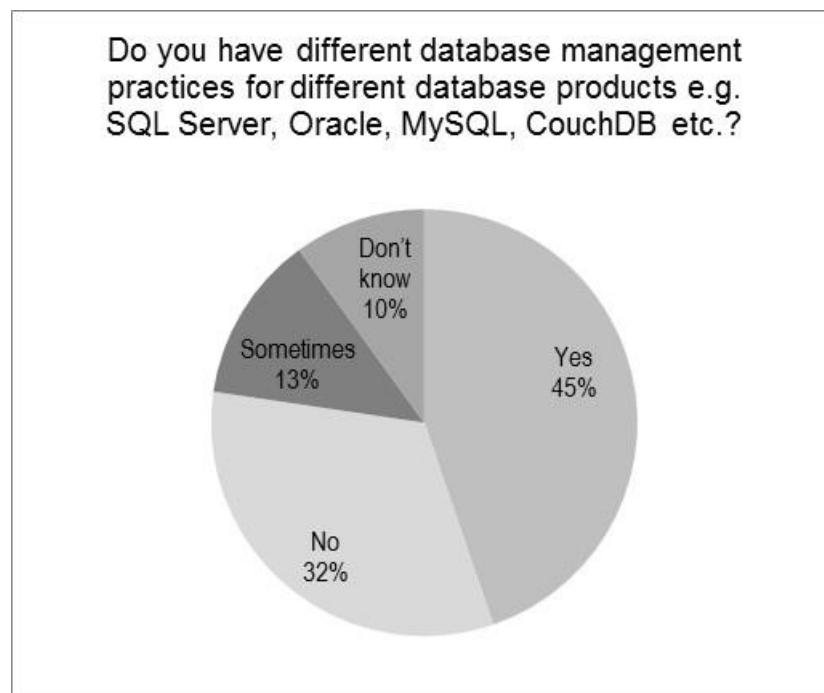
59% of respondents stated that the type of database management that can be carried out was governed by the database software features (Figure 4.53). 72% said database application scalability was a requirement, although 52% of the respondents did not have procedures to manage scalability.

62% did not have procedures to select different database engines for the task. 52% did not have procedures for reviewing new database engine changes, while 51% had procedures in place for managing virtualized databases.



**Figure 4.53 Practices and procedures for the main application**  
**Q71 N=223**

45% of respondents (Figure 4.54) had different management practices for different database products used. Although the output from different database products often had the same output, the means of achieving the goal may vary.

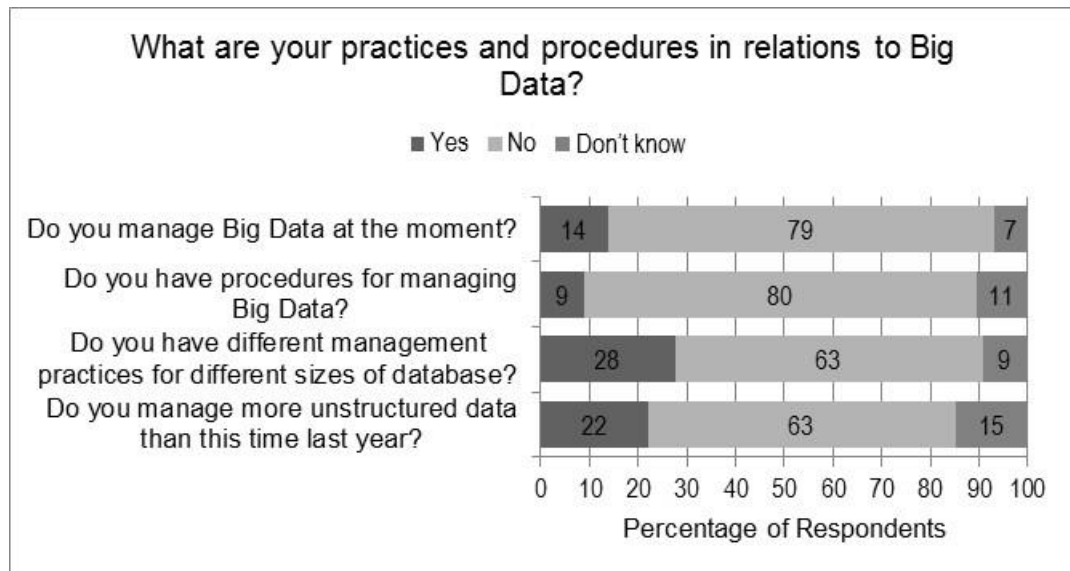


**Figure 4.54 Different management practices for different database products**  
**Q72 N=221**

The size of data stored in a database often resulted in a change in management techniques, so large and smaller databases were handled differently. There were a large selection of database products used by the respondents (Figure 4.15) so it was important to know whether any database management practices were different. 63% of the respondents were not managing more unstructured data than last year, although this compared to 22% who were managing more unstructured data (Figure 4.55).

There were very few database administration practices and procedures for managing 'Big Data' (Figure 4.55). From Figure 4.13 only 1.3% of the respondents had a database over 100TB in size. The question defined Big Data as a general term used to describe the large volume of unstructured and semi-structured data that cannot be processed using conventional methods.

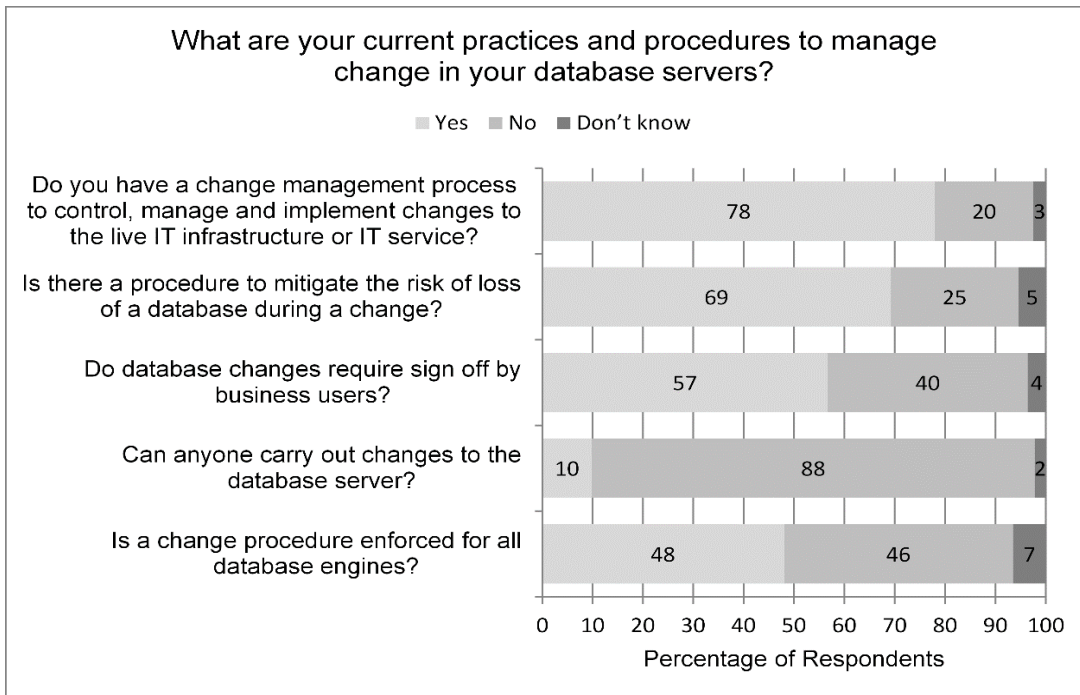
79% of respondents did not manage Big Data (Figure 4.55). 81% did not have any procedures for the management of Big Data. 63% did not have different management practices for different sizes of database (Figure 4.55).



**Figure 4.55 Practices and procedures for big data**  
Q73 N=223

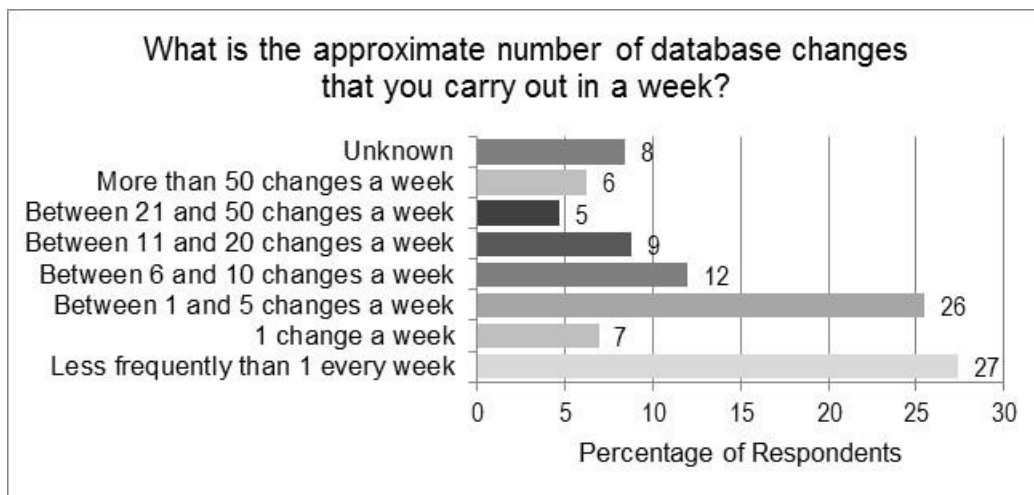
#### 4.2.14 Change Management

Most respondents reported having practices and procedures to manage changes for database servers. 57% reported that database changes required sign off (agreement) by business users. 88% reported that changes to the database server could not be carried out by just anyone. 48% of respondents reported that change procedures were enforced for all database engines, while 46% did not enforce such procedures (Figure 4.56).



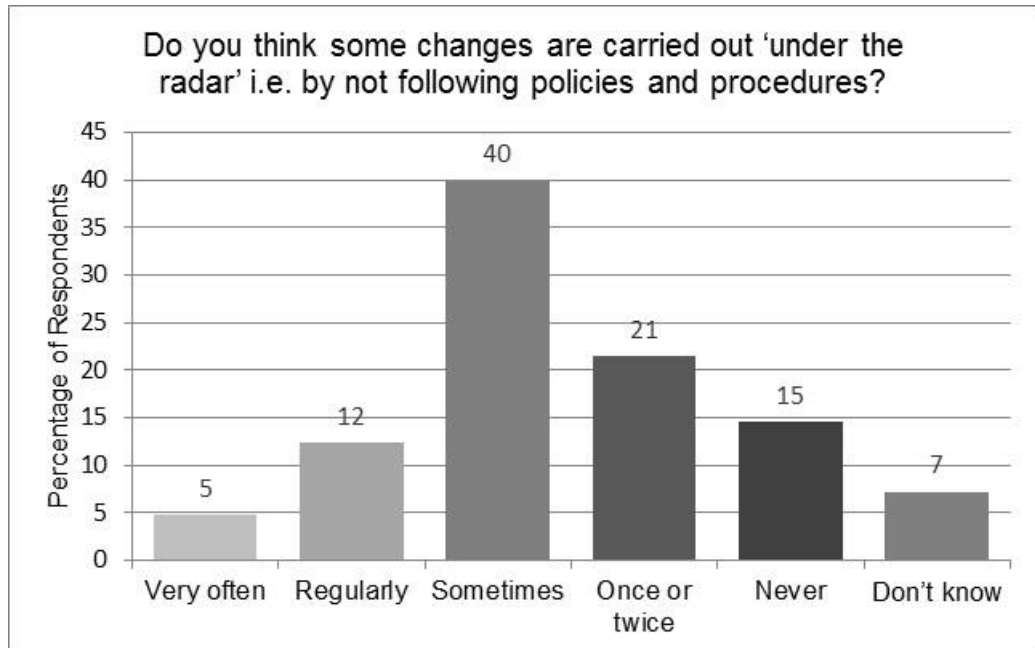
**Figure 4.56 Practices and procedures for change management**  
Q38 N=277

Regular changes were made to the database environments in respondents' organizations (Figure 4.57). 27% of respondents said changes were carried out less often than weekly, while 6% made more than 50 changes per week.



**Figure 4.57 Approximate database changes a week**  
Q39 N=274

Formal change processes were not always used. 40% of respondents reported that sometimes changes occurred without following policies and procedures, while 5% reported that this happened very often (Figure 4.58).

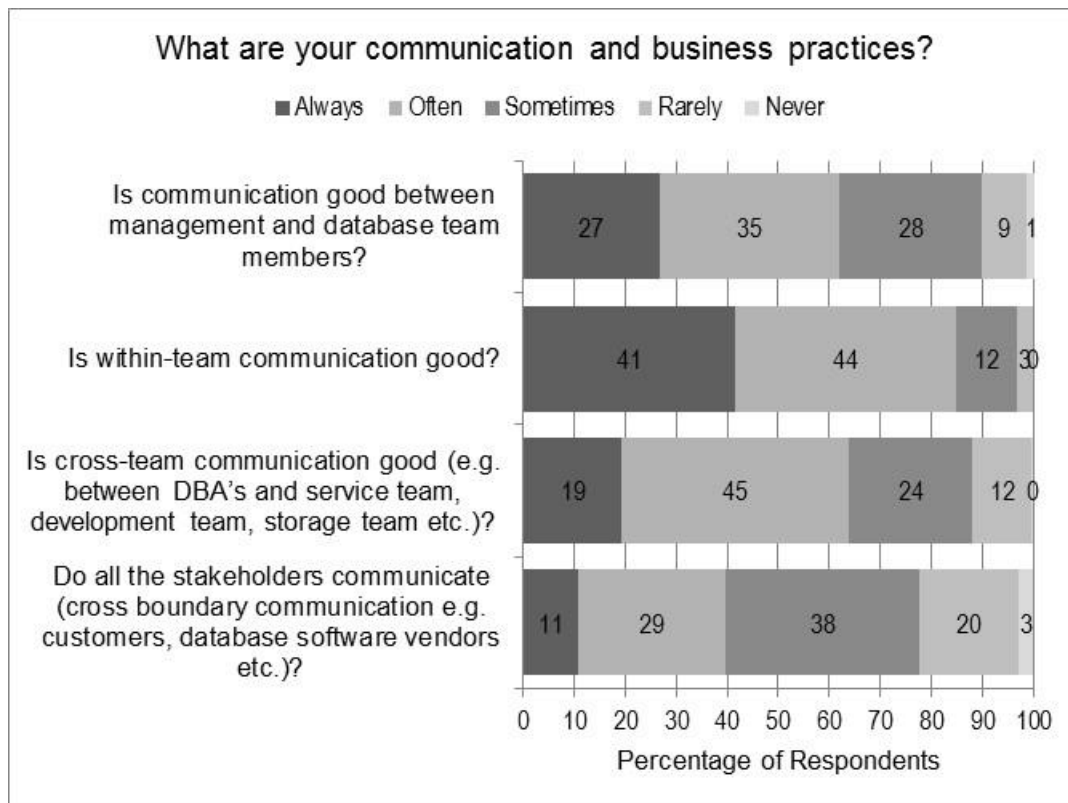


**Figure 4.58 Changes not following policies and procedures**  
Q40 N=276

#### 4.2.15 Organizational Culture

The working team practices of the respondents gave insight into the working conditions, communication, control, strategy and budget within the organizations (Figure 4.59).

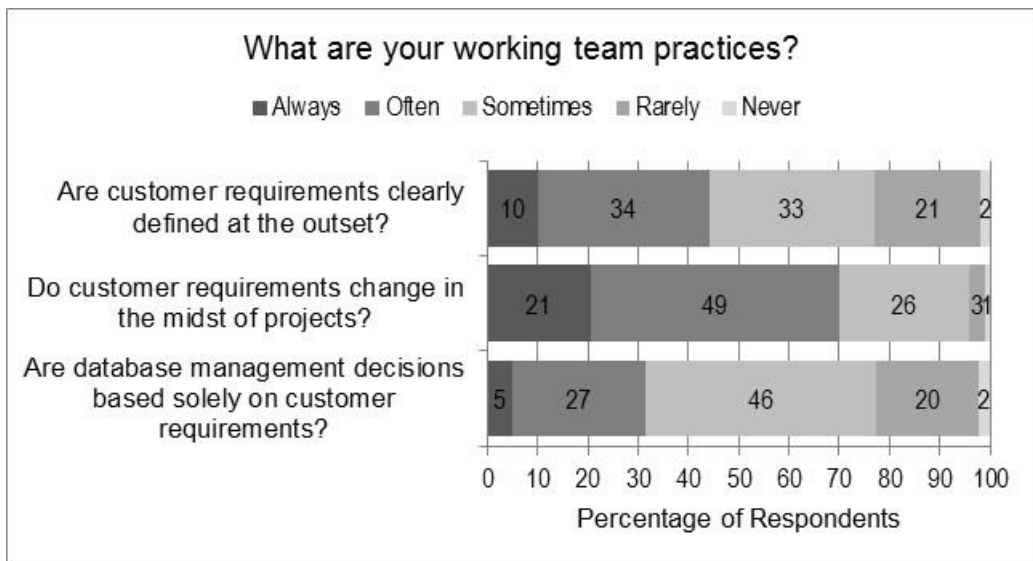
Communication between management and database team members as well as cross team communication was often or always good. Within-team communication was seen as always good for 41% of respondents, whereas only 11% of respondents stated cross-boundary communication was always good.



**Figure 4.59 Communication and business practices**  
**Q67 N=228**

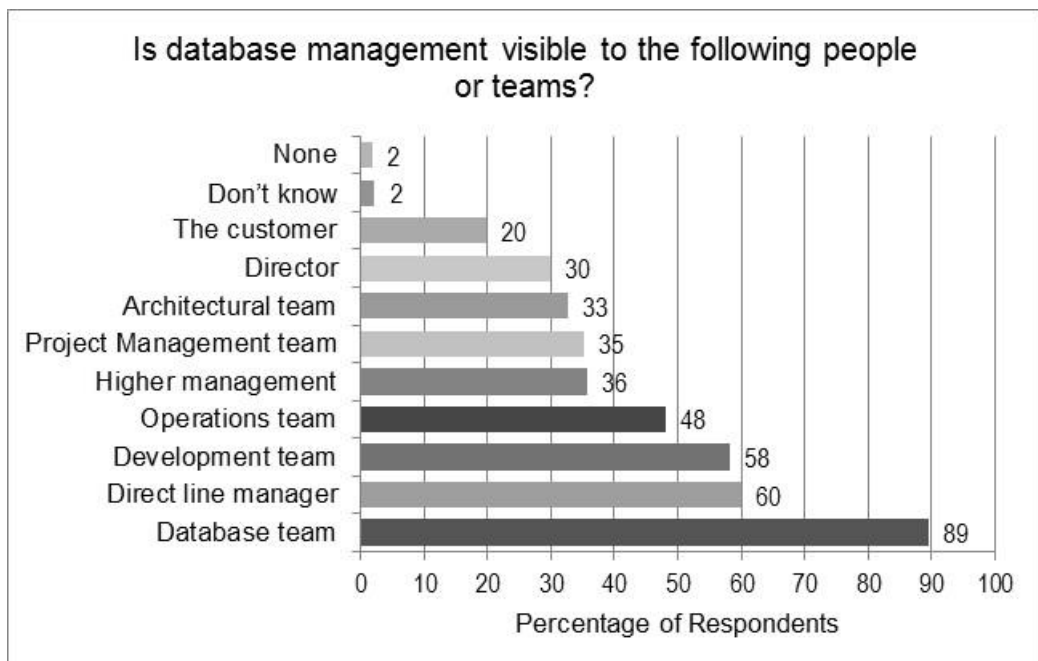
5% of respondents stated that database management decisions were based solely on customer requirements (Figure 4.60). Nearly 50% of customer requirements often changed in projects and only 10% of respondents said customer requirements were always clearly identified at the outset.





**Figure 4.60 Working team practices**  
Q66 N=228

Database Management was clearly visible in the database team. For the direct line manager it was 30% less visible and when it reaches the director the respondents answered there was a 60% reduction in visibility (Figure 4.61).



**Figure 4.61 Database management visibility**

**Note: percentages do not total 100% because respondents could check all that apply**  
Q69 N=227

Whether or not the database software product selection was constrained due to the employee in house skill set was uncertain: for 48% of respondents it was a constraint, while an equal number said it was not (Figure 4.62). 55% of respondents stated the budget didn't determine what database platform was used, although 57% of respondents stated financial reasons influenced the version of the database software.

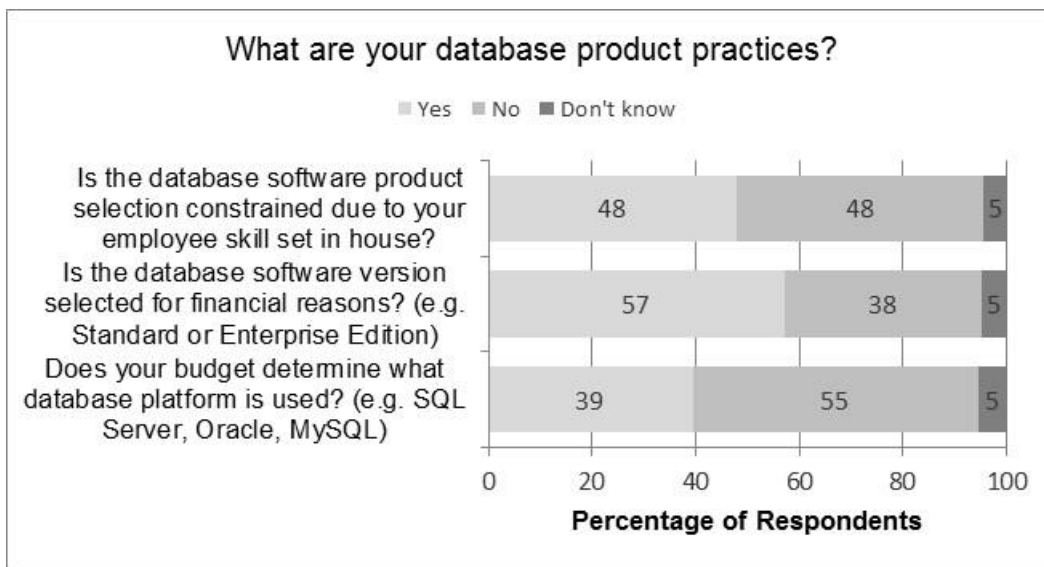
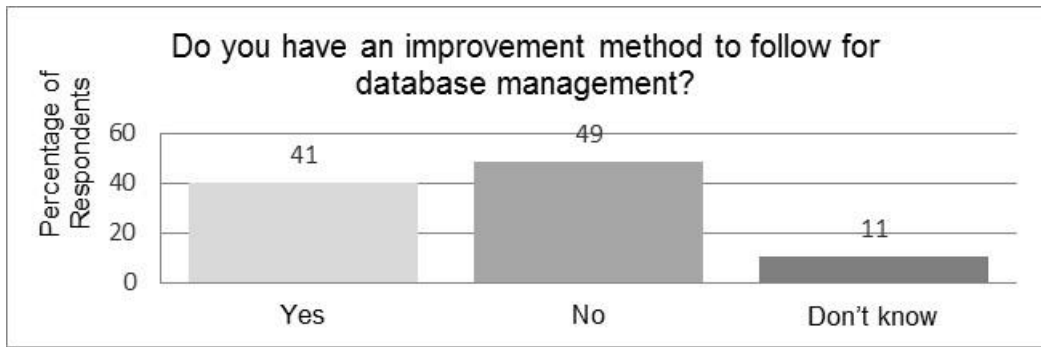


Figure 4.62 Database product practices  
Q68 N=224

#### 4.2.16 Improvement Methods

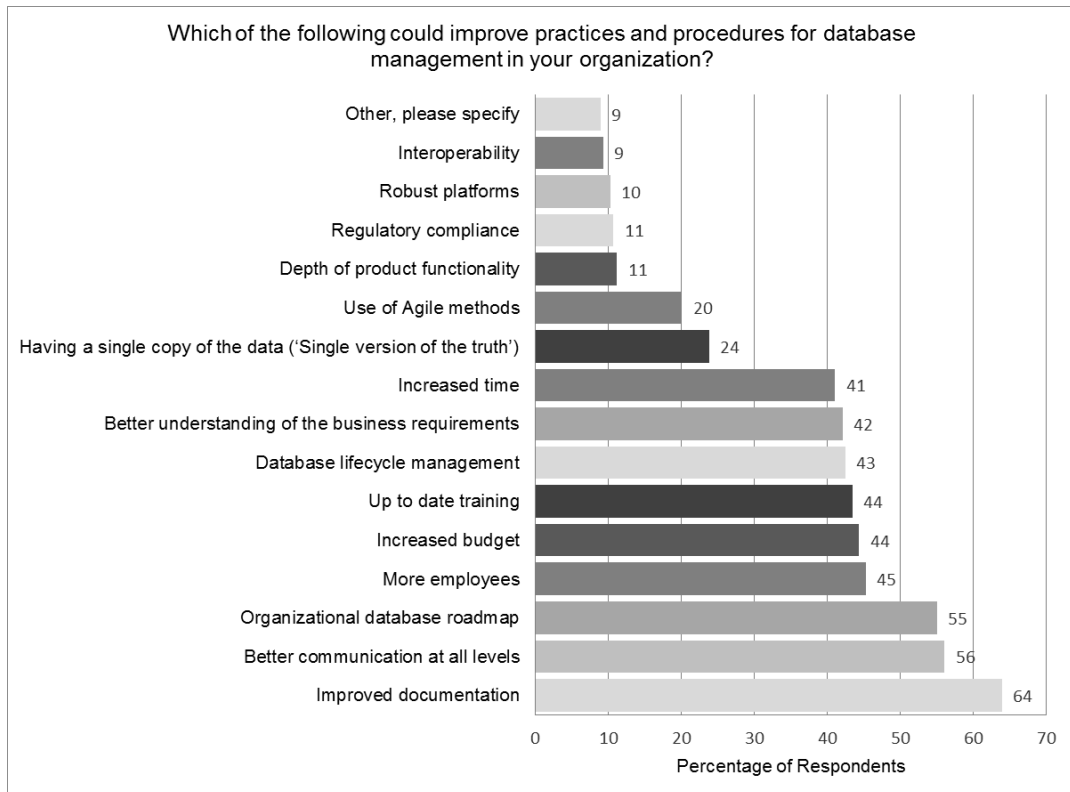
Two quantitative survey questions specifically asked about improvement methods and whether any of the suggested list would help with improvement.

Improvement methods (Figure 4.63) are used in under half of the respondent's cases.



**Figure 4.63 Do you have an improvement method to follow for database management**

**Q71 N=223**



**Figure 4.64 Improvement method to follow for database management**

**Note: Percentages do not total 100% because respondents could check all that apply**

**Q80 N=214**

There were various items which could help improve practices and procedures in the organization. Improved documentation (Figure 4.64) was suggested by 64% of respondents as the biggest improvement. Better communication and having an organizational roadmap were the next highest suggested by the respondents.

Other comments added to Q80 (which had a free-text section) included:

*“Training, standardization or processes. Greater sensitivity of coders to existing indexes”*

*“Employee turnover”*

*“Formal policies”*

*“More visibility with other silos – I only see production, not what goes on in dev, test etc.”*

*“All the various groups understanding WHY it's important”*

*“Consultation with Subject Matter Experts”*

*“More rigorous testing”*

*“There is no such thing as a ‘single version of the truth’”*

#### **4.2.17 Future Vision**

A question was asked in the survey to determine whether the respondents had a view of the future of database management. The most frequently occurring words in the survey are shown in Figure 4.65.



**Figure 4.65 Business vision of database management**

**Q81**

The highest occurring words (those with above 5 occurrences) were: data (46); database(s) (28); management (21); cloud (21); automated (8); structured (6); business (6); solutions (6); services (6); storage (7); think (7); servers (5); time (6); different (6) and systems (6).

The main areas highlighted following thematic analysis were automation, change, cloud, data, NoSQL, management, people, development and technical.

### 4.3 Connecting Quantitative and Qualitative phases

This section provides the link from the quantitative (wide and shallow) approach to the qualitative (narrow and deep) data collection, to enable the information that has been captured to be explained, with the help of stories about managing database systems. Whereas the quantitative research sought to answer the first research question, the qualitative research sought to answer the second and third research

questions. In order to answer these questions in the focus groups, some semi-structured questions were required to lead the participants through the investigative discussions. The questions for the next stage of the research were derived from the answers received to various questions in the quantitative research.

The data from the initial quantitative survey indicated that the successful management of database systems required a knowledge of the complex interactions between the technical components and the actors. The following quantitative data results led to the in depth investigation of the complex interactions which revealed certain information not previously considered to be of major importance.

Ten questions were chosen to lead the participants through discussions of matters arising from database operations. The ten questions were derived from the analysis of the first survey results. The initial approach, following on from best practice usage revealed by the quantitative results, gave a more in depth understanding of certain areas. These areas were important best practices, selection of database engines, requirements gathering, database lifecycle management, technical layers, managing cloud databases , complexity compromising implementation, creation and control, cloud boundary communication and strategic planning.

In each of these areas a new qualitative research question was posed. The question order was considered to ensure the semi-structured focus groups followed a logical pattern of an introduction, a middle and an end. The reasoning behind why each new question was chosen is detailed below.

The ten qualitative questions were derived from the quantitative results. The explanation is given in each of the following tables.

### 4.3.1 Importance of Best Practices

Figure and Question Number	Quantitative Question	Interesting points to investigate further
Figure 4.33 -Q27	What are your current practices and procedures for database management?	The answers to these questions about practice and procedures for core database maintenance activities were mostly between 60% and 88%. Why were these processes and procedures high? Are these core activities more important than others?
Figure 4.36 -Q31	What are your current practices and procedures to maintain availability of your database servers?	Availability on the whole is between 45% and 63%. Is this more or less important than other management tasks. Is keeping the servers up and available important?
Figure 4.46 -Q34	Do you have a database software patching policy (e.g. for service packs, security updates, hotfixes, critical patches)?	56% patch database software, which included security patches. How important do people think procedures are for protecting the data?
Figure 4.47 -Q42	What are your current practices and procedures across your systems for data management? :-  Do you have procedures to follow to keep data for legal reasons?  Is there a policy in place to keep historical data for a specific number of years?	There are legal reasons (69%) and historical reasons (71%) which have a high number of policies and procedures associated to them. Where do these fit into the management?
Figure 4.35 -Q56	Whose practice is followed for database storage configuration?	Why was practice followed for storage configuration from different people?
Figure 4.10 -Q77	What issues can occur with following best practice? :  Do you think it is important to have best practices?	Best practice was reported as being important by 94% of respondents. It would be interesting to know more about why this was important.
Figure 4.48 -Q48	What are your practices and procedures for data management?:  Do you have your own data management practices and procedures?	48% of respondents have their own data management practices and procedures. This would be a factor worth discussing further for managing databases.

There seemed to be many different practices and procedures within development, database operations and the architecture.

The use of data management practice and procedures were varied across most sectors of industry. A qualitative question was derived to seek a better understanding of the choice of practices and procedures for managing database systems:

*Question 1: Do you think some best practices and procedures are more important than others for managing database systems? If so, what are the most important ones?*

### 4.3.2 Selection of Database Engines

Figure and Question Number	Quantitative Question	Interesting points to investigate further
Figure 4.15 -Q11	What database applications do you use?	51 different database applications were used of all types. Why this variety of usage?
Figure 4.54 -Q72	Do you have different database management practices for different database products e.g. SQL Server, Oracle, MySQL, CouchDB etc.?	Interesting that only 45% of respondents had different practices. Is there a core set of management practices that are shared for different products?
Figure 4.14 -Q12	What type of database engine do you use?	98% of respondents used relational databases. The type of engines is interesting. Does this mean practices and procedures were similar or the same? What was the choice factor?
Figure 4.53 -Q71	In relation to the main application that you are involved with, what are your practices and procedures from a database perspective?  Do you have a procedure to select different database engines for the task required?	62% of respondents don't have a procedure to select different database engines and 52% don't have procedures to review new engines. What did this mean for selection?  59% of respondents mentioned software features governed management. Did this affect selection of different database engines?



	<p>Do you have a procedure to review new database engine changes?</p> <p>Do these included database software features govern the type of management that can be provided?</p>	
--	---	--

Many different database engines were selected, used and managed. The qualitative question below was asked to clarify the basis of selection:

*Question 2: What best practices and procedures do you think should be considered when selecting different database engines?*

### 4.3.3 Requirements Gathering and Design

Figure and Question Number	Quantitative Question	Interesting points to investigate further
Figure 4.24 -Q18	Do you use any of these architecture frameworks for database design?	60% of respondents did not use common industry standard architecture frameworks. What methods were used?
Figure 4.25 -Q19	Does your organization use the following processes at the architecture stage?	Only 47% of respondents had processes for requirements gathering although 58% documented the solution. Did respondents think processes were not an important part of architectural design?
Figure 4.26 -Q20	Are any processes followed at the design stage?	It was surprising how low the processes followed are for design. 55-66% of respondents reported for database design for OLTP and DW were followed which were high in comparison to the other options stated. Hardware and data structure 36-38%. Why?

It appears not all respondents used processes for requirements gathering or architectural reference. The following qualitative question was asked:

*Question 3: What kind of requirements gathering and architectural design processes for the hardware, data and databases do you think are important? Why are these important?*

### 4.3.4 Database Lifecycle Management

Figure and Question Number	Quantitative Question	Interesting points to investigate further
Figure 4.64 -Q80	Which of the following could improve practices and procedures for database management in your organization?	Database lifecycle management was suggested by 43% of respondents to improve practices and procedures for database management. Why did only 43% of respondents suggest database lifecycle management would improve practices and procedure for database management?
Figure 4.24 -Q18	Do you use any of these architecture frameworks for database design?	60% of respondents indicated no architectural frameworks were used for database design. The frameworks listed were prescriptive and may not have been for low level design. It would be interesting to delve into more depth to understand why no frameworks were used?
Figure 4.31 -Q26	How do you install and configure your database server?	Manual installation was carried out by 63% of respondents. Best practice using automation was not mainly used which seemed strange. Why was this?
Figure 4.28 -Q22	Are any processes followed at the development stage?	43% of respondents followed the database development cycle and had standard testing processes. Why was this lower than the source control and standard coding practice?
Figure 4.30 -Q28	How are the majority of database servers managed?	The majority of database servers are managed individually or through central tools with only 22% self-managing. Would best practice have assisted management?
Figure 4.56 -Q38	What are your current practices and procedures to manage change in your database servers?	Practices and procedures on the database servers were managed well under change. Why did this work and is there anything else that was required?
Figure 4.38 -Q47	What IT service management frameworks do you use, if any?	35% of respondents did not use any framework which was fairly high. Why was this?

Various identified practices and procedures were used in the database lifecycle.

The quantitative analysis, revealed a number of unanswered questions concerning lifecycle management which needed further investigation. The following qualitative question was derived to look deeper into the situation:

*Question 4: In what ways do you think that best practices and procedures could assist management of the database lifecycle?*

### 4.3.5 Technical Layers

Figure and Question Number	Quantitative Question	Interesting points to investigate further
Figure 4.14 -Q12	What type of database engine do you use? Please select all that apply.	99% of respondents used relational database engines. With all the increasing new types of engine and rise in data analytics will this split change the usage pattern? Why is this important?
Figure 4.17 -Q23	What database platforms are used?	3 types of platforms are used physical, virtual and cloud, with cloud consumed services at 18% of respondents. Will the split of 3 add complexity between them?
Figure 4.16 -Q25	What separate database environments do you have for supporting database applications?	There were a mix of environments used by respondents. Will this add complexity between them?
Table 4.1 -Q15	What percentage of your database servers use on-premises database software (run on computers on the premises, in the building) or outsourced database hosting?	On premises or outsourced database hosting were used by the majority, 72% of respondents. Were the rest using cloud providers or something else and what layers were involved?
Figure 4.32 -Q32	Are the following security policies enforced?	Security policies were enforced heavily across all areas. Did this mean it didn't matter how many technical layers there were, networking security, physical server security and data centre access?
Figure 4.49 -Q45	Are data requirements driving database management procedures in your organization?	71% of respondents reported data requirements were driving database management procedures. Did this mean data affected the technical layers?
Figure 4.62 -Q68	What are your database product practices?  Is the database software product selection constrained due to your employee skill set in house?  Is the database software version selected for	The software version could affect or be unaffected by the underlying technical architecture. 57% of respondents stated this was chosen for financial reasons and 48% constrained due to employee skills. Why were employee skills not improved? Will adoption of best practices be affected by these complexities? Is the correct product chosen for the required job?

	financial reasons? (e.g. Standard or Enterprise Edition)	
Figure 4.34 -Q52	What storage types do you use?	There were a variety of storage types used. They all had differing facets and features for configuration. How did this affect the technical layers and operations of databases?

There are many technical layers involved in management of database systems.

The software choice driven from the business and staffing skills can be governed by the business funding. Other questions relating to poor skill set indicated there was a lack of training of staff, but are systems too complex or change too rapidly? There was a proliferation of choices. The platform and engine had different technical architecture layers. The qualitative question asked was:

*Question 5: What complexities between technology layers, do you think, affect the operation of databases?*

### 4.3.6 Managing Cloud Databases

Figure and Question Number	Quantitative Question	Interesting points to investigate further
Figure 4.45 -Q59	Do you have practices and procedures in place in the following areas to manage cloud databases?  The areas: Security; Where the data is stored (e.g. in which country); Availability; Recoverability; Scalability; Expansion and contraction of resources; Access patterns; Reliability; Cost; Management tools; Interoperability between database vendors; Reductions in administration time; Agility; Control	Very few practices and procedures were in place when the survey was taken. This was a new area and not many respondents were using it. When do best practices become established?
Figure 4.43 -Q58	What (if anything) do you use cloud database services for?	The majority didn't use cloud database services. There was a mix in usage of environments for the rest of the respondents. Why are cloud services not used for databases?

Figure and Question Number	Quantitative Question	Interesting points to investigate further
Figure 4.42 -Q57	Do you use any of the following forms of cloud database options?	The majority didn't use cloud database options. For the rest various different types of cloud service were used. Will this affect best practices and add to the complexity?
Table 4.2 -Q16	What percentage of your databases use cloud database services (databases which are accessible via public, private or hybrid cloud instantly, on-demand, e.g. SQL Azure)?	30% were using cloud database services for some databases. Does this add complexity and affect adoption of best practices?
Qualitative quote -Q82	If you are using the cloud for database services, what is the reason behind this? E.g. Database as a Service (DBaaS) for Self Service database functionality or Self-managed database servers running on Infrastructure as a Service (IaaS) (optional question)	<p><i>"Cost and flexibility"</i></p> <p><i>"Wouldn't consider"</i></p> <p><i>"Network pricing, reliability &amp; bandwidth"</i></p> <p><i>"Do not &amp; will not trust vendor clouds"</i></p> <p><i>"Security &amp; cloud data are held by many organizations"</i></p> <p><i>"Less in depth management, low onboarding cost"</i></p> <p><i>"scalability"</i></p> <p>How can these issues be addressed? Would business want to use cloud for everything?</p>

Cloud technologies were and are still growing and developing features which include security data protection and tools to enable ease of management. Each type of cloud service has differing risks, issues and possibly methods for best practices and procedures. With this evolution from on-premises to cloud or hybrid scenarios, it was important to understand the best practices and procedures and any complex interactions in adoption.

Only a few cloud practices and procedures were used. Positive and negative responses were shared from the current adopters of cloud. The qualitative question asked was:

*Question 6: Describe any complexities that exist with the adoption of best practices and procedures when managing cloud databases?*

### 4.3.7 Complexity Compromising Implementation

Figure and Question Number	Quantitative Question	Interesting points to investigate further
Figure 4.40 -Q66	What are your working team practices? Are frequent malfunctions dealt with in a reactive role (e.g. fire fighting)?	Almost half of the respondents reported responses to malfunctions dealt with reactively. Did complexity affect best practice as a lot of work was reactive?
Figure 4.41 -Q66	What are your working team practices? Do you put long term fixes in place for regularly occurring issues to future proof the database applications?	66% of respondents always or often put long term fixes in place for regularly occurring issues. When long term fixes were put in place did that alter best practice?
Figure 4.58 -Q31	Do you think some changes are carried out 'under the radar' i.e. by not following policies and procedures?	57% of respondents very often, regularly or sometimes carry out changes under the radar. Were the best practices too complex or did the changes compromise the ability to manage the system?
Figure 4.19 -Q60	How do you receive database training?	85% read articles when required. Did this mean implementing best practices were only considered when a problem had occurred?
Figure 4.20 -Q66	What are your database product practices? Does the company foster an environment to encourage certification?	Certification was generally not encouraged (25% rarely, 28% never), why not?
Figure 4.22 -Q61	How often do you have the opportunity to undertake formal training courses?	23% of respondents never have the opportunity to undertake formal training which may compromise ability to implement best practices and procedures. Is this a reason for many failures?
Figure 4.23 -Q63	How often do you have the opportunity to attend database conferences, workshops or seminars?	24% were never able to attend database conferences. Conferences allow issues and best practices and procedures to be discussed and solution shared. Did this reduce the quality of database management?

Implementation of best practices and procedures can be undertaken in a number of ways. This can be formally following best practices and procedures or under the radar ignoring practices and procedures. Deployment of the tasks can be in a reactive or proactive often related to the organization culture. This type of culture may be a result of the complexity not being well understood.

Implementation could also be affected by knowledge, skills and training. Learning tracks can be through formal training courses, conferences, certifications and informal reading. The rapidly changing technology area probably added complexity in the interaction and adoption of best practices. If people didn't understand the new technologies, which could be helped through training, it could compromise the ability to implement best practices and procedures. The qualitative question asked was:

*Question 7: Was there ever a time when you felt the complexity of database systems compromised your ability to implement best practices and procedures?*

#### 4.3.8 Creation and Control of Best Practice

Figure and Question Number	Quantitative Question	Interesting points to investigate further
Figure 4.11 -Q79	Please select to what extent best practice is currently controlled within your company in each of the following areas	Control throughout the lifecycle was variable, with database security controlled the most by 51% of respondents. Are complex interactions affected by different levels of control? Is adoption of best practice different for different areas where control levels vary?
Figure 4.12 -Q70	Who controls database choices?	There was a change in control of database choices when cloud database software was chosen. It was controlled by the Head of IT Operations. For other choices usage was controlled by the database administrator / database manager. Does this change in control affect database systems and add complex interactions?

Control of database systems can be through automation or people-driven mechanisms. The environment may also add influences through government legislation, or the structure within large corporations, with organizations using different teams and managers to control at different levels. The different levels of control and change in control within the organization may be affecting complex interactions and best practice. The qualitative question to look at creation and control of best practice was:

*Question 8: Who you think should create and control database best practices and procedures?*

### 4.3.9 Cross Boundary Communication

Figure and Question Number	Quantitative Question	Interesting points to investigate further
Figure 4.59 -Q67	What are your communication and business practices?	11% of respondents stated cross boundary communication was always good. 19% of respondents stated cross team communication was always good. These were low percentages and may affect management of database best practices and procedures. Does communication play a part in complex interactions? Does communication affect adoption of best practices and procedures?

Communication is required when working with groups within the organization. Organizational communication is also important to allow for the most effective use of the people and teams skills and abilities. This communication can affect perspectives, conflict and culture. Tasks where time and cost is a factor may affect design and operation of database systems if not communicated well. Cross boundary and cross team communication was reported as not always being good. A qualitative question to investigate communication was:

*Question 9: How, if at all, do cross boundary communications among stakeholders affect best practices and procedures?*



### 4.3.10 Strategic Planning

Figure and Question Number	Quantitative Question	Interesting points to investigate further
Figure 4.64 -Q80	Which of the following could improve practices and procedures for database management in your organization?	55% of respondents thought organizational database roadmaps could help improve database management. How can roadmaps be developed in a complex field?
Figure 4.9 -Q76	Does your organization follow any database best practice guidelines?	Some level of best practices were followed for 90% of respondents. Does a plan affect this?
Figure 4.61 -Q69	Is database management visible to the following people or teams?	Database management was visible clearly only within the database management team. The strategic plans were drawn up by higher management. Does this affect adoption of best practice?
Figure 4.18 -Q29	What service availability is required for your databases servers?	There were various service availability targets that the respondents had. The targets can significantly affect the design choice. Were these taken into account in the strategic plans?
Figure 4.6 -Q17	What percentage of your time is spent managing database servers?	Only 7% of the respondent's time was all spent managing database servers. Would strategic plans affect the rest of the time? Did the strategic plan affect the amount time the respondent's had to manage their databases? What are the complex interactions? Does this affect complex interaction?

Strategic planning within database systems is constructed of several parts that relate to the business tier and objectives, the management of database systems, the best practice guidelines for the technical management and the tasks carried out by the people. Thus strategy improvement and planning may need further investigation. The last qualitative question asked for the focus group research, to draw the investigation to a close, was:

*Question 10: What effect can a database management strategic plan have on best practices and procedures for the management of database systems?*

## 4.4 Summary

The quantitative survey findings from 453 worldwide respondents have been presented in this chapter. The questions represented in the survey covered the full end to end database systems management. Best practice usage in action was documented in the findings from the wide spectrum of respondents. The findings were mixed, with best practices used in some cases but not in others. Best practices and procedures were used to some extent in the management of database systems.

The findings from the quantitative survey raised further questions for investigation. These led to the development of 10 questions that would form the structure of the qualitative focus groups' discussion. These questions were aimed at giving a better understanding of the operation of databases in each of the above areas. This could bring insight into the issues involved, enabling a better informed answer to be suggested to the second and third research questions.

At this phase of the research certain overall factors in the operation of databases could be grouped together as follows. Control was a most important factor that affected the successful running of a database. It involved software application control and management communication with and between the actors. Data was the basic element of the database, the type, size and required output were necessarily important. Also knowledgeable and trained staff were key to providing efficiency and effectiveness when using suitable software. All efforts could fail in the face of unpredictable events and could not be entirely planned for, except by back-ups and recovery procedures.

The next chapter presents the data from the qualitative research obtained from the focus groups. The narrow and deep analysis provided an in depth understanding of the management process.

## Chapter 5: Qualitative Findings, from Analysis to Synthesis

### Synthesis

#### 5.1 Introduction

The quantitative survey data provided a wealth of information about best practices and the current management of database systems. The data provided many insights into the nature of database systems but did not provide an explanation of management decisions. The mixed method approach undertaken as part of this research addresses the shortfall in understanding by following the quantitative survey with a set of qualitative focus groups.

Two of the research questions sought to examine the complexity and interactions:

*2. What are the complex interactions that are an integral part of the management of database systems?*

*3. Is the adoption of best practices and procedures affected by the complex interactions that are an integral part of the management of database systems?*

The questions formed the basis of this next phase of investigation, the qualitative stage. This stage of the research collected data from a number of focus groups that were held in Europe (Bath, Cambridge and Amsterdam). Various different types of focus groups were tried: face to face focus groups, an asynchronous email group, and an asynchronous forum focus group. The latter two are sometimes known as virtual focus groups (Liamputtong 2011, p.12). Five face to face focus groups, two asynchronous e-mail groups and one asynchronous forum were organised. A total of 29 responses were received. Liamputtong (2011, p.44) suggested that four to five

focus groups might be sufficient; eight groups were organised. Collecting data from the focus groups was continued until such a point that the data started to repeat themes.

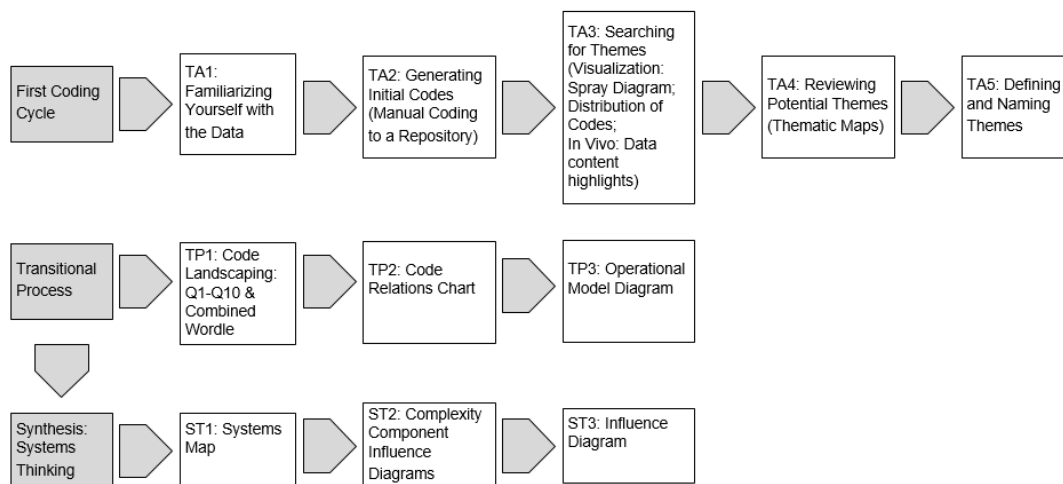
Ten questions (listed in Section 5.4.3 and Appendix B), derived from the quantitative analysis, were created to help structure the discussions and enable the participants to share their experiences.

This chapter is a presentation of the qualitative results, traversing through the method and providing examples of the data. The research data was examined in a holistic manner considering the system of interest and the behaviour of the components. Some of the data demonstrated that there was considerable advantage to using best practices, and some highlighted a number of issues where careful consideration was required. In addition the complex interactions were demonstrated through a number of examples within the data.

The full analysis of the qualitative data is dealt with in Chapters 5 and 6. Chapter 5 examines the results in relation to the method. Chapter 6 discusses the findings relating to each research question and includes the findings of the quantitative data.

## 5.2 Qualitative Analysis Process

The qualitative analysis process undertaken is described in Chapter 3, the methodology chapter; and a summary is shown in Figure 3.3.



**Figure 5.1 Qualitative analysis process**

The summary of the qualitative analysis process depicts the high level view of each of the three stages of the process.

Stage 1. The first coding cycle analysed the data using thematic analysis. Thematic analysis is a foundational analytic method for identifying and analysing patterns in qualitative data (Clarke & Braun 2013, p.120). A key component of thematic analysis is identifying codes and themes within the data. Themes and codes are used in many qualitative research methods. In the context of this research, themes and codes are defined in Chapter 3. Thematic analysis relating to practices and procedures utilized by the database community offered much information of value. Further details of the process are given in Section 5.4 below.

Stage 2, a transitional stage, followed the thematic analysis to bridge between the analytics and the following synthesis stage. The transitional stage started with a review of the data corpus using code landscaping, looking at the prevalence of textual words visually in word clouds. This was followed by an examination of the code relations, prevalence of codes to map patterns. Then an operational model

was created, to view a visual representation of system relations. Details of the process are given in Section 5.5 below.

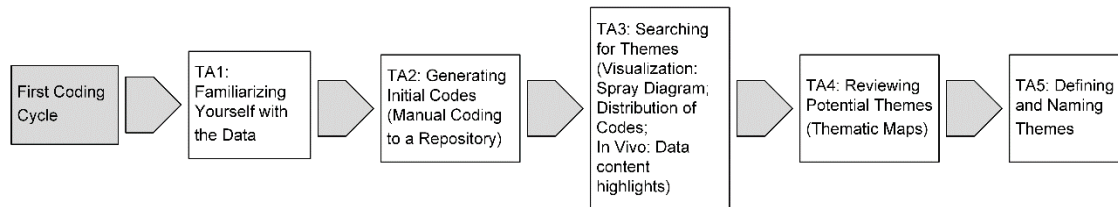
Stage 3, the synthesis stage, enabled interrelationships to be identified, and highlighted emergent properties. Systems thinking was used in this stage to gain insight into the complexity of the situation. This stage focused on understanding the whole system. The strategy adopted was systems thinking and the specific use of systems diagramming. In the systems thinking part of the research the codes were treated as if they were components of the system. Reviewing the results in a holistic way enabled the structures that existed within the data to show the complex situation and interrelationships. Details of the process are explained in 5.6 below.

### 5.3 Participant Demographics

The participants of the focus groups had a variety of professional roles and experience. In total there were 29 participants. The largest groups of participants were database administrators (DBAs); Business Intelligence (BI) workers and database architects formed the next largest number of participants. The self-employed consultants came from a variety of backgrounds. The focus group participants were selected to ensure diversity in the types of database roles. This was important to obtain a balanced view from different areas within the field. The participants were mostly very experienced within the database field, having worked in the field for over ten years. The comments raised focused more on the small to medium enterprise although three of the participants were from large enterprises. Consultants may have worked in both types of environment. The fact that not all participants were from the UK, with some being from the USA and Ireland, added further diversity to the data obtained.

## 5.4 First Cycle of Coding

The first steps of analysis are shown in Figure 5.2.



**Figure 5.2 First coding cycle**

This first cycle of coding incorporates all the initial methods used in the initial coding of the data (Saldana 2013, p.58). Thematic analysis (Braun & Clarke 2006) was the primary method used in this cycle. Additional tools (spray diagrams, distribution of codes, in vivo coding and thematic maps) were also used to help understand the data.

The high level steps of Thematic Analysis (TA) are:

- TA1: Familiarising yourself with the data
- TA2: Coding
- TA3: Searching for Themes
- TA4: Reviewing themes
- TA5: Defining and naming themes

A coded data example from the data collected in this research is shown in Table 5.1. Braun & Clarke (2006) defined the 'data corpus' as all the data collected for the research; a 'data item' is an individual unit of data that was collected. A 'data set' refers to all the data items collected that are being used for a particular study.



**Table 5.1 A data item from the data corpus**

Code	Theme	Data Item
Best Practice	Business Best Practice	<i>“but isn’t it requirements’ to why those best practices don’t change ,it is the business requirement of secure data but the implementation of how you, what you have to do to achieve that written goal is very different” (Q6 3.2 line 59)</i>

A code can be related to many themes. All data in the data set was coded systematically in this manner.

### 5.4.1 Familiarizing Yourself with the Data

TA1: Familiarizing Yourself with the Data

Familiarisation with the data, the first stage of thematic analysis, began with meticulous transcription and collation of the focus group data. The two focus groups from the asynchronous email focus group and online forum group were collated and moved to the same text format as the transcribed face to face focus groups. The face to face focus groups were rigorously transcribed with a verbatim account of all the verbal comments and were checked back against the original audio recording for accuracy. The transcripts from the focus groups were initially reviewed for clarity following transcription and then reread to become immersed in the breadth and depth of the data. It was also important, whilst becoming familiar with the data, to identify possible patterns; this would assist when creating codes and themes.

### 5.4.2 Generating Initial Codes and manual coding to a repository

TA2: Generating Initial Codes (Manual Coding to a Repository)

Generating initial codes, the second step of thematic analysis began with transcription of the focus group interviews, followed by a process of data segmentation.

The qualitative data was first analysed relating to each of the ten questions. A few elements of data were removed where the focus group dialogue leapt into an in depth technical discussion relating to technology and was not related to the research questions.

The text was cut into paper segments and manipulated into clusters of similar areas. The clusters, potential codes, were outlined in pencil. Best practice was one such cluster. These clusters were created from the data during a process where the researcher became immersed in the depth and breadth of the data collected. All of the data from each of the ten questions was sorted and clustered. As such it became possible to see all the comments relating to best practice from all of the ten questions. Figure 5.3 provides an example, from a subset of data, of the initial sorting that was undertaken for two of the questions, and shows early development of the 'best practice' code. Best practice is a confused and contested concept.

<b>Question 1 Best Practice Cluster</b>	
Best practice empower	Industry best practice
Coding best practice	Company best practice
Best practice what people do	Follow best practice without realising
Best practice influences	Accepted practitioners best practice
Best practice isn't defined	Hundreds of best practice rules
Ethereal best practices	Some best practice more important
Best practice categories	Unsuited best practices for use cases
Guidelines	Google Search for Best Practice
Implementation to make procedures industry based	
Best practice set a baseline of known performance and configuration	

<b>Question 2 Best Practice Cluster</b>	
Require different best practices	Working group should select best practices
Adhere security best practice	Best practice driven by business need
Best practice not considered	Best practice and procedures might affect the purchase decisions
Best practice versus actual best practice	

**Figure 5.3 A subset of the initial raw data for Best Practice within Questions 1 and 2**

The coding was influenced by data-driven themes rather than theory-driven themes. The codes identified features in the data. Systematically working through the entire data set produced results shown in the example in Figure 5.3.

The data in the transcripts, together with the initial code assignments, were then meticulously transcribed and migrated into Microsoft Excel for loading later into SQL Server. To anonymize the data and to ensure ease of reference later, each participant was given a number. The Participant Identification Number (PID) was created with the concatenation of the Participant ID and Focus Group ID: where the Participant ID is 1 and the Focus Group ID is 1 the PID is 1.1 (Appendix F).

The extracts of data were coded for as many potential themes and patterns as possible. All the data was retained at this stage to ensure the story told by the participants was clear.

This stage ended when the data from all of the questions had had initial codes generated. The dominant story was beginning to appear in the data analysis, although many tensions, relations and inconsistencies in viewpoints were found in

the extracts. In Chapter 3 inconsistencies were mentioned and Braun & Clarke (2006) stated this was not unexpected although still important.

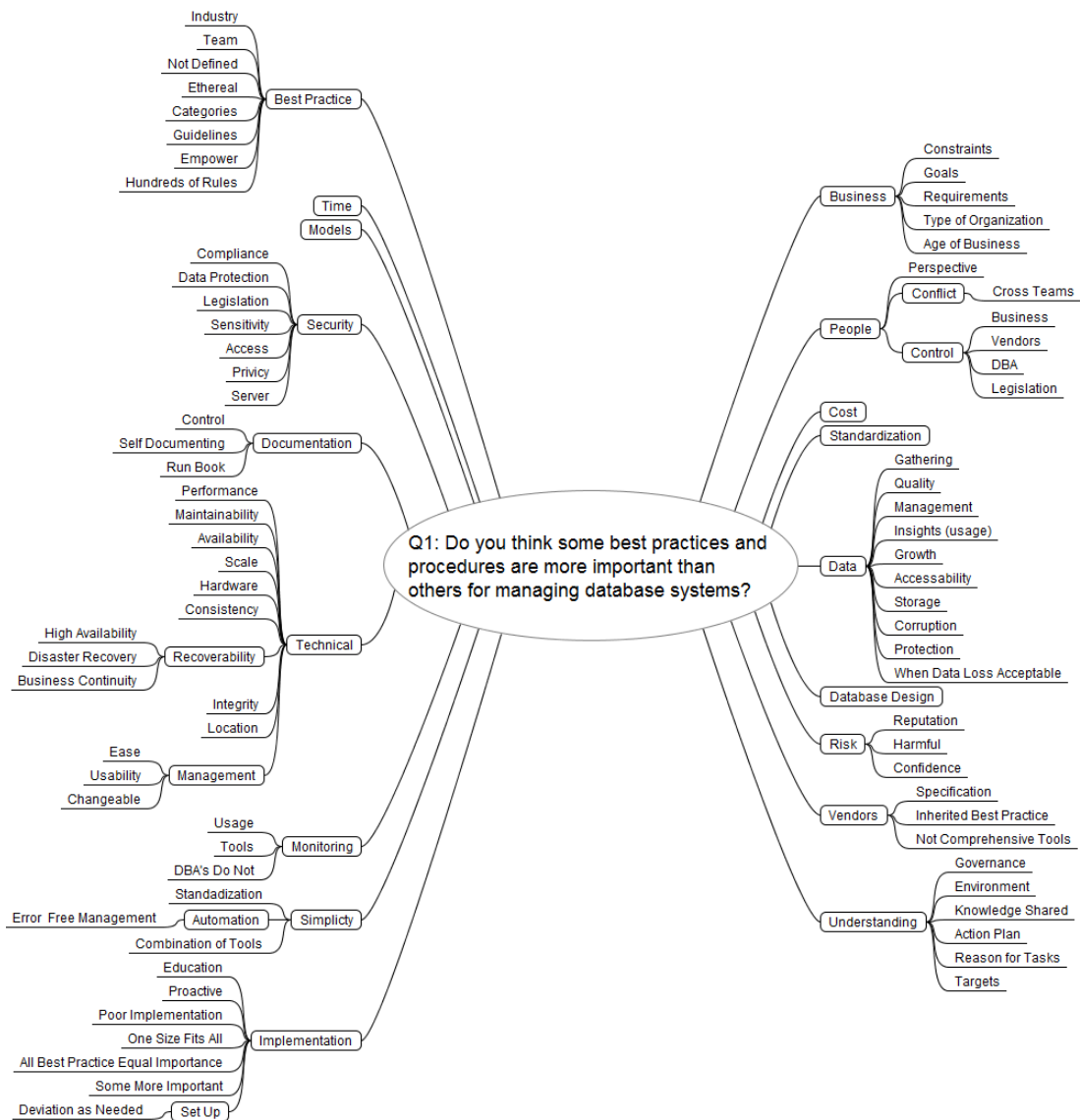
### 5.4.3 Searching for Themes

TA3: Searching for Themes  
(Visualization: Spray Diagram,  
Distribution of Codes.  
In Vivo: Data content highlights)

The third step, involved reviewing the structure and relationships of the data in each question for the initial codes and themes.

#### Visualization: Spray Diagram

Visual representations of the data codes and themes, using spray diagrams, were used to help with the creation of the overarching themes and understanding their relationships. The spray diagrams were a part of the exploratory heuristic investigation and not part of the final codes and themes. They are conceptual maps of the situation showing connections between the potential codes and summarised potential themes. The spray diagram for focus group Question 1 is shown in Figure 5.4; Question 1 asked whether some best practices are more important than others when managing database systems. This chapter follows through with 'best practice' as an example of the analysis process; it was necessary to be selective due to the lack of space to share all the data discussions.



**Figure 5.4** Spray diagram of Question 1 codes and themes from the data corpus

A code is (a short word or phrase), shown in an oval, and the themes each capture something important about the data, related to that code. The spray diagrams, Appendix D, represent an analysis of the responses for each of the ten questions. These spray diagrams were used for the early stages of analysis to help understanding the codes and potential themes that were derived from the data by the researcher, for each question.

The spray diagram for each question has many codes (e.g. see Figure 5.4). One of the codes is “documentation” with three themes: “control”; “self-documenting” and “run book”. These three themes were drawn from the data transcripts:

“Researching and creating a “run book” to manage”

“Following run book”

“Document control”

“Document usage”

“Self documentation”

Another code, “best practice”, raised themes in the Question 1 data as follows:

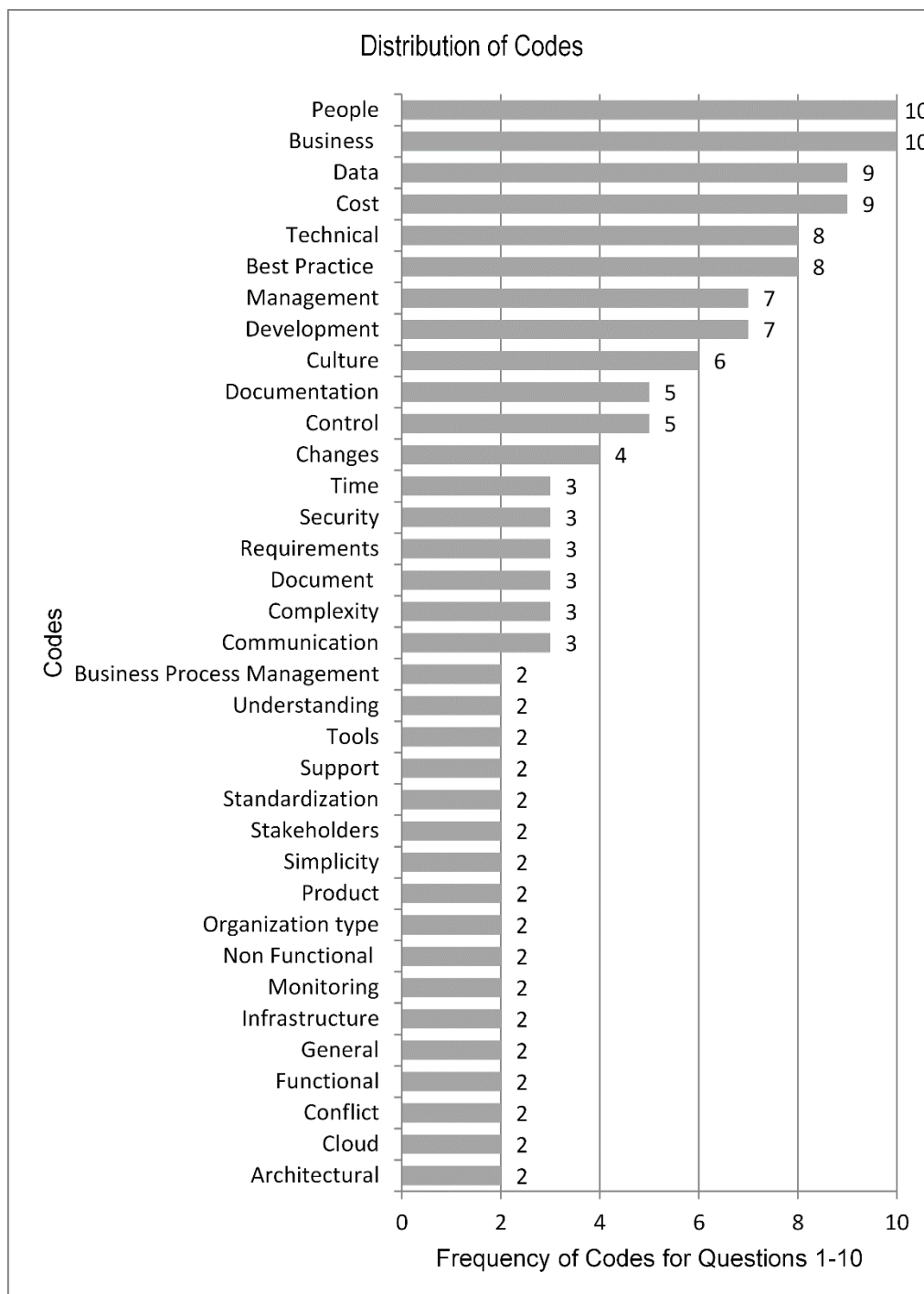
“industry”; “team”; “not defined”; “ethereal”; “categories”; “guidelines”; “empower”; “hundreds of rules”. These were based on the data transcripts for Question 1. This could then be considered with best practice summary themes in the other questions in Appendix D. When the spray diagrams were completed the visualization of data complexity became clearer and the spray diagrams helped to show some order within the data.

### **Visualization: Distribution of Codes**

The analytical technique used to review the codes within the data was a code frequency report. The report provided a frequency count of the codes across the ten questions (see details below). Although the codes were early interpretative summaries of the data, this method helped illuminate and mark some codes for redefinition.

The distribution of the codes from questions 1-10 was calculated based on whether the code occurred in each question. If a code appeared more than once in a

question it was counted only once, in order to gain a visual picture of recurring codes across the questions. The codes were refined later in the analysis. The code frequency bar chart is shown in Figure 5.5 for the most frequent codes. The codes in the spray diagrams were reviewed and the revised iteration summarized in Figure 5.5. The codes shown in Figure 5.5 are not the final codes but required further iteration.



**Figure 5.5 Distribution of codes within all 10 questions with greater than 1 occurrence**

This initial distribution of codes was used to help refine the codes that were finally produced. Some codes were later consolidated with other codes. It was clear from Figure 5.5 which codes were more prominent than others and that key codes across the data set were “people” and “business”. The codes which appeared in five or



more questions were: people; business; data; cost; technical; best practice; management; development; culture; documentation and control.

### **In Vivo : Data Highlights**

The aim of using the In Vivo coding technique was to gain an overall picture of the data highlights from direct quotes in each of the 10 questions. This was used to give a high level view of the participants' own words. The words and phrases that were considered significant, at this early stage, from each of the ten questions follow.

These extracts were selected through reviewing the data for each question, looking for key comments. These quotes do not relate to the codes but were used as a review to check that no meaning was lost. The quotes presented relate to the data corpus in each question discussed in the focus groups.

*Q1: Do you think some best practices and procedures are more important than others for managing database systems? If so, what are the most important ones?*

“best practice categories”  
“best practice different for different use cases”  
“requirements sets”  
“guidelines”  
“monitor usage”  
“researching & creating a “run book” to manage”  
“following run book”

*Q2: What best practices and procedures do you think should be considered when selecting different database engines?*

“strategic application”  
“decision up front”  
“evolution on case to case basis”  
“right tool for job”

*Q3: What kind of requirements gathering and architectural design processes for the hardware, data and databases do you think are important? Why are these important?*

“designing up front”

“start up initial planning”

“start up investigations”

“early requirements discussions”

“rookie mistake to treat first cut as final requirements”

“review & modify best practice”

*Q4: In what ways do you think that best practices and procedures could assist management of the database lifecycle?*

“communication”

“run book”

“visibility at all levels”

“control”

“order from chaos?”

“lifecycle plan”

“guidance”

*Q5: What complexities between technology layers, do you think, affect the operation of databases?*

“Database engines end up as edge cases for the storage admins, sysadmins, licensing admins, etc.”

“biggest thing that comes to mind is it’s not a technology problem, but a people problem”

“needs three domains: people, process and technology. Some authors have added a fourth (Business).”

*Q6: Describe any complexities that exist with the adoption of best practices and procedures when managing cloud databases?*

Problems:-

“that complex systems (people, process, technology or business) sometimes have needs that hamper conventional best practice”

“playing by someone else’s’ rules”

“loss of control”

“data protection a problem”

“security and quality risk”

*Q7: Was there ever a time when you felt the complexity of database systems compromised your ability to implement best practices and procedures?*

“ practices and procedures should achieve consistency”

“best practice is never really best practice – it is just a best solution in a particular set of circumstances”

*Q8: Who do you think should create and control database best practices and procedures?*

“discussion/ agreement about what will actually work”

“arbitrate & get agreement for everyone”

“set by cross party team”

“architects, developers, DBA’s, sysadmins”

“a Senior IT business developer etc. staff member to control”

*Q9: How, if at all, do cross boundary communications among stakeholders affect best practices and procedures?*

“put aside personal differences”

“cross group / stakeholder collaboration”

“stakeholders must come together to resolve issues”

“solve conflicts”

“create communication (cross boundary) to provide holistic picture”

*Q10: What effect can a database management strategic plan have on best practices and procedures for the management of database systems?*

“a road map”

“development plan for 10 years”

“includes flexibility”

“modification of plan may be needed”

“right tools”

“strategic focus”

The In Vivo Coding gave increased depth in understanding of each of the ten questions; when looked at holistically, this helped build up the picture of the complexity.

Table 5.2 is an example of the code book that evolved throughout the qualitative analysis.

**Table 5.2 Code book early generation of codes and themes**

PID	Transcript (Data Item)	Initial Themes	Line No.	Codes	Job	Time	Size	Place
8.1	The single most important best practice is designing a solution to meet the business's goals for minimum allowable data loss. If the business wants to lose no more than, say, 5 minutes of data, then we can't design a solution that allows for a higher level of data loss.	Best Practice design Meet business goals Minimum data loss, Constrained design	119	88 92 91 11	F	M	S	N

In the example for PID 8.1 listed in Table 5.2 the numeric codes are: 11 = Database Design, 88 = Best Practice, 92 = Business and 91 = Data. The data transcripts in bold were coded and potential (initial) themes highlighted.

Each row represented a participant entry, with the columns for demographics defined as

Job: What is your job area?

Time: How long have you worked in the database field?

Size: What size of organization do you work in?

Place: In which country do you work?

References back to the actual data are quoted hereon throughout the thesis in a consistent format such as: (Q1 8.1 line 119). This refers back to the Question Number + PID (Group ID + Participant ID) + Line Number (in the spreadsheet). The example above is for Question 1, participant 1 from Focus Group 8, line 119 from the spreadsheet contained within the code book. The key is in Appendix E.

The code book was then all ported into Microsoft SQL Server to allow further data analysis.

At the end of TA3 the data corpus had all been reviewed to create initial themes.

#### 5.4.4 Reviewing Potential Themes

TA4: Reviewing Potential Themes (Thematic Maps)

This is the fourth step of thematic analysis that is to review, combine, refine, separate or discard themes. Some themes did not have enough data to support them, some required combining two themes into one (due to there not being enough data to have separate themes). Some new themes occurred. Thematic Maps were used to help carry out further analysis in this stage.

The example below for 'best practice' illustrates how this approach was used in this research. The codes and themes for best practice from the questions were combined (Table 5.3). A further review of coded extracts was undertaken, candidate themes were collated together and thematic maps were then developed. Table 5.3

details the themes and the subthemes for ‘best practice’ that were created from the data.

**Table 5.3 Combined themes for the best practice example**

<b>BEST PRACTICE – CODE</b>	
<b>Themes</b>	<b>Subthemes</b>
<b>Best practice not followed</b>	Best practice is not followed for greenfield projects; best practice not considered; not chosen to implement best practice; technology processes in department not adhering to own best practice; not have any best practices or procedures; the best practices don't exist yet; articulate and identify when/why chosen to not implement best practice
<b>Breaking best practice</b>	Breaking the best practice that is stopping implementing my best practice; Best practice doesn't fit someone will go around it; what else that you should do that breaks best practice
<b>Issues with following best practice</b>	Application compromised ability to follow best practice; where best practices and procedures are lacking it can lead to unnecessary complexity; if hadn't broken it, would have been stuck with the best practice handed down; identify poor practices blocking implementing wider best practices; hard to build best practices, consistency outweighs best practice; best practice not over prescriptive; best practice not long winded; best practices prevent applications working; can actively disrupt best practice; unsuited best practices for use case; some best practice misinterpreted; best practice should have been done at the beginning
<b>Best practice facet</b>	Best practice categories derived standards; hundreds of best practice rules; best practice organic; best practices going to guide everything that happens; best practices are how we run our systems now; best practice never really best practice - it is just a best solution in a particular set of circumstances; some best practice more important; some best practices completely different with a set core being the same; no generic best practice; follow best practice without realising; lots of best practices; worldwide perspective
<b>Define best practice</b>	Define own best practice; best practice isn't defined; best practice defined by set up; tailor best practices to match; treat best practices as a general guideline; best practice is the gold plated element; best practices are a guideline not a requirement; treat best practices as a general guideline; compromise on what is best practice; not universal truths but guidelines (think: maxims); very strongly affect best practices and procedures (BPP); since BPP reflect a set of empirically derived ways of optimal operations; ethereal best practice; unhelpful to call things a blanket best practice; approaches not best practices; what you do with that in perfect world; some more important, that is best practice; dislike term best practice; input on best practices

BEST PRACTICE – CODE	
Themes	Subthemes
<b>Lifecycle best practice</b>	Best practice not value in lifecycle; best practices and procedures for database lifecycle; best practice application have own lifecycle
<b>Best practice requirements</b>	Requirements best practice; general best practices gathering requirements; require different best practices; best practice versus actual best practice; how best practice fit; particular best practices; want best practice; best practice; never compromising my ability to implement best practice; best practice to use in built tools
<b>Best practice reduce risk</b>	Best practices and procedures avoid relearning mistakes; best practices reduce exposure to risk; future release shouldn't effect best practices for today's systems best to concentrate focus on best practices and procedures

Thematic maps were created to investigate the conceptual data patterns and relationships between the themes for a given code. For each code, with respective themes, the data from all of the 10 questions were combined. There are no prescriptive rules for constructing thematic maps as they are a tool to help researchers mapping out their themes. An example of part of a thematic map for the best practice code is shown in Figure 5.6.

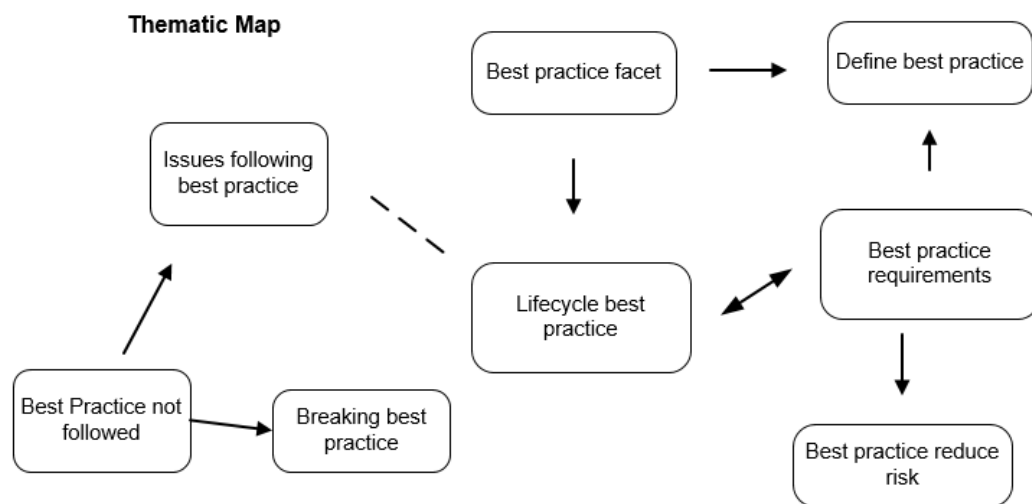


Figure 5.6 Best practice example thematic map

The thematic map in Figure 5.6 identifies a selection of themes related to best practice. The themes “best practice not followed”; “breaking best practice” and “issues following best practice” could be connected. The three themes are described below. Participants thought best practices were not followed for Greenfield projects (Q2 4.1 line 50). The participants thought breaking best practice was the only option if it stopped them implementing their own best practice (Q7 2.2 line 48). If the best practice was inappropriate people would go round it or it would be circumvented (Q8 5.4 line 85). Issues following best practice could occur for inherited systems (Q7 7.1 line 103), when tasks were not completed at the beginning of projects (Q7 3.1 line 62) or in some cases people just believe best practices are harmful and thus not used (Q7 3.1 line 62).

The ‘Best practice facets’ theme was about derived standards, rules, the best solution and following best practices without prior definition. The participants defined these best practices as guidelines (Q8 5 line 78), empirically derived ways, where they were set up by the organization (Q9 7.4 line 80), or where best practice was defined by themselves (Q2 2.2 line 18). Best practice requirement could be driven from definitions, with different best practice required by the project (Q2 5.4 line 65), using in-built tools which was more organic (Q10 2.2 line 15) or participants stating that checking best solution works with current best practice, but noting they may need to change over time (Q10 8.2 line 111). Best practice should avoid relearning mistakes (Q4 8.1 line 78), reduce risk (Q4 7.4 line 75) and future releases should not cause any problems with these best practices (Q10 8.3 line 113).

The “lifecycle best practice” theme was tentatively connected to ‘issues following best practice’ due to concerns over the lifetime durability of the hardware performance (Q3 6.1 line 79). Best practices lifecycle may be affected by the best practice requirements.



In this step themes were reviewed to more accurately represent the data and an understanding of how they fit together was gained. The resultant themes tell the story of the data.

#### 5.4.5 Defining and Naming Themes

TA5: Defining and Naming Themes

Thematic analysis TA5 followed where each theme was defined and refined. The themes helped tell the story of the data and provided vivid extracts and patterns across the data.

These themes have grown organically and changed throughout this exploratory phase.

Many best practice themes and data quotes are detailed below. This is just part of the data.

##### **Theme: Business best practice**

Business requirement of secure data:

*“but isn’t it requirements’ to why those best practices don’t change it is the business requirement of secure data but the implementation of how you, what you have to do to achieve that written goal is very different”*  
(Q6 3.2 line 59)

##### **Theme: Industry or company best practice**

*“most places have very similar requirements that then becomes if not then an industry best practice at least a company best practice”* (Q1 2.2 line 40)

##### **Theme: Operational best practice**

*“Best Practices help to ensure that databases are on hardware that is scaled correctly, they are not riddled with bugs, they play to the strengths of RDBMS selected and are secure and available for SLA. Best Practices should not however be overly prescriptive or so long winded that they either stifle innovation or adoption of new technology or are just ignored.”* (Q4 5.3 line 65)

Automation scripts to adhere to standard and best practice:

*“without a level of control people make mistakes and things can get missed whether it is an automated process or some companies make a choice ours is the installations whatever the other thing is you can have like you said at the beginning a PowerShell script that deploys a database to a specific standard that is used and it adheres to best practice”* (Q8 2.1 line 30)

### **Theme: Life cycle best practice**

Best practices and procedures could assist management of the database lifecycle:

*“At the end as it is the data that drives what best practices to set up”* (Q4 5.5 line 63)

Takes years to work out what works as a general practice and best practices do not exist initially:

*“In newer database engines, the best practices just don't exist yet. It takes years to figure out what works well as a general practice (rather than what worked well for one guy.)”* (Q4 8.1 line 77)

Best practices and procedures could all contribute to and assist with management of the database through every stage:

*“They should contribute to every stage, as all stages require guidance and support from tool selection, through analysis, design, build and operation.” (Q4 1.2 line 2)*

**Theme: Changing best practice**

*“In mature databases, best practices and procedures avoid relearning mistakes. Unfortunately, the best practices and procedures have to remain agile because hardware and software continues to change.” (Q4 8.1 line 78)*

*“Best practices are in and out of vogue” (Q7 5 line 83)*

*“Large complex systems often get moved away from the best practice configuration because the best practices are built for 95% of the systems out there, not the 0.05% which are the biggest systems on the market.” (Q7 8.3 line 112)*

**Theme: Best practice facets**

Not meeting expectations. Best practice is just a best solution in a particular set of circumstances:

*“best practice never really best practice- it is just a best solution in a particular set of circumstances, when situation is different pragmatism dictates choose something different, unhelpful to call things a blanket best practice, falling short, procedures and practices should achieve consistency, manageability, tailored to situations” (Q7 1.4 line 4)*

**Theme: Best practices reduce risk**

*“In general, best practices reduce exposure to risk, minimize firefighting, and ensure strong performance and easier recoverability / business continuity.”*

(Q4 7.4 line 75)

Best Practices when followed can avoid risk and problems. Understanding the reasons for these practices is not required.

*“I think a lot of people do this kind of thing without realising what they are asking but what they are doing is implementing best practice.”* (Q 1

2.2 line 44)

**Theme: Understanding best practices**

*“Poorly understood best practices can make it all worse. Best practices should be treated with care. They are not universal truths but guidelines (think: maxims)”* (Q9 7.5 line 81)

**Theme: Design best practice**

*“Frequently it is the application design, especially with 3rd party tools / applications, which are frequently not designed to work with the required best practices, or can even actively disrupt them.”* (Q7 5.1 line 91)

**Theme: Best practice not followed**

*“If you don’t have any communication you tend not to have any best practices or procedures”* (Q9 3.4 line 40)

**Theme: Breaking best practice**

The complexity of database systems does not compromise my ability to implement my best practices

*“because you work round it the complexity of something is never compromising my ability to implement best practice because I break the best practice that is stopping me implementing my best practice” (Q7 2.2 line 48)*

Another participant stated:

*“If a BP doesn't fit, then eventually someone will go around it or it'll get overridden by senior management.” (Q8 5.4 line 85)*

**Theme: Best practice control**

Best practices empower, giving control:

*“the best practices are the ones that empower users to engage with the data and extract value from it.” (Q1 5.5 line 109)*

**Theme: Issues with following best practice**

*“where best practices and procedures are lacking it can lead to unnecessary complexity. I think these situations often arise when one inherits a system.” (Q7 7.1 line 103)*

Best practice should have been done at the beginning:

*“we are not seeing complex systems where things are not the best ideally they should have done it right at the beginning it is another half a*

*day fixing it would have been good but suddenly you are looking at another massive piece of work to roll something back.” (Q7 3.1 line 62)*

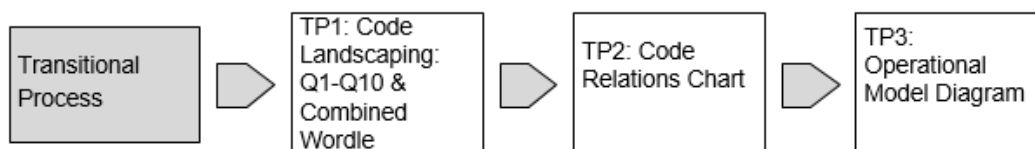
Some can be harmful and self-management of risk could help:

*“As a general rule - I think "best practices" are harmful, not beneficial. This is caused by the widespread belief that when following best practices - people tend to stop thinking for themselves. That being said - I do think that default configuration should be documented and automatically applied across the real estate.” (Q1 7.2 line 114)*

The rest of the data, codes and themes were documented in the codebook.

## 5.5 Transitional Process

The transitional stage moves from defining themes using thematic analysis, to the synthesis phase. The tools used in the transitional stage helped in switching from in-depth data analysis to viewing the data holistically. This stage looks at the post-coding transition by further reanalysis of the combined dataset to provide a clearer focus to enable the analysis to move to the next level (Saldana 2013, p.187). This leads on to the proposal of an operational model diagram presenting the results in terms of the systems which together form the database system. The process is shown in Figure 5.7 below.



**Figure 5.7 Transitional process**



The word frequency counts of the highest occurring words are listed in Appendix C. The prominence of the words provides an indicative view of the data, and the word clouds for each question provide a specific insight into the data. For each question the data from the discussion always contained a few of the key words that were provided in the question. These word clouds, created from the data corpus of each question, provided a picture that validated the codes and themes from a holistic viewpoint. They also enabled a clearer vision of results for the individual questions. Then, combining all the questions, it was possible to gain a view of the overall most prevalent words. The results of the word clouds are shown below with the words in bold where they were used in the question:

- Question 1 Data, business, **best practice**, security, access, maintainability, practices, control, implementation, different, recovery, **requirements**, loss
- Question 2 Platform, **engine**, business database, requirements, vendor, cost, solution, process, separate, practices, **best practice**, functional, management, support
- Question 3 Processes, business, **requirements**, scale, design, **hardware**, **data**, performance, growth
- Question 4 **Best practices**, **lifecycle**, tools, procedures, designs, help, support, standard, requirements, documentation, operational, chaos, quality, availability, database
- Question 5 **Complexity**, **databases**, **technology**, performance, teams, storage, different, **layers**, problems, standards, things, people, open, components, operations, data, application



- Question 6 Practices, **cloud**, data, different, **best practices**, business, change, control, databases, security, cost, procedures, private, used, requirements, control, servers, multiple
- Question 7 **Systems**, data, change, **best practices**, business, implement, time, work, know, **procedures, complexity, complex**
- Question 8 Senior, developers, people, business, needs, person, team, **create**, needs, security, responsible, **best practices**, DBA group, owners, operations, staff, architecture, skills, **control**, drive
- Question 9 Understanding, **stakeholders**, different, **best practices**, teams, business, data, practice, people, things, **communication**, company, vendor, affect, design, cross, loss, security
- Question 10 Time, **strategic, plan**, change, business, **best practices**, plans, **database system, management**, goals, versions, vision, new, requirements, implements, thing, know, affect

## 5.5.2 Code Relations

TP2: Code Relations Chart

Code relations are coded sections of the text that are near or have a close proximity to each other. These interconnections can present patterns and relationships. To clarify the use of the terms used in this thesis: An interaction of components does not describe how they affect each other; an influence has direction. A influences B gives the direction of the influences such that A would be unchanged. That is unless B also influences A, in which case the influences are two directional i.e. they influence each other; when components are interconnected that may or may not interact. N.B. The direction of the arrows on the influence diagrams indicate the direction of influence.

The code relations are coded sections of the text which are related such that one code affects another, that is to say one code influences another. This analysis was done for all of the data corpus. The creation of the data map was an interactive process building on the codes, themes and data quotes, some of which are shown in Section 5.4.5.

An example of a few relationships can be seen in the relationship list below and Table 5.4.

**Relationship list**

BEST PRACTICE -> DESIGN  
BEST PRACTICE -> DATA  
BEST PRACTICE -> ENGINE  
BEST PRACTICE -> SECURITY  
BEST PRACTICE -> TECHNICAL

For example BEST PRACTICE -> DESIGN (read as best practice influences design) is derived from

*“best practices are going to guide everything that happens within it and all you are going to do is design what your best practices are as part of it” (Q10 2.1 line 34)*

Database design was discussed here in relation to the exact design of the operations and best practices used for start-up companies. Design, in the context of the research, is design of the tables and internal structures of the database. It is possible to interpret the quotes in more than one way. For the more ambiguous quotes, the researcher’s in-depth knowledge of the field was used as a basis for interpretation. One possible interpretation is that for the start-up case there is no

knowledge of how specific best practices will work in practice, and a good place to start is by implementing a selection of recommended best practices from vendors and community leaders. Thus best practice influences the design chosen for the database management system.

This relationship between best practices and design demonstrated there was an interconnection where best practices guides what happens. These relationships in the data corpus start to form the data map. Part of the data map displaying the connections from the data corpus is shown in the data map in Table 5.4

**Table 5.4 Data map - An example displaying influences between components from the data corpus**  
**Key: Codes = Components**

	Component B										
Component A	APPLICATION	DESIGN	DEVELOPMENT	DATA	CLOUD	ENGINE	SECURITY	TECHNICAL	BEST PRACTICE	~	TOTAL
APPLICATION			1			1		1	3		14
DESIGN				1				2	4		18
DEVELOPMENT						1					6
DATA					1	1	2	5	1		27
CLOUD							1	1	5		19
ENGINE							1	1	2		14
SECURITY				4	1	1		2			17
TECHNICAL	1	1		1	1				3		40
BEST PRACTICE		1		2		1	3	2			57
~											

The 'Component A' rows in Table 5.4 of the data map influence the 'Component B' columns. The relations between the codes show the number of times a pair of codes are connected throughout the data corpus. For the example; Application has three influences, one to Development (in the aforementioned relationship list), one to Engine and one to Technical. The Application row has a total of 14 influences in the data corpus from the components shown in the columns. Where more than one connection existed in the data, the total was increased in the relevant column (e.g. Data to Technical has a total of 5 interconnections). This could indicate that these

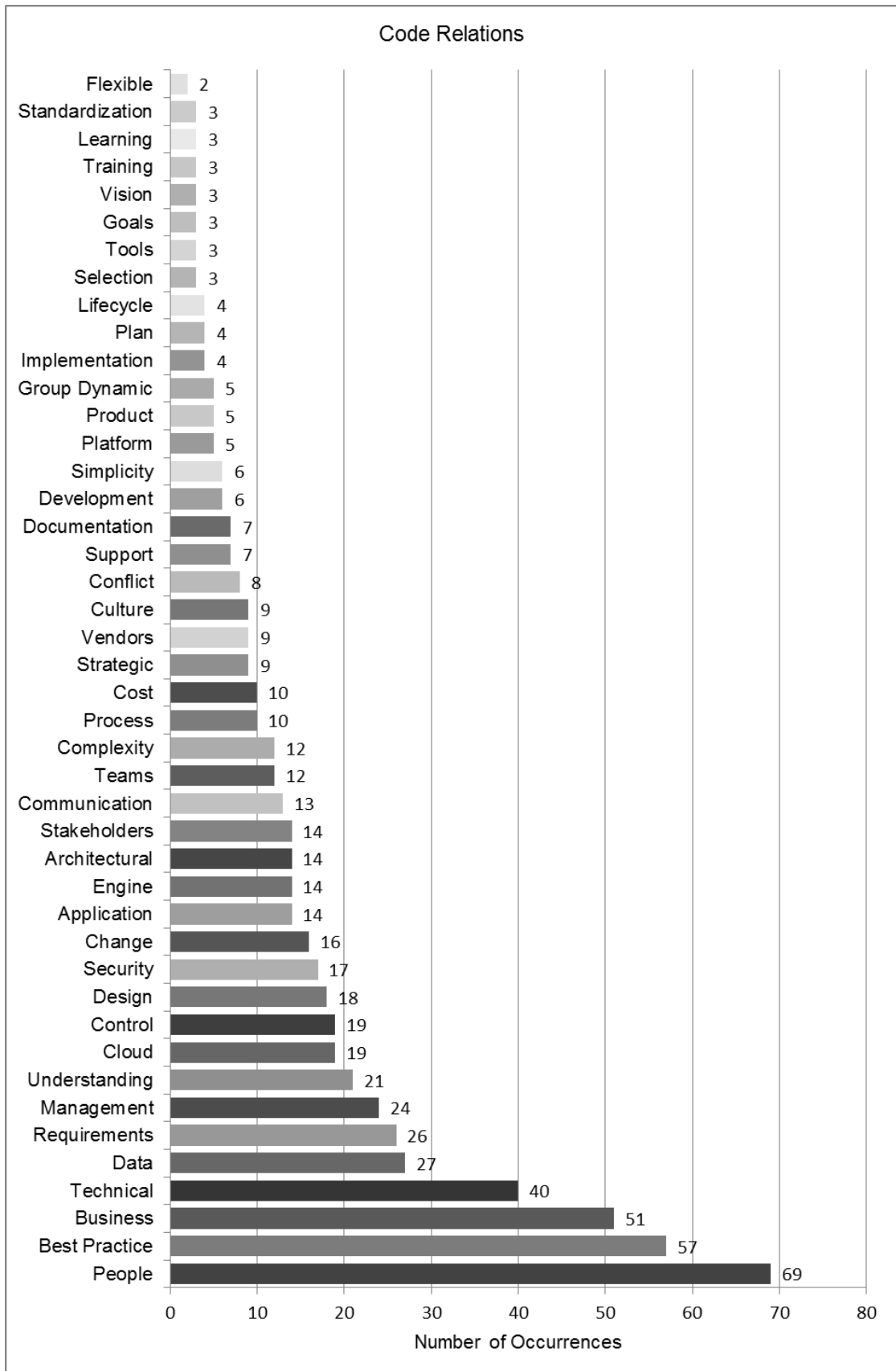
codes have greater importance. The data map has many duplicate interconnections – those with four and above interconnections are shown in Table 5.5.

**Table 5.5 Data map top influences**

	COST	TECHNICAL	DATA	SUPPORT	REQUIREMENTS	COMMUNICATION	CONTROL	BEST PRACTICE	COMPLEXITY	UNDERSTANDING	MANAGEMENT	PEOPLE
BUSINESS	4							4				
DATA		5							4			
SECURITY			4									
TECHNICAL				4					5			4
PEOPLE			4	4	6	5	4	8		5	4	
CONTROL		4						5				
UNDERSTANDING		4										4
COMPLEXITY		4										
DESIGN								4				
CLOUD								5				
CHANGE								5				
BEST PRACTICE											12	
MANAGEMENT								7				

The large number of influences indicated that these areas were effectively the most complex.

The code relations chart is included in Figure 5.9 to display the total counts of influences which are presented in the ‘Total’ column of Table 5.4, the Data Map. For example, the code relations chart has a component called ‘technical’, which influences 40 components as shown in Figure 5.9. It is possible to see visually the components that have the highest number of direct influences.



**Figure 5.9 Code relations**

The spray diagrams from each of the ten questions (Appendix E) and the structure of database systems outlined in Chapter 1, Figure 1.2, were reflected upon. The data analysed in the transitional process was then collected into meaningful groups, as shown in Table 5.6.

These logical groups were created from reviewing the data and the groups with the highest number of influences in the data map. These groups raised some initial questions, in particular whether 'best practice' could be in its own group. Best practice is the topic of the research and for this reason accounted for the high frequency of the words used. On review it was placed in the management group. The word 'data' for the same reason was used more often and at this stage was included in the technical group. An extra validation check was applied by looking at the 300 most frequently used words in the data corpus (see Appendix C). Although the word counts provide no real meaning they do offer some level of validation to make sure the analysis has not diverted from the raw data. The components were added to the top four groups: people; business; technical; and requirements.

The components were allocated to groups through a combination of methods: the general proximity of words using fuzzy matching, inflectional variant terms including singular / plural and a general meaning of some text. This was followed by a further process of iteration and reflection. The computer application used in this research was Microsoft SQL Server.

The findings from the data presented in the previous sections are all drawn together to summarise the most prevalent codes. The distribution of codes (Figure 5.5), code landscaping (Figure 5.8), data map (Table 5.5), code relations (Figure 5.9) were assigned to the relevant groups. This allocation route highlighted three areas: that management was more than just a component of one of the four groups; that

requirements wasn't a group on its own and it sat better in a group labelled 'architectural'; and that other groups became visible, especially App Dev, operations and knowledge. The findings from these sections are shown where the interconnections were more prevalent in Table 5.6.

<b>Business</b>	<b>Management</b>	<b>Knowledge</b>	<b>People</b>	<b>Operations</b>	<b>Architectural</b>	<b>App Dev</b>	<b>Technical</b>	
Business Cost	Best Practice Management		People Culture Control	Documentation		Development	Data Technical	Distribution of Codes (in 5 or more questions)
Business Company Plan Time	Best Practice	Know Think	People	Procedures Change	Systems Requirements	Design	Cloud Security Technology Data Storage	Code Landscaping (Data Corpus word counts above 42)
Business Cost	Best Practice Management	Understanding Complexity	People Control Communications	Change Support	Requirements	Design	Technical Data Security Cloud	Data Map (Inter-connections above 4)
Business	Management Best Practice	Understanding	People Stakeholders Control	Change	Architectural Requirements	Application Design	Technical Cloud Engine Data Security	Code Relations (Inter-connections 14 and above)

**Table 5.6 Findings from code landscaping, data map and code relations**



### 5.5.3 Operational Model Diagram

**TP3: Operational Model Diagram** The operational model diagram (Figure 5.10) helps disentangle the complex thread in the form of a network diagram. This is the final output from the transitional stage. It is a simple visualization based on relationships between the codes and is used as the transition from the real world to the systems world. The codes (components) have been grouped together into logical groups; the operational model shows connections between these groups derived from the codes and the data set. This shows an emergent sequence or network of codes to visually supplement the analysis. The groups have been named as systems, i.e. conceptual constructs based on the real world enquiry.

**Table 5.7 Data map systems summary showing the total number of interconnections**

	AppDev	Technical	Architectural	Operational	Business	People	Knowledge	Management
AppDev	1	6	1	4	4	2		12
Technical	2	23	4	15	17	15	8	38
Architectural	1	9	7	13	3	4	1	13
Operational	1	6	5	6	4	7		15
Business	3	12	8	10	15	12	5	14
People		21	15	19	16	33	13	37
Knowledge		5	4	3	2	8	2	3
Management	1	18	4	15	14	9	3	44

Table 5.7 is a summary of the total influences in each system to enable a visual display of the highest number of interactions in the system. The interactions between the systems are explored further in the systems map (Figure 5.12) and influence diagrams (Figure 5.21; Figure 5.22; Figure 5.23; Figure 5.24; Figure 5.25; Figure 5.26; Figure 5.27) that follow.

Out of the 8 systems, 4 systems do not influence each other which indicates a lack of communication:

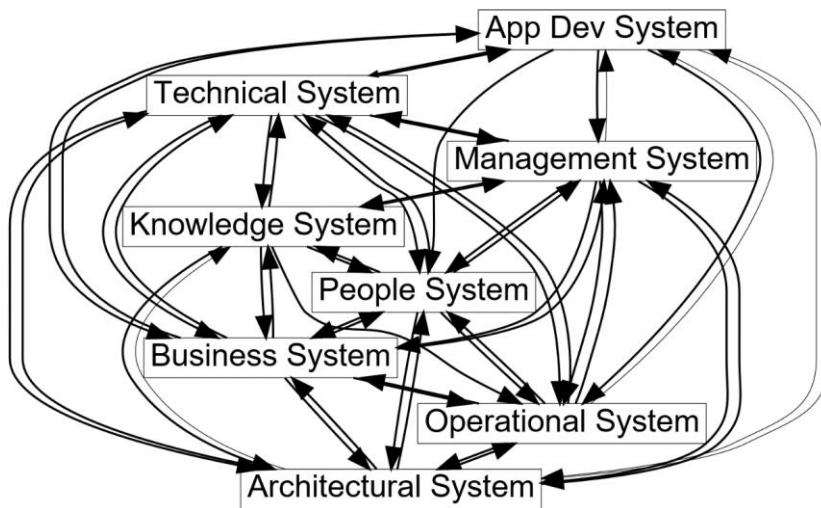
- The App Dev System has no connection to the Knowledge System.

- The Operational System has no connection to the Knowledge System
- The People System has no connection to the App Dev System
- The Knowledge System has no connection to the App Dev System

Five other systems only have 1 connection:

- The App Dev System has 1 connection to the Architectural System.
- The Architectural System has 1 connection to the App Dev System.
- The Architectural System has 1 connection to the Knowledge System.
- The Operational System has 1 connection to the App Dev System.
- The Management System has 1 connection to the App Dev System.

Figure 5.10 shows the weaker influences (1 connection) in lines shown in a non-bold font.



**Figure 5.10 Operational model diagram**

All the systems have some components within the system which are connected to other components within the same system. The operational model (Figure 5.10), based on the total number of influences, provides a visual representation of the

interconnections involved in the management of a database. Figure 5.10 also provides the initial view of the holistic analysis in a systems form.

This operational model shows that there is a one way connection only for the aforementioned systems. This lack of two way connections between some of the systems may be due to the lack of interconnectivity or to no direct feedback in the system. Feedback is where performance of the output is shared with the beginning of the process to enable it to be modified and improved. Feedback should occur through documentation and communication. However, from the data, there is no evidence how effective this is. Some systems connections have only one connection which showed a weaker connection between systems.

The transitional section has moved the understanding of the data from codes to components, looking at the data in a more holistic way. The data analysis has shown the key areas that were more prevalent in this data.

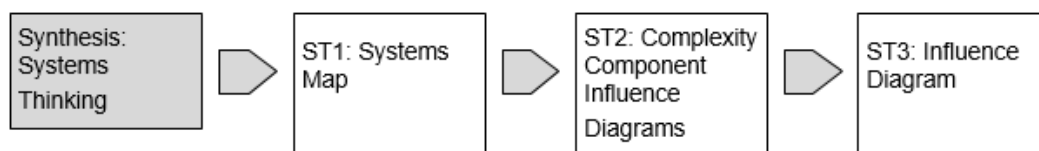
At the start of the transitional section, the word counts of codes in the data corpus were used for the word cloud. Then the codes' influences were recorded in the data map. The total number of influences per code was depicted in the code relations chart. These codes known as components were put into logical groups to produce the data map (Table 5.7). Table 5.7 shows the understanding gained from the transitional phase, showing the components with the highest number of connections. This culminated in the operational model which marked the transition from the real world to the systems world.

## 5.6 Synthesis: Systems Thinking

Within the thematic analysis the codes and themes were created from the data corpus. The transitional process began the transition from the reductionist approach

to a holistic approach. This synthesis, systems thinking stage, presents the results of the data and identifies the interconnectedness of the components. This research diverges and expands its scope from traditional thematic analysis to move from analytical research, where the data is considered as whole to be broken down, to synthesis thinking where the components are considered as a part of the whole (Ackoff 1981a, pp.16–17).

The systems thinking stages of analysis are shown in Figure 5.11 below.



**Figure 5.11 Synthesis: systems thinking**

To understand the switch in presentation of outcomes of the analysis I reiterate the following definition of a system:

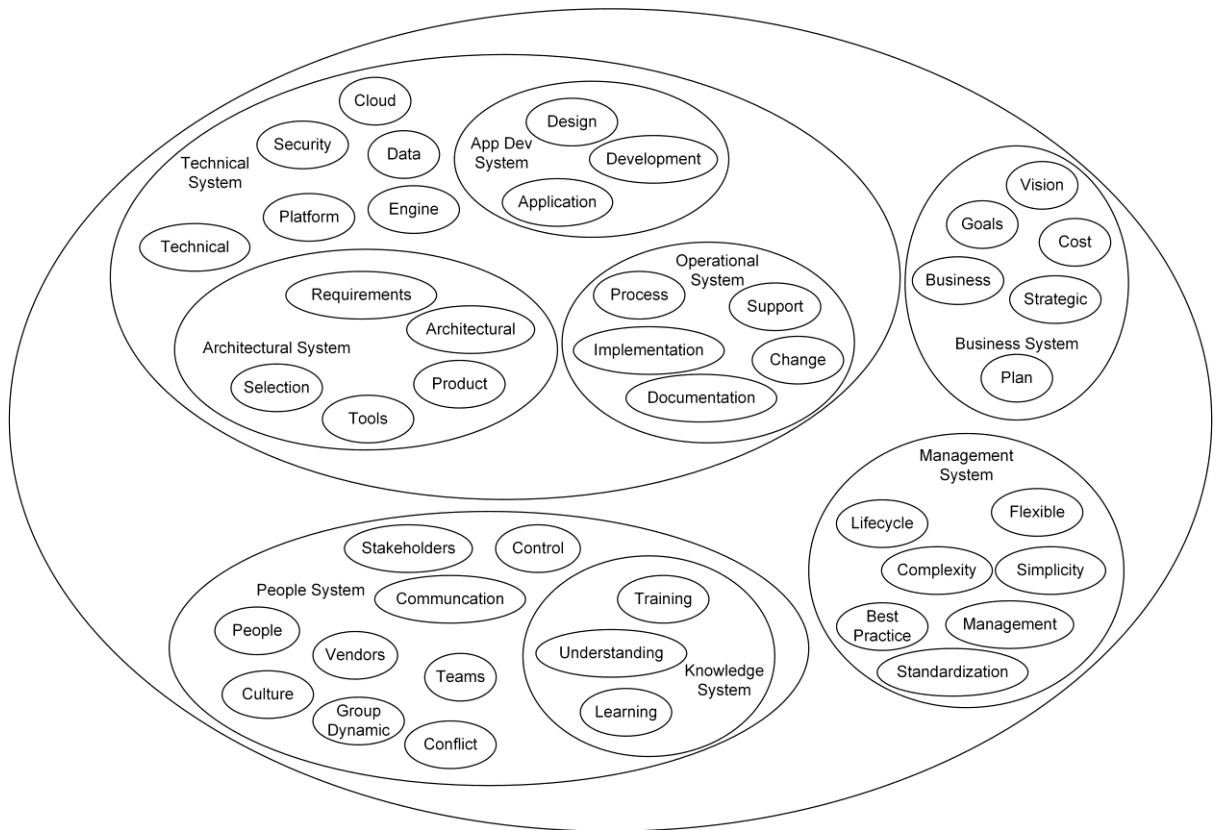
*“A system is a set of two or more elements that satisfies the following three conditions. (1) The behaviour of each element has an effect on the behaviour of the whole. (2) The behaviour of the elements and their effects on the whole are interdependent. This condition implies that the way each element behaves and the way it affects the whole depends on how at least one other element behaves. (3) However subgroups of the elements are formed, each has an effect on the behaviour of the whole and none has an independent effect on it.” (Ackoff 1981a, p.15)*

### 5.6.1 Systems Map of the Codes

ST1: Systems Map

A systems map (Bell *et al.* 2012) is a snapshot of the components of the system and environment. The systems map contains individual components and some groups of components. The system map diagram shows the structure of the components of the system to demonstrate the structural elements in a pictorial view.

Data from the thematic analysis highlighted significant areas of concern in the operation of databases today so a holistic view of the system was considered to offer a better understanding.



**Figure 5.12 Systems map of the management of database systems**

The systems map was generated by including all the codes in the data map (Appendix F) and placed these codes in the previously created logical groups. The

creation of the groups is explained in Section 5.5.2, as well as how the codes were allocated into the groups. The systems map (Figure 5.12) groups the components (formerly codes), into systems and subsystems. From the data presented in Table 5.6 and Figure 5.10, the analysis created eight systems. The eight systems were considered due to the significant gap in total number of interconnections. Those considered were the architectural system (51), operational system (44), App Dev system (30) and knowledge system (27). As an example, the architectural system (51 interconnections) is the sum of the components within it. The individual numbers of interconnections are: requirements (26); architectural (14); product (5); tools (3) and selection (3).

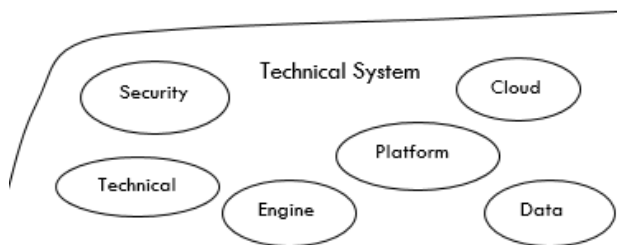
The outcome of this iteration through the data was the placement of those four systems within the larger systems of 'technical' and 'people'. This shows the components connected together within the boundary of the management of database systems. The management of database systems map consists of 4 main systems: technical system; people system; business system and management system. The technical system has three sub-systems, architectural, App Dev and operational. The people system has a knowledge subsystem.

The systems map helps to provide understanding of the management of database systems and how the components fit together. It is used in chapter 6 to provide insight on the structure of the system and the relationships between the main systems. To understand the situation it is important to understand the purpose of each of the components in the system. The systems and components are defined in subsequent sections.

## Technical System

The technical system contains all the elements pertaining to technology, hardware, software, architectural design, application development (App Dev) through to the operational delivery.

There are some components that are in the technical system but not in a subsystem.



**Figure 5.13 Technical system**

A definition of each component in this system follows:

- |           |  |
|-----------|--|
| Security  | The internal and external security requirements of data and databases.   |
| Technical | This component includes scalability, upgrades, backup configuration, technical layers (database, operating system, storage, antivirus, server backups, licensing). |
| Engine    | The database engine, for example SQL Server, Oracle, MySQL, MongoDB, other vendors.  |

**Platform** This relates to the database engine plus the operating system, plus the storage, plus relevant networks. It can also include the physical, virtual or cloud hardware.

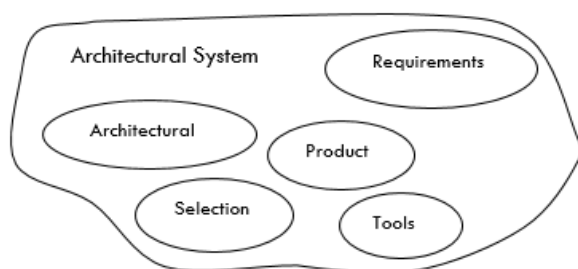
**Cloud** Shared computing resources that are not local servers that may or may not have predefined “database as a service” offerings.

**Data** Facts or raw information such as numbers, letters, structured and unstructured data.

## Architectural Subsystem

### 5.6.2 Architectural Subsystem

This subsystem contains all the elements involved in the technical, architectural design of the database systems.



**Figure 5.14 Architectural subsystem**

A definition of each component in this system follows:

**Architectural** This component relates to architectural design. The actual components of the database architecture which looks to



address availability, recovery, performance. Also architectural frameworks.

**Requirements** These are business requirements of the database design and the technical scope of hardware, software and teams to manage the system.

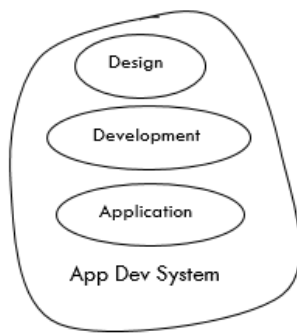
**Product** The features or capabilities of the database engines.

**Selection** All areas that are a part of how and why a database engine is selected.

**Tools** The tools that are available within the product or external to the products to help management of the database system.

### **App Dev Subsystem**

This subsystem contains components related to the design and development of databases that are used for special applications, some custom built, others as part of larger products.



**Figure 5.15 Application development (app dev) subsystem**

A definition of each component in this system follows:

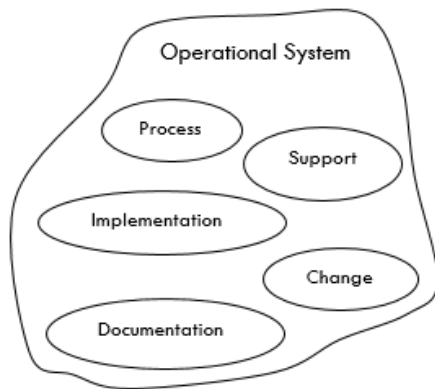
**Design**            The design of the data structure and database.

**Development**    The development languages and code components for creation, inserting or modification.

**Application**      These are related to databases from prebuild products like SharePoint and Systems Centre. Configuration is limited as databases should be treated as black boxes. Applications could also be custom built.

### **Operational Subsystem**

This subsystem includes all the elements connected with supporting the database systems once the databases are configured or installed.



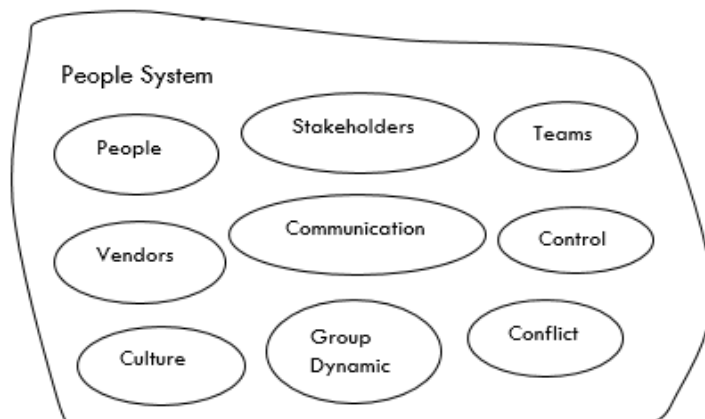
**Figure 5.16 Operational subsystem**

A definition of each component in this system follows:

Process	The processes involved in deploying and managing database systems and whether or not processes are followed.
Support	Providing support includes patching, deployments, alerting, monitoring, troubleshooting, taking backups, recoverability, performance tuning, disaster recovery and operational availability.
Implementation	Implementing changes, applying best practice, standard models, adding hardware or deploying software.
Change	Stability, rate of change, risk, managed change, planned and unplanned changes, adaptation strategy.
Documentation	Type of documentation, whether or not documented, control of runbooks, best practice, and documented process.

## People System

The people system looks at all the groups and individuals involved in the design and operation of database systems. It also looks at how those individuals interact and group dynamics within the organization.



**Figure 5.17 People system**

A definition of each component in this system follows:

People	This is the individuals who are a part of the system.
Stakeholders	Could be the business, suppliers, customers, vendors, public.
Teams	Could be database team, windows team, support team, development team, architect team, application team.
Vendors	Organizations such as Microsoft, Oracle, MongoDB.

Communication	This can be within team, cross team, outside the organization, difficulties, non-verbal, roles, hierarchical and horizontal.
Control	Who controls the components within the management of the database, database system, best practices used, the budget, structure and environment.
Culture	Reactive firefighting and proactive behaviour. Factors such as working in silos.
Group dynamics	Groups working together, business knowledge in groups, decision making, autonomous working, objectives, types of people.
Conflict	Within team, between teams, managers, territorial, budgeting, management, external conflict.

### Knowledge Subsystem

The system looks at training, understanding, learning and skills

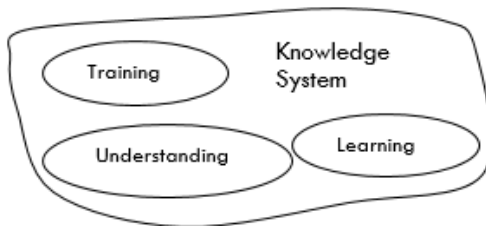


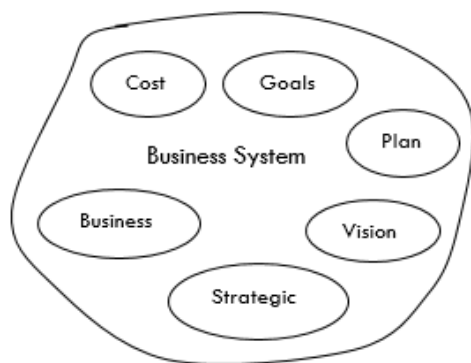
Figure 5.18 Knowledge subsystem

A definition of each component in this system follows:

- Training            Training provided if any and what type.
- Understanding    Know how the database engine works, how it fits in with other technical components. Comprehension of importance to database management, lack of understanding, strengths and weaknesses.
- Learning           Difficulties, speed of learning, knowledge transfer and skills.

### **Business System**

The business system focuses on the business and how it competes with other businesses, where the business is going and budgets.



**Figure 5.19 Business system**

A definition of each component in this system follows:

- Cost                Budgets, time to deliver, financials.

Goals	What needs to be done to meet the strategy and clarity on goals.
Plan	Lifecycle, disruptive technology, roadmaps within and external to organization.
Vision	What the business is trying to achieve with clarity and foresight.
Strategic	Competition in market place, organization threats, challenges, technological change, government regulations.
Business	Risk, changing business model, business purpose, reputation and constraints.

### Management System

This system looks at the components that form part of the entire lifecycle of the management of database systems. Best practices, complexity and efficiencies, standards and flexibility and simplicity.

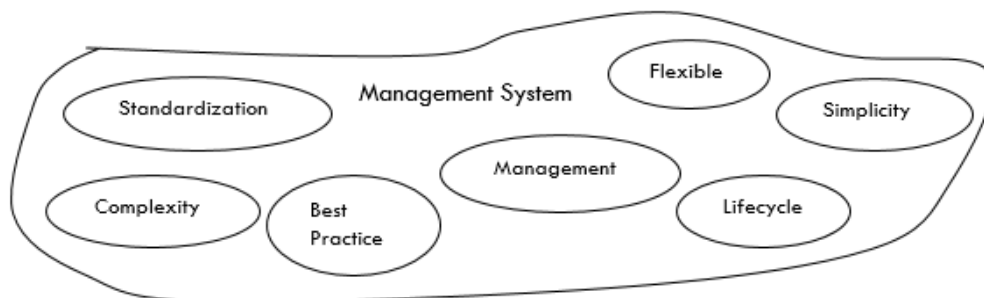


Figure 5.20 Management system

A definition of each component in this system follows:

Standardization	Productivity, automation, economy of scale, effectiveness, follow standards set, government regulations, open standards.
Complexity	Many perspectives, magnitude of components, many problems, interlinking components.
Best practice	Practices the respondents and participants thought were the recommended ways of doing something. Whether or not best practices were carried out, pitfalls and outcomes.
Management	Operational factors, technical factors, best practice management and creation, functional tasks, how to deploy strategic plans, centralised.
Flexible	Can easily be changed, grow and adapt to environment changes.
Simplicity	Simple architectures, ease of use, agile.
Lifecycle	End to end data and database management.



### 5.6.3 Database Management Complexity

ST2: Complexity  
Component Influence  
Diagrams

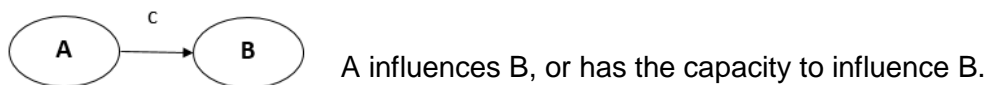
The systems map in Figure 5.12 showed a visual representation of the database management system.

This section builds on the systems map and elucidates the influences between the components.

This part of the thesis uses visual representations in addition to words, to illustrate the type of complexity in the database management system. These visual representations take the form of influence diagrams.

An influence diagram (Bell *et al.* 2012) is used to represent the main structural features of the database management situation and the important relationships that existed among them. It explores the interrelationships of the system and its components, or to express a broad view of how things are in the environment.

In this research, influence diagrams were used to present an overview of the areas of activity constituted within the database system that were required to manage the database lifecycle. They included the organization and the people and their main interrelationships. The notation used is given below.



C is a label that relates to the part of the quote from which the specific influence has been drawn.

The actual textual quotations from the transcripts are listed in *italics*.

The interconnections presented are based on six examples within the data set. These examples demonstrate: a component influencing the system; influences between two systems; and influences between systems and subsystems. It is not possible to present all the influences to this level of detail due to space constraints. The purpose of the study was to explore the complexity within database management systems.

The following areas were chosen as examples to examine in-depth complex interactions: best practice; management; technical and people; requirements and architecture; understanding, knowledge and skills; and business and change. The six areas were chosen for the following reasons:

- “Best Practice”, “Management” and “Technical” were chosen as key elements of the research. Best Practice was chosen as it is the main theme running throughout the research. The two systems with the highest number of interconnections from the data map in Table 5.7 were the management system (44 interconnections) and the technical system (38 interconnections). Within “Management”, “Best Practice” had the highest total (12 interconnections)
- “Business and Change” contained the next highest number of influences and changes, shown in Figure 5.9 (Code Relations).
- “Requirements and Architecture” was chosen as a group because it is one of the three subsystems in the technical system, shown in Figure 5.12.
- The “Understanding, knowledge and skills” were chosen because the knowledge system has a low number of connections between the other components shown in Table 5.7 and it is a subsystem in the people system shown in Figure 5.12.

In each of the 6 examples, common codes appear: “Technical”; “Business”; “People”; and “Best Practice”. “Data”, the next highest related code, appears in 5 of the 6 examples. ‘People’ has been used as a generic term to collect person-related aspects, such as comments about staffing, or when respondents used generic terms such as ‘some users’ or ‘without anyone driving it’.

These examples are interpretations of the influences from the quotations and shown in the diagrams.

### 5.6.4 Best Practice

There were numerous quotes referring to this aspect of the database system. The influence diagram (Figure 5.21) shows multiple interactions.

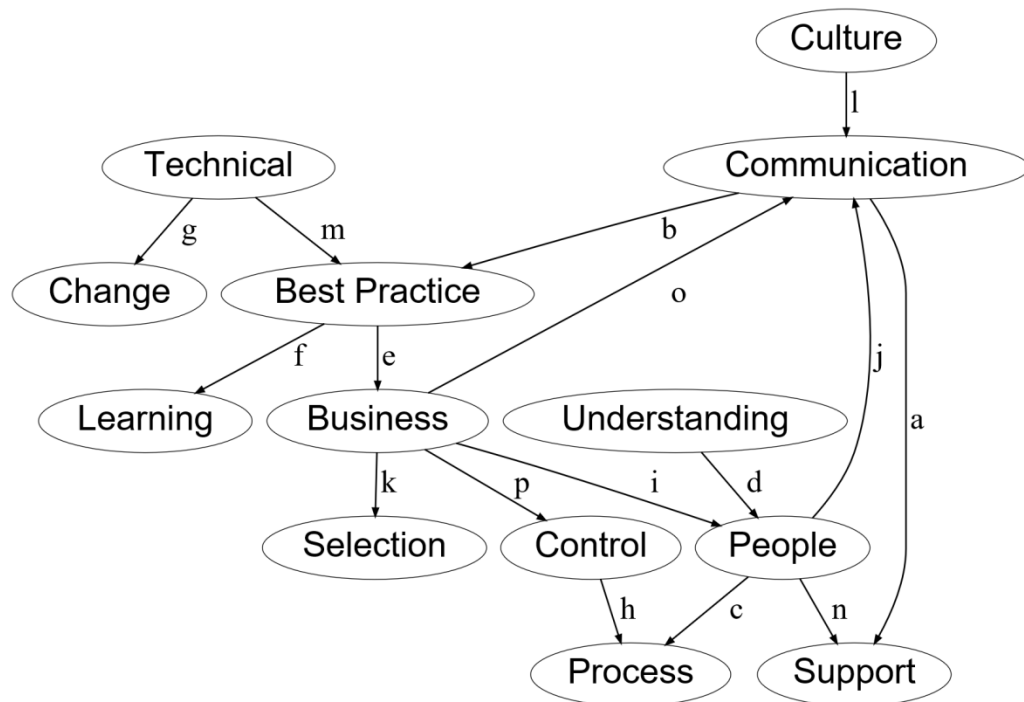


Figure 5.21 Best practice influence diagram

Best practice is reliant on people defining process and communicating. Control of processes helps prevent mistakes and best practices reduce risk, improving supportability through the reduction in ‘firefighting’. Best practice can help with

business continuity if people know there is a process. Although communication is important, operational teams who provide support are often “the last people to know”. The influence diagram, Figure 5.21, is explained using quotes from the transcripts:

*“Supportability operations (n) are the last people to know (a). The business leads (o) and drives what is selected (k). IT lost out on control (p) to the business. It is hard for the IT department and reaches them too late. (i)” (Q2 5.4 line 58)*

Link ‘n’	people ► support	People can influence when operational support teams are informed.
Link ‘a’	communication ► support	Lack of communication influences the support of database systems.
Link ‘o’	business ► communication	The business influences what is communicated to the people.
Link ‘k’	business ► selection	The business can influence the selection of new systems required.
Link ‘p’	business ► control	The business can decide what software products are selected possibly due to cost or what is currently in use.

Link 'i'    business ► people    IT people can be affected by the lack of control and influences by the controlling owner.

*“If you don’t have any communication (j) you tend not to have any best practices or procedures (b)” (Q9 3.4 line 40)*

Link 'j'    people ► communication    Communication could be influenced by the people, business or culture.

Link 'b'    communication ► best practice    A lack of communication could influence whether best practice or procedures exist.

*“A customer may not have any processes in place to account for a scenario (like DR!), or may have a process (c), but the people do not know the process. (d)” (Q5 8.2 line 106)*

Link 'c'    people ► process    People such as customers influence what processes exist for difference scenarios.

Link understanding ► A lack of understanding influences whether  
'd' people people know the process and can use it.

*"In general, best practices reduce exposure to risk (e), minimize firefighting (l), and ensure strong performance and easier recoverability / business continuity." (Q4 7.4 line 75)*

Link best practice ► Best practice can influence the level of risk the  
'e' business business has for things such as database performance and business continuity.

Link 'l' culture ► The culture within the business could be that of  
communication firefighting to correct errors rather than being proactive and this could influence what is communicated to people working on database systems. Also the culture of silo system means little communication of best practices.

*"In mature databases, best practices and procedures avoid relearning mistakes (f). Unfortunately, the best practices and procedures have to remain agile because hardware (g) and software continues to change." (Q4 8.1 line 78)*

Link best practice ► Best practice and procedures can influence learning  
'f' learning by preventing relearning mistakes.

Link technical ► Technical advances influence the volume of change  
'g' change technical hardware and software continual change  
influences the best practice and procedures which  
should be agile.

*“without a level of control people make mistakes and things can get missed whether it is an automated process (h) or some companies make a choice ours is the installations (i) whatever the other thing is you can have like you said at the beginning a PowerShell script that deploys a database to a specific standard that is used and it adheres to best practice (m)” (Q8 2.1 line 30)*

Link 'h' control ► process Control is required to manage process and a  
lack of control influences process.

Link 'i' business ► people The businesses influence the choice of what is  
or is not automated.

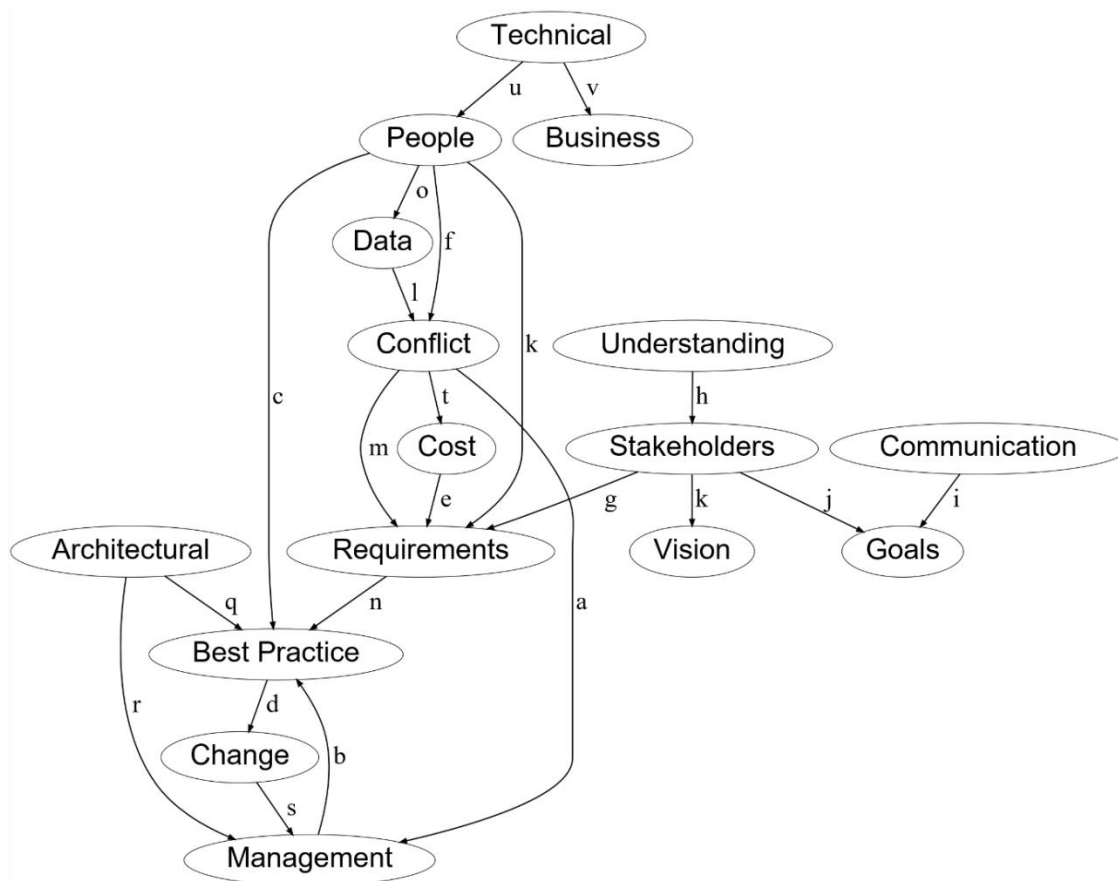
Link technical ► best The technical design influences best practice.  
'm' practice

The influences on best practice illustrated by the above quotes from the focus group data give evidence of the effect on database operations. Best practice is shown to be reliant on good communication between all the actors, that is, customers, stakeholders and people involved in operational teams. As well as communication, the culture that exists in business is affected by the way control manages processes

such as best practice, disaster recovery and change. Understanding of these processes and the ability to use them by the operational teams is important. Best practices can prevent relearning mistakes. The technical design and whether or not a process is automated may be controlled by the business or database managers and affects processes such as best practice.

### 5.6.5 Management

The following quotes have been used to construct the influence diagram shown in Figure 5.22.



**Figure 5.22 Management influence diagram**

Conflict occurs in many ways and is a key area for management. This can start from stakeholders having their own requirements and agenda. This can often be countered by the suppliers or business trying to meet expectations such as cost.



Conflict caused by operators can spread into the actual data causing loss of confidence in it. Conflict can occur between teams, database specialists and admins. Best practice may not apply between layers.

*“Always look through a critical lens at ‘best practices’, some may not be very applicable to your particular use case whilst other may be very relevant (u). There may also be a conflict (a) between ‘best practice’ for database specialists and those for Admins: Little point in having a system that is perfect from a database perspective but cannot be maintained. (b)” (Q1 1.5 line 5)*

Link      conflict ►  
'a'      management      The defined best practice used for management of database systems can be influenced by disagreement and conflict between what management is required by other administrators.

Link      management ► best  
'b'      practice      Management and the system influence the type of best practice selected, as the use case needs to be relevant to the action or steps undertaken.

Link      technical ► people  
'u'           Technical configuration for databases can be different to other technical configurations, which can influence how people behave and interact.

*“I'm a consultant who gets called in when the server's (v) on fire. It's not reliable enough or fast enough (d). Because of my job, I can't blindly implement best practices (c). Any change inherently carries risk. (s)” (Q7 8.1 line 115)*

Link 'd'	best practice ► change	Best practice influences the outcome of a change which aims to improve server performance.
Link 'c'	people ► best practice	People influence whether or not best practices are implemented.
Link 's'	change ► management	Any change may influence management of the database server as reliability and speed may or may not be fixed.
Link 'v'	technical ► business	Technical failure influences the business to bring in specialists to resolve the problems.

*“Often we need to cover ourselves as the customer (f) may have conflicting (t) requirements around cost (e) and time and we need to make sure that they are aware that one constraint may affect other system qualities (r) such as best practices (q) and quality risks (n).” (Q6 8.2 line 103)*

Link 'f'	people ► conflict	Different people influence whatever conflict arises.
Link 't'	conflict ► cost	Conflicting requirements could influence the cost and time spent depending on the business choices made.

Link 'e'	cost ► requirements	Cost may influence requirements as only certain options might be possible.
Link 'r'	architectural ► management	Architectural constraints influence the type of management required to be carried out.
Link 'q'	architectural ► best practice	Constraints placed on the database system architecture could influence best practice.
Link 'n'	requirements ► best practice	Constraints placed on the database system requirements could influence best practice.

*“Stakeholders have their own agenda (k), have their own requirements (g), their own job description. To achieve that, not be in line with your aims, important to communicate (i) with all of them, understand their requirements (h) to protect yours (m) , get a consensus, communication to meet goals (j)” (Q9 3.1 line 43)*

Link 'k'	stakeholders ► vision	Stakeholders influencing the vision as they have their own agenda.
Link 'g'	stakeholders ► requirements	Stakeholders in the database system influence the requirements by specifying

what they want the system to do, achieve or performance metrics.

Link 'i' communication ► goals Communication can influence how the goals are met.

Link 'h' understanding ► stakeholders Understanding, influences how people protect their requirements and ensure stakeholders' requirements can be met.

Link 'm' conflict ► requirements Stakeholders handle conflict to protect their requirements.

Link 'j' stakeholders ► goals Stakeholders having or setting their own goals to achieve, but consensus by communication is suggested.

*“This is the classic case of how applications turn rogue. Stakeholders must be united in the vision (k). If there are rogue operators creating customised interpretations on the data (o) that conflict with the core data set confidence (l) in both systems is significantly impacted.” (Q9 5.5 line 63)*

Link 'k' stakeholders ► vision Stakeholders influence the vision by being united in their goals.

Link 'o'    people ► data    People in the form of rogue operators influence the data quality and whether it exists.

Link 'l'    data ► conflict    Data can influence conflict if it is wrong or has a lack of quality.

The selection of quotes from the qualitative data relating to management are illustrated in the influence diagram (Figure 5.22). There are a number of people involved in setting up and operating a database and the best results overall are probably achieved via management deciding on best practice (link b). Initial requirements and the configuration of a database system have a number of constraints. These can be cost, architectural, data quality and overall understanding.

### 5.6.6 Technical

Using the following quotes led to the influence diagram shown in Figure 5.23, an important part of the database system.

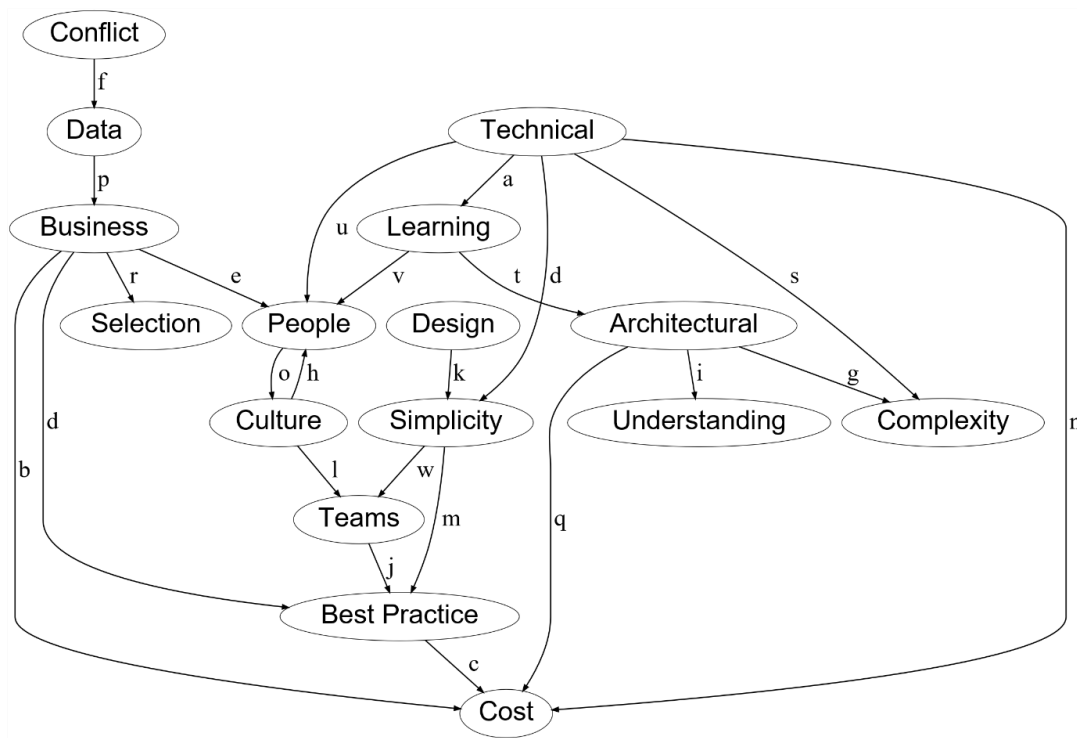


Figure 5.23 Technical influence diagram

The selection of database engines can be influenced by frameworks that are too complex or take too long to set up. Simplicity is a core factor for implementation to meet business goals, with price an influencing factor. Learning difficulties and staffing challenges are problematic in addition to the personal differences of the business and people. Simplicity can mean many technical layers, different teams and different best practices. Teams need to work together to achieve the goals rather than work against each other.

*“With that in mind, everybody who works with these things is dancing at the edge of their comfort zone (o) and beyond. The staffing challenges (u) and learning difficulties (v) are the biggest problem facing databases.” (a) (Q5 8.1 line 101)*

Link 'o' people ► culture People's level of understanding and ways of working may not be up to date.

Link 'u' technical ► people Finding people with knowledge of the current technology is hard.

Link 'v' learning ► people Learning influences people's behaviour in implementing new ideas.

Link 'a' technical ► learning Technology chosen influences what is required to be learnt or the technology in use influences what people have to learn.

*“most of my work involves finding the fastest, cheapest, (b) easiest compromise to implement in order to accomplish the business goals. The business has to be able to make money and avoid loss - and unfortunately, a lot of best practices and procedures (c) ignore costs. If we all had unlimited time, manpower, and money, we'd all build systems according to best practices (d), but like Steve Jobs said, artists ship.”*

(Q7 8.1 line108)

Link 'b' business ► cost Meeting the business goals influences the need to find fast cheap options.

- Link 'c' best practice ► cost ► Best practice influences cost and doesn't necessarily produce profit. Not using them may help to keep costs down.
- Link 'd' technical ► simplicity ► Technical designs influence simplicity. Best practice may require unlimited money.

*“which leads to more discussion and probably a slower implementation of best practices and probably a compromise on what is best practice. But at the end of the day it needs to be a decision that is made by the business (e) in the best interest people have to learn to put aside their personal differences and essentially the data is there for someone to use (f) it to keep the company going or keep the project going to sustain whatever enterprise (r) is built around the data (q)” (Q9 6.1 line 66)*

- Link 'e' business ► people ► The decisions of the business that influences people's interests.
- Link 'f' conflict ► data ► Conflict of personal differences influencing whether data is used.
- Link 'r' business ► selection ► Selection by the business can influence the technical data.
- Link 'q' architectural ► cost ► The architecture design around the data that influences company profitability.



*“it all links into the requirements and how they are met. Basic complexity is increased just by adding a layer (s). I’ve heard of requirements analysis, is there such a thing as implementation analysis? Verbally added: Adds complexity to implementation. You need requirements analysis for implementations. Should there be top down design approach for the layers? Easy to add constraints of the other layers at the beginning, harder at the end (q). It is ivory towers (h), you can’t ignore the real world and all the components and people (i). Why is there no time to implement best practices (c), don’t know current frameworks (t), current frameworks are too complex (g) to use or too time consuming to implement. (n)” (Q7 1.5 line 7)*

- Link ‘s’ technical ► complexity      Could be the number of technical layers in the architecture increasing the complexity of the tasks.
- Link ‘q’ architectural ► cost      Could be architectural frameworks are influenced by cost as they are time consuming to deploy.
- Link ‘h’ culture ► people      Could be the ivory towers, that exclusivity of singular working, vision and rigid structure influence people being ignored.

- Link 'c' best practice ► cost      Could be best practice adoption or simply the time and cost could prevent best practice being utilized.
- Link 'i' culture ► teams      Could be culture of the ivory towers that influences how the teams behave.
- Link 't' learning ► architectural      Could be no learning of framework occurs which influences the architectural design.
- Link 'g' architectural ► complexity      Could be the architecture influences the complexity and the cost, with the time increasing to implement the framework.
- Link 'n' technical ► cost      Could be the technical frameworks complexity influencing the implementation due to length of time taken.

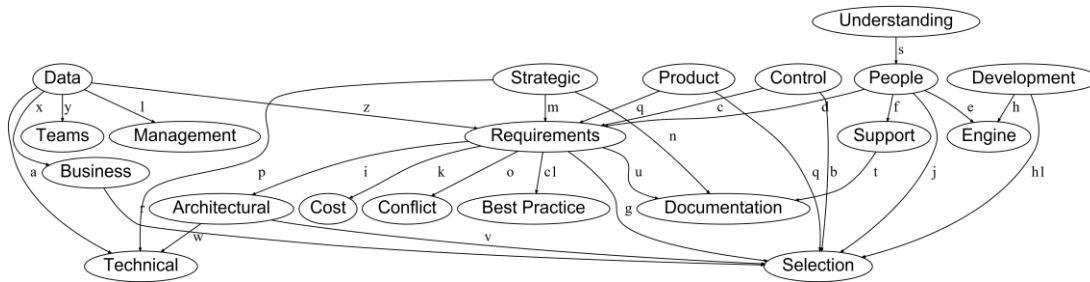
*“Simple fact (w) you have layers (k) - different teams (j)- differing best practice (m)- can end up with teams pulling (l) rather than working together. ” (Q5 4.1 line 70)*

Link 'k'	design ► simplicity	The simple fact of the design that can influence how many different teams have to work together.
Link 'j'	teams ► best practice	The design of the layers could influence simplicity.
Link 'm'	simplicity ► best practice	The simple fact that there are layers that influences the number of best practices.
Link 'w'	simplicity ► teams	The simple set up and multiple layers that can influence how many different teams have to work together.
Link 'l'	culture ► teams	The lack of simplicity could influence the business outcome with teams pulling rather than working together.

People have to cope with fast changing technology and complex technical layers in the architecture which need to be learnt if database systems are to improve. The goals and aims relating to business, technical, cost and best practice can usually only be achieved by compromise. Simplicity of technical designs are suggested. Personal differences should not affect business decisions. The profitability of businesses will influence selection of data, the architectural cost and possibly, some of the people. A culture of rigid structure and singular working cannot work well because databases operate with specialist teams due to the complexity of individual systems.

### 5.6.7 Requirements and Architecture

The interrelationships between these complex interrelationships were drawn from the following quotes and subsequently illustrated in Figure 5.24.



**Figure 5.24 Requirements and architecture influence diagram**

Data is a core component that draws together technical components (hardware and software) in conjunction with the business. The requirements define linking components such as architecture through to support. The customers are sometimes unrealistic in their expectations and the requirements are reduced. Architecture skills are required in the engine. The need for skills in the business sometimes mean existing known architectures and technical engines are selected.

*“Coordination is essential (y) since the data system is the touch point (z) between the hardware (a), software, and business teams. (x)” (Q9 7.6 line 82)*

Link ‘y’ data ► teams

Data being a factor that influences the business teams.

Link 'z' data ► requirements Data being a factor that influences the requirements for the coordination.

Link 'a' data ► technical Data can influence the hardware and software.

Link 'x' data ► business Data influencing the core business.

*“Technology selection should be performed within the context of an organisation's Enterprise Architecture (b). Technology choices (c) may be made for tactical reasons (p) (in the short term) or to align with the strategic applications portfolio, but procurement should be controlled to avoid unnecessary proliferation of disparate technologies (w). Assessment of product capabilities against the required capabilities, etc. (q).” (Q2 1.2 line 2)*

Link 'b' control ► selection Control within the business influencing the selection to prevent a proliferation of disparate technologies.

Link 'c' control ► requirements Control of technology choices influences requirements for tactical reasons.

Link 'p' requirements ► architectural The requirements of the product capabilities influence the architecture design.

Link 'w'    business ► selection    The business influences requirements to prevent a proliferation of disparate technologies.

Link 'q'    product ► selection    The product capabilities influence the selection based on what features the business requires.

*“They pare back requirements to what the customer actually needs from the wish list. (d)” (Q3 5.2 line 70)*

Link 'd'    people ► requirements    The customer and business people influence what requirements are actually needed for the business purpose.

*“In practice, I've found that the selection (j) of engine (e) often follows considerations such as package support (f), enterprise diktat (r) or in-house development platform (h). Also, most database engines to offer similar mainstream feature sets (h1) and allow most site policies to map onto them pretty well. An important factor will always be whether skills (s) to support a particular engine already exist on-site (g).” (Q2 7.1 line 74)*

Link 'j'    people ► selection    The engine selection will influence and be influenced by the support provided or what is available within the organization.

Link 'e'	people ► engine	The support provided within the organization influences the engine chosen.
Link 'f'	people ► support	The people can influence the support given.
Link 'r'	strategic ► technical	The business may have a strategic position that it imposes control over the technology used.
Link 'h'	development ► engine	Development work in house may be carried on a specific engine due to in house expertise or the hardware already in use must be used as the business may not have the funds to by a new hardware to support a new engine.
Link 'h1'	development ► selection	Development could be influenced by the feature set included with the product and want to use a particular function.
Link 's'	understanding ► people	People in house may already have the skills for working on a particular engine or may need to invest time in learning the new engine.
Link 'g'	requirements ► selection	The skills may influence what product is selected as the business might only want to

consider engines that the staff currently have expertise in.

*“We also need to re-evaluate existing thinking and apply today’s architectures to it. I think many solutions being built today ignore many of the modern advances in hardware and software (j) – people stick with what they are comfortable with (v) and have built before rather than architect (i) towards more modern application patterns” (Q3 5.5 line 78)*

Link ‘j’    people ► selection    People’s knowledge influence the selection of modern architectures.

Link ‘v’    architectural ► selection    The constrained architecture choices available influence the selection of engine and solutions.

Link ‘i’    requirements ► cost    The new requirement application patterns are not yet understood by people within the business and this influences the budget and time to deliver any new database systems.

*“Often we need to cover ourselves as the customer may have conflicting requirements (o) around cost (k) and time and we need to make sure that they are aware that one constraint may affect other system qualities such as best practices (c1) and quality risks. (l)” (Q7 8.2 line 111)*



- Link 'o' requirements ► conflict The requirement may influence conflict and cost with the time required to deliver a system increasing.
- Link 'k' requirements ► cost That certain requirements such a time influence cost as more people may be required to deliver the project within the timescales.
- Link 'l' data ► management The quality of the data externally and internally influence how database management is undertaken.
- Link 'k' requirements ► best practice Having multiple requirements may influence delivering high quality database systems as they take time and people.

*"In my "run book" article, I state that it's strategic (n) to know the purpose (u) of the system and its ongoing requirements (m) to properly manage (l) a data system (t)." (Q10 7.6 line 106)*

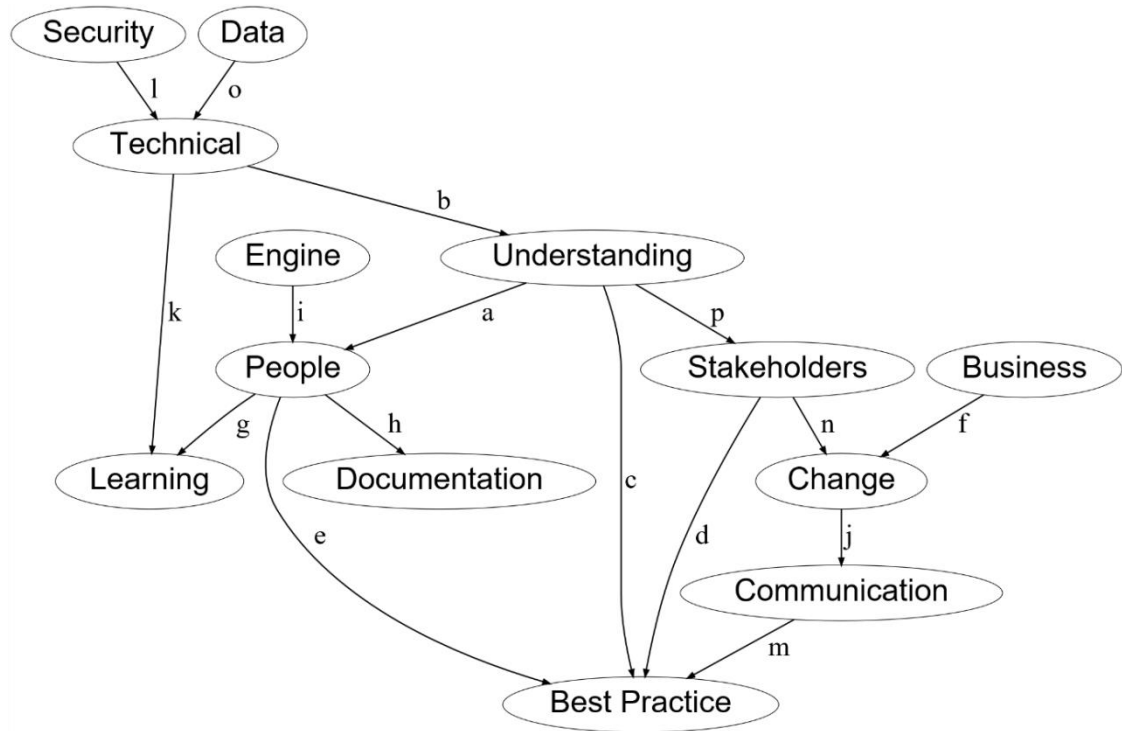
- Link 'n' strategic ► documentation The strategy deployed and purpose of the system could potentially influence the documentation produced or the use of standard documentation in existence.

Link 'u'	requirements ► documentation	The requirements influence what is documented.
Link 'm'	strategic ► requirements	Strategy influencing what is required for management.
Link 'l'	data ► management	The data in the system influences what needs managing and supporting.
Link 't'	support ► documentation	The support documentation of the data systems influences what is documented based on the needs of the operation.

The data in the system is a key influencing factor for technical, business, the teams and potentially an influencing factor for the architecture design. Control influences requirements such that suitable architectural design is set up. Control by influence selection can override choices made by business or product selection where necessary for cost or efficiency. The customer/ business requirements are also influenced by the constraints of what is possible. Thus in some cases conflict can arise. People influence the support that is available within the organisation and affect the selection of the engine and product when new skills may need to be learnt and understood. Also new developments offset selection. Business Strategy (strategic) can influence the technology used. People influence architecture design by potentially staying with what they already know. Multiple requirements could influence best practice. Strategy, requirements and support for the operation should be evident in documentation. Data basically influences what is to be managed.

### 5.6.8 Understanding, Knowledge and Skills

The quotes give rise to complex interactions between understanding knowledge and skills and are a further part of the database system and are shown in Figure 5.25.



**Figure 5.25 Understanding, knowledge and skills influence diagram**

Issues can occur when best practices are not understood. Also a lack of access or knowledge of the infrastructure can cause issues. The environment connected to the database can determine the best practices. Understanding staffs' strengths and weaknesses can be an asset because sometimes they want to challenge their skills which can affect the quality of the documentation. A lack of understanding by stakeholders shows why certain best practices are required.

*“Understand the current staff's strengths and weaknesses (a) - for example, their comfort level with existing database engines (i), and their ability to learn new ones.(k)” (Q2 8.1 line 89)*

Link 'a' understanding ► people      The level of understanding may influence the staff strengths and weaknesses.

Link 'i' engine ► people      Engine types can influence how the staff perceive their comfort with the product.

Link 'k' technical ► learning      People's technical ability can influence whether or not they can learn new engines.

*"The lack of access (l) or knowledge (b) about the infrastructure under the database" (Q6 8.3 line 104).*

Link 'l' security ► technical      The infrastructure technical components can influence the levels of understanding and knowledge about the database system.

Link 'b' technical ► understanding      Security access influences what technical work can be carried out or skills obtained.

*"I think understanding your environment will dictate what best practices (c) are if your financial and accuracy is important. I hold a lot of data including paediatrics medical records and availability and if that went down for 2 days it is not the end of the world but if we lost it that part of it is a massive disaster.(o)" (Q1 3.1 line 86)*

Link 'c' understanding ► best practice      Understanding the environment can influence what best practices are deployed.

Link 'o' data ► technical      Be the data is important to the business and influences the technical components used.

*“Poorly understood can best practices (c) make it all worse. Best practices should be treated with care (e). They are not universal truths but guidelines (think: maxims) (m)” (Q9 7.5 line 81)*

Link 'c' understanding ► best practice      A lack of understanding could influence best practice used or selected.

Link 'e' people ► best practice      People working with best practice can influence the quality of best practice.

Link 'm' communication ► best practice      Communication of the purpose of best practice could influence how they are used.

*“Database staff want to move on to bigger and better projects (g) that challenge their skill levels, not keep redoing the exact same task and perfecting the best practices and procedures. As a result, the people writing (h) an organization's procedures are usually winging it.” (Q8 8.1 line 102)*

Link 'g'      people ► learning      People wanting to continue to improve which could influence what they learn next.

Link 'h'      people ► documentation      People influence the quality of the documentation as they may not understand everything they document.

*“If some stakeholders aren't aware (p) of why certain Best practices are being enforced (d) then it can often be misinterpreted (j) as 'DBA never lets us do anything', or seen as foot dragging from IT (n) whenever a business change is required. (f)” (Q9 5.4 line 61)*

Link 'd'      stakeholders ► best practice      Stakeholders' lack of knowledge may influence why best practice is misinterpreted.

Link 'j'      change ► communication      Changes may influence how the database administrators are perceived with possible misinterpretations as a “DBA never lets us do anything”.

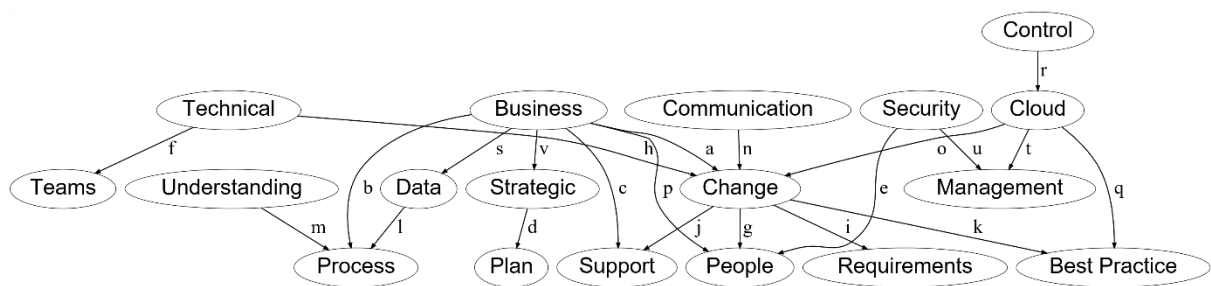
Link 'n'      stakeholders ► change      Stakeholders could influence change by misinterpreting the business change or by not carrying out the business change.

Link 'f'	business ► change	The business can influence the stakeholder's tasks when changes are required.
Link 'p'	understanding ► stakeholders	The level of understanding influences the stakeholder's actions because the stakeholders may not know why certain best practices are enforced.

Technical factors may influence learning when new technical features are to be introduced. People's level of understanding may be need to be updated and the engine to be deployed influences which people are suitable for the job. The security infrastructure influences technical factors which in turn influence the need to understand. Understanding communication affects people's use of best practices. Data needs to be secure hence technical components must be suitable. People may influence their own learning. People influence documentation which if comprehensive can improve practices. Stakeholders can influence practices and affect change but are only influenced by their understanding. Change should influence communication for people to accept its necessity.

### 5.6.9 Business and Change

Figure 5.26 depicts change and management plans.



**Figure 5.26 Business and change influence diagram**

Business models are sometimes affected by rapid change and the support requiring people and processes may not be able to keep up. More teams need to be consulted for changes. Exacerbated by the number of increasing technical layers all best practice need a level of review as the change carries risk.

*“Strategic plans (d) may defer to immediate tactical needs (v), and the longer term value may be less obvious to the people on the ground so compliance (e) may need to be managed differently (u).” (Q10 1.2 line 2)*

- |          |                         |   |
|----------|-------------------------|---|
| Link 'd' | strategic ► plan        | Strategic and tactical needs may not be aligned in the present by strategic decisions and influence the plans undertaken. |
| Link 'v' | business ►<br>strategic | The business can influence the strategic plans developed  |
| Link 'e' | security ► people       | The security set up influences people's allowable actions whether they understand this long term aim or not.              |



Link 'u' security ► The security needs influence the management  
management required to be undertaken.

*“The business model may be changing (a) too rapidly for the processes (b) and people (p) that are in place to support them. (c)” (Q5 8.2 line 108)*

Link 'a' technical ► The technology changing too rapidly for the teams to  
teams keep up.

Link 'b' business ► The business influencing how quickly processes  
process need to change.

Link 'p' business ► The business can influence the work people need to  
people do.

Link 'c' business ► The business can influence the support that can be  
support provided.

*“As more layers get added (DBA, Network, SAN, Virtualisation, Cloud, BI) to DB project the more teams (f) or individuals (g) need to be consulted for any proposed change. (h)” (Q5 5.4 line 83)*

Link 'f' technical ► Technical components added in the way of layers  
teams influence the need to consult teams and people.

Link 'g' change ► people Changes can influence both the teams and people in their decisions made during consultation.

Link 'h' technical ► change New technical architecture components influence changes made to the database system.

*“Personally speaking, we have a general strategic IT plan of which database management is a part as it should be. But such plans tend to be fairly static if not abstract; change happens quickly and new requirements constantly arise (i), so operational (j) or tactical factors are of much more concern. I suspect that this situation is not uncommon. Given this situation, I find it best to concentrate, place the focus of best practices and procedures (k) on those factors.” (Q10 7.1 line 101)*

Link 'i' change ► requirements Change influences existing plans and requirements.

Link 'j' change ► support A key factor is the change influencing support and how those operational or tactical factors are to be handled.

Link 'k' change ► best practice Change influences best practices and procedures and these could be focused on the help needed to manage the change.

*“Too often silo systems are built because an organisation has no visibility (m) of the extent of the processes (b) it has – and therefore where the data (s) is actually used.(l)” (Q3 5.1 line 77)*

Link ‘m’    understanding ► process    People create silos which influence the lack of visibility of process.

Link ‘b’    business ► process    Could be the organization has no visibility and has no understanding of process.

Link ‘s’    business ► data    The business influencing what data is required for its normal operations.

Link ‘l’    data ► process    Data within the system influences the process depending on where it is used.

*““Cloud” usually means “someone else’s black box.” (r) You’re playing by someone else’s rules, and since they keep changing their systems (t), they’re changing the rules (o), too. It’s hard to build best practices (q) when the underlying mechanisms are evolving so rapidly (p) and you’re not privy to the changes.(n)” (Q6 8.1 line 100)*

Link ‘r’    control ► cloud    Control is limited in influence for cloud provider database management systems.

Link 't' cloud ► management With the technology component changing rapidly in the cloud these influence what type of management can be provided.

Link 'o' cloud ► change The cloud products and services are changing that influence the changes that are required to successfully manage database systems.

Link 'q' cloud ► best practice The management of someone else's hardware such as cloud influences what best practices are required.

Link 'p' business ► people The business not being privy to rapidly changing mechanisms influences what action the people need to take.

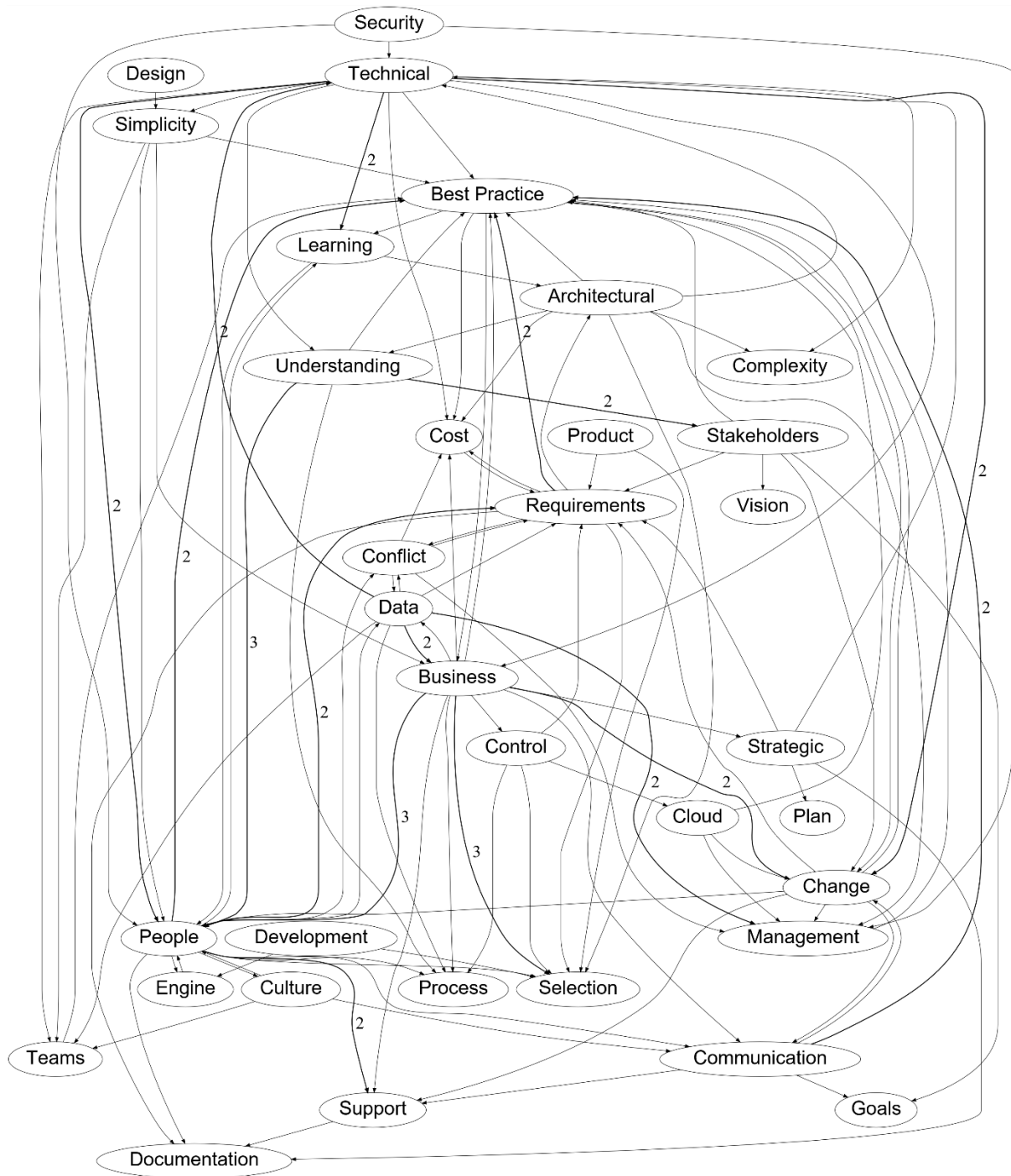
Link 'n' communication ► change A lack of communication of these changes influences how the changes are dealt with or managed.

Strategy may be influenced by the business, and hence a plan is then affected by the strategy. Security influences management and the people involved may find their actions limited. The business influences the people, teams and operational support and rapid changes in technology may require changes of process. Change is influenced by the addition of more technical layers. Change can be major and influence requirements, support and best practice. Data and business may influence process which if not understood can result in silo systems. Also the business may

need to so influence data required. When control is passed to cloud, it is the cloud that influences change, best practice and management. But in turn lack of communication can influence how the business and people deal with change.

### 5.6.10 Global Influence Diagram

ST3: Influence Diagram	The complex interactions shown in the preceding diagrams (Figure 5.21; Figure 5.22; Figure 5.23; Figure 5.24; Figure 5.25; Figure 5.26) when mapped together systematically produce the diagram (Figure 5.27) below, which illustrates the complexity and nature of the interconnections of managing database systems. This is not all the interactions. It is only the ones from the previous figures. All the interactions from the entire data set are illustrated in Chapter 6, Figure 6.10.
------------------------	--



**Figure 5.27 The database system influence diagram consolidated (from: Figure 5.21; Figure 5.22; Figure 5.23; Figure 5.24; Figure 5.25; Figure 5.26)**

The total number of occurrences of influences are summarised from Figure 5.27 in Table 5.8.

**Table 5.8 Total number of occurrences of influences**

Number of Occurrences	From	To
3	Business	Selection
3	Business	People
3	Understanding	People
2	Technical	People
2	Technical	Learning
2	Technical	Change
2	Requirements	Best Practice
2	Data	Technical
2	Data	Business
2	Data	Management
2	People	Best Practice
2	People	Support
2	People	Requirements
2	Communication	Best Practice
2	Understanding	Stakeholders
2	Business	Change

Figure 5.27 as well as demonstrating the nature of the complexity also shown in the list where the influences occur more than once in the six sections. This is not showing every data quote from the entire research but from this state it is possible to visualize which components provide more influences and potentially have a greater impact on the system. Complexity exists throughout all the areas relating to the management of database systems. In this subsystem it is interesting to note that the highest influencers are business and understanding with selection and people

equally being influenced. In this subset of data the technical components do not have as high an influence. The highest number of occurrences are highlighting socio-technical influences.

## 5.7 Summary

The chapter presented the findings of the qualitative data obtained in the focus groups from the participants. The data was analysed using Thematic Analysis embellished with additional tools. Other methods used were spray diagrams, distribution of codes, in vivo coding, code landscaping and the operational model diagram. These methods provided an analytic way of looking at the data. The concluding results presented at the end of the transitional section show the components that have the greatest effect on the system.

The research then looked to synthesis and systems thinking to gain a holistic view of the situation. The tools used were a systems map and influence diagram. These diagrams help with understanding the behaviour of each component and the effect each component has on a whole.

The complex interactions that take place in database systems are illustrated by the influence diagram, Figure 5.27. The adoption of best practices and procedures is affected by the complex interactions and this had been evident from numerous quotations made by the participants which were carefully recorded.

The next chapter will discuss the data findings from both the quantitative and qualitative research and propose the CODEX. The CODEX (Control Of Data EXpediently) is a blueprint to help with the management of database systems.



## Chapter 6: Discussion of Database Management and the Complexity of Delivering a Best Practice Solution

### 6.1 Introduction

This chapter discusses the main findings of the research and is a synthesis of both the quantitative and qualitative analysis. This chapter also incorporates bridging, the consensus between quantitative and qualitative, to help understand the transition between the mixed methods used. Systems thinking has been applied to understand this complexity in order to improve practice. Whereas chapter 5 focused on the analysis method, this chapter focuses on answering the specific research questions.

The research undertaken set out four questions which are discussed in the analysis. The four sections of the chapter each relate to one of the research questions. The first research question is discussed in the section 'best practice usage'. The second research question is discussed in 'Complex interactions in the management of database systems'. The third question is discussed in the section 'Adoption of best practice affected by complex interactions'. The fourth question is discussed in the section 'Improvement and innovation', which also presents the CODEX.

An output from the analysis is an influence diagram showing the complexity of database systems management. Adoption and effects of best practice are discussed in the chapter, leading to a proposed innovation, the CODEX (Control of Data EXpediently). The CODEX is a compact blueprint that will enable improvement in the management of database systems through identifying components that are affected when a certain change occurs. Given that best practices are so ill defined,

differently understood and at different levels, this is not an attempt to propose a new form of best practice. However, the CODEX could be the building block of a multi-layered system, one that enables organizations to reflect upon their practices in database management in the light of their past experiences (and those of others) and improve them. The CODEX will have an effect on the environment and implementations for other teams as this practice is related to the database sphere only.

In order to select an operational practice for a database system, comprehensive understanding of the many choices of software and hardware that could be assembled together is needed if best practice is to be achieved. Best practice must be selected to meet customer requirements concerning outputs and cost, among many others. The knowledge and understanding of up to date technology and software is not easy, due to the increasingly rapid development taking place.

Companies in the same sector of business often follow recommended best practice guidelines, but it has been shown that slight differences in structure could mean that best practice for one context, in another could result in unsatisfactory, inefficient, time consuming and in fact the worst practice to be followed. The complex nature of databases requires experienced staff with expertise and understanding for the best operation and results.

## 6.2 Best Practice Usage

The first research question discussed in this chapter is:

To what extent are best practices and procedures utilised by the database community?

This study of the management of database systems examined how best practices were used and were part of the process. Best practices were connected through every part of the database systems from inception to decommissioning. The concept of best practice underpinned this research, with the premise that it is fundamental to the effective management of database systems. A survey was undertaken as part of the quantitative data collection. The questions posed in the survey were designed to capture the real world situation of the respondents, covering all aspects of the database lifecycle. The questions dealt with the whole system, composed of architecture, development, operational management, security, cloud, cross engines, database management, data management, organizational culture and training for database engineers. This section (6.2) discusses the key findings relating to best practice usage.

### **6.2.1 Best Practice**

It is apparent that 'best practice' had many different meanings for the respondents, and best practices were always changing. There is a diverse set of places to find best practice guidelines, and many respondents' organizations had created their own 'best practices' (Figures 4.7, 4.8 and 4.9). Sanwal (2008) argued "*best practicism*" is the errant belief that there are really certain practices that are best and they will yield better reality or improved leadership. There were a number of issues that could occur in following best practice (Figures 4.10 and 4.11). It was striking that 94% of the survey respondents thought it was important to have best practices, despite their drawbacks. Control of best practices within an organization was shown to vary due to different aspects of the database architecture. As Becker (2004) suggested, some best practices could not be separated from their organization.

The qualitative research provided additional discussions in the focus groups with further opinions relating to the usage of best practice.

Requirements and implementation of best practice can vary between business, industry and companies (Q6 3.2 line 59, Q1 2.2 line 40). Operational best practice *“should not however be overly prescriptive or so long winded that they either stifle innovation or adoption of new technology or are just ignored.”* (Q4 5.3 line 65).

Gonnering (2011, p.98) contended that best practice adoption had a role in improving quality, although it should not just be a method for solving a problem, as learning was important as well.

The control of defined standards, automation and best practice can help improve the quality of work and stop mistakes (Q4 5.3 line 65) (Q8 2.1 line 30). The identification of best practices can take years to find the best working solution (Q4 8.1 line 77) and best practices should be applied throughout the lifecycle (Q4 1.2 line 2). One respondent thought *“it is the data that drives what best practice to set up”* (Q4 5.5 line 63). Change is a continuing factor even in mature databases.

Prahalad (2010) argued that best practice only allowed business to progress to a point and after that point it was next practice that formed the innovation for change.

Agility is required as *“hardware and software continues to change”* (Q4 8.1 line 78).

An aspect of best practice could be to reduce risk, improve consistency and manageability of situations. One respondent said that *“Best practice is never really best practice - it is just a best solution in a particular set of circumstances”* (Q4 1.4 line 4).

Gonnering (2011, p.97) argued that best practices could be used for prescribing a vision for the organisation rather than a blueprint. A lack of understanding of best practice incorporated in the design could contribute to management issues (Q9 5.1 line 91) (Q7 5.1 line 91).

Best practice is only for ordered simple cause and effect scenarios (Kurtz & Snowden 2003; Snowden & Boone 2007). Following best practice for certain areas could have the effect that *“people tend to stop thinking for themselves”* (Q1 7.2 line 114). Wagner *et al.* (2006, p.255) suggest best practice could be used as a political weapon against value judgements and could often constrain choices. However, another respondent thought best practices brought order with no surprises and brought awareness with it. It *“would make things visible at all levels”* (Q4 1.1 line 1). *“It is worth stating that we are all still learning”* and although new recommendations might be found to improve a situation *“the way in which best practices are applied may have their own lifecycle at a site”* (Q4 7.1 line 73).

Thus a comprehensive list of important issues need to be carefully considered when decisions are made concerning best practice. Some of these issues have been discussed in previous literature.

*“The well-travelled existing path is often littered with discarded, useless best practices and the organisations that have fallen victim to them”* (Sanwal 2008).

These results underpinned the research which led to the development of further questions to explore the management of database systems. In sections 6.2.2 – 6.2.7 the topics discussed are the key findings from the quantitative survey. The quantitative survey was designed to obtain a baseline of the practices and procedures in use whilst managing database systems today.

## **6.2.2 Database Management and Data Management**

The management of databases required diverse knowledge and skills which were continually changing, and core technical practices were numerous. Current practices and procedures in data management were important to understand in

relation to database management due to the close connection between data and databases (Figure 4.49). Aiken *et al.* (2011) highlighted trends which linked together components of data management.

A few respondents followed data lifecycle management policies (Figure 4.48), although just under half the respondents' organizations had their own data management practices and procedures. Management of data was regarded as a cost rather than an asset (Aiken *et al.* 2007, p.49)

Little time was spent solely on managing database servers (Figure 4.6), which could indicate that there were many other tasks required, in addition to the management of the database server. The number of practices and procedures created or adopted could be affected by the time spent managing the database servers.

Database management decisions were mostly based on customer requirements (Figure 4.60), and not what the product manufacturer or industry necessarily recommended. This could potentially impact what best practices were adopted.

Leading database administrative challenges reported, by 47% of respondents, for Oracle databases were diagnosing performance problems (Mckendrick 2013).

There was a lack of choice of management processes to match the different sizes of data (Figure 4.53). This could result in the wrong type of solution, the wrong tools being used or inappropriate management.

There are a large number of database software providers available (Maslett 2012) . Software features were fairly important in product choice (Figure 4.44) and this could govern the type of database management available through the tools which come as part of the database products.

Documentation usually contained the architecture design (Figure 4.25), the development design, the configuration and specific practices and procedures. Most respondents considered that documentation was important to improve practices and procedures.

Data governance and master data management were frequently lacking, and data quality procedures were only partially used (Figure 4.47). This is a key area for database management to provide good quality information.

A small group of respondents reported growth in the management of unstructured data over the previous 12 months (Figure 4.55); however the majority of respondents did not currently manage such data. Other research has identified that the volume of unstructured data is growing (Gantz *et al.* 2007; Gantz & Reinsel 2010; Gantz & Reinsel 2013).

Best practices and procedures appeared to be widely adopted for data security, with the exception of procedures to transfer data between servers (Figure 4.50). There had been a number of major security issues where data in transit between different geographic locations and within offices had been lost (BBC News 2007).

### **6.2.3 Operational Management**

The operational state of databases was probably the most well managed area of the database system (Figure 4.33), due to the potential impact to the business if they became unavailable. There is a risk to business if databases become unavailable. This could affect internal employees' ability to perform their role, external publicity or could have significant financial impact due to loss of revenue. Challenges reported in an Independent Oracle Users Group survey (IOUG) (Mckendrick 2011a) for operations included: an increase in the number of databases, databases of larger size, a reduction in older systems being retired, as well as more features and

functionality being included in newer systems. Stonebraker *et al.* (2013) also discussed the operational challenges of new features which add to the complexity.

The availability of data and database servers was a very important factor (Figure 4.18 and 4.36). However, recovery time objective (RTO) or recovery point objective (RPO) were not defined in many cases (Figure 4.33), so when the database became corrupt or unavailable, recovery defaulted to an individual's best endeavours. This could result in lengthy outages of availability.

About a third of respondents did not use an IT Service Management framework (Figure 4.38); such frameworks can help ensure best practices are followed.

Problem management methods were often not used (Figure 4.39). Frequent malfunctions were often dealt with in a reactive way (Figure 4.40). Dealing with malfunctions in a reactive way could be an indication that improvement within the system might not be taking place. Senge (1990) argued that a way of dealing with problems was to shift the burden when the underlying problem was difficult to fix.

An emerging feature of databases was that the state continually changes as changes in the real world are reflected within it, through database structure, data, design or architecture. It would seem important to understand the rate of change carried out on databases, which may require greater need for processes. Although, practices and procedures were set out for changes and were regularly adopted (Figure 4.56), respondents did not always follow such practices and procedures (Figure 4.58). This could result in future incidents or best practice not being applied. Roche (2013) maintains there was a 'focus on engineering and quantified metrics' in a DevOps role that empowered management and architects to help improve the data quality in the release and development stage. This, Roche argued, had changed the culture of the operations engineer, with shared responsibility between



development, quality assurance and operations, and with skills being broader, spanning multiple disciplines.

A third of respondents had no patching policy even though patching takes place (Figure 4.46). This could affect the security of the data stored. Security policies were in place for the majority of respondents, from physical datacentre access through to data access (Figure 4.32).

Installation and configuration of servers was carried out by using the “manual installation wizard” for 63% of respondents (Figure 4.31). This lack of automation could affect the quality and speed of delivery (IT Revolution 2015). The majority of database servers were managed individually (Figure 4.30).

#### **6.2.4 Architecture, Design and Development**

Architecture, design and development are major elements that are a part of the management of database systems. In the survey on database lifecycle management, Mckendrick (2016) states that a pressing challenge for 39% of respondents for managing database environments was “testing new technologies and infrastructure solutions for databases” and “keeping databases at current update or patch levels”.

Customer requirements and the design architecture were key areas in which changes caused service downtime, once systems had moved into a production environment. Data structure within the databases, if not optimal, could affect the quality of the data. Agrawal *et al.* (2009) state the importance of this process, referring to “*architectural shifts in computing*” – in their view, fundamental software changes were being triggered by advances in hardware, data management and cloud computing. There were several findings from the survey, presented in Chapter 4, which are discussed below.

The use of mandated processes for requirements gathering, at the architectural stage, was low (Figure 4.25). Customer requirements were rarely defined at the outset and customer requirements could be changed in the midst of projects (Figure 4.60). Decisions about database management were not always based on customer requirements. Requirements gathering was an important stage in the database lifecycle.

There was a lack of use of core architecture frameworks for database design, although documented design patterns were used (Figure 4.24). Some of this lack of use could be explained if the usage was for an application from a software vendor, where the database design came complete or partially complete and only on-going support was required.

Processes at the design stage were rarely used for NoSQL, NewSQL, Cloud, and In Memory databases and for database sharding (horizontally scaling of a database). About half of the respondents did not follow data structure and hardware manageability processes at the design stage (Figure 4.26).

Although database scalability was a requirement for organizations there were few procedures to manage this (Figure 4.53). Elastically scalable database systems have increased due to global business (Abadi 2012) so having procedures to manage scalability is therefore important.

Agile database management techniques could be used to help improve the effectiveness of database management. Instead of taking a longer term view of the database platforms, shorter sprints with a set time limit for repeatable work patterns, could be used. Agile could be used for database development - VersionOne (2013) found that 37% of respondents were using Agile for “76-100% of projects”. However, the VersionOne survey also showed that Agile was not without its own

problems such as “*lack of upfront planning*” (34% of respondents) and “*a loss of management control*” (31% of respondents).

This survey found that the usage of agile development methodologies was high; indicating the belief that an interactive and adaptive approach was effective (Figure 4.27). Armour (2015) discussed the benefit Agile has had in helping to manage complex unpredictable situations.

Half of the respondents had no standard testing processes (Figure 4.28); this could result in applications being deployed which could cause the databases to perform slowly, be insecure, or grow uncontrollably. About half of the respondents’ organizations had no defined database development lifecycle (Figure 4.28) thus it might be difficult to engage at the appropriate time to ensure an effective secure design that is maintainable.

### **6.2.5 Cloud**

Cloud database services are becoming a popular choice as they offer a cheaper service, have automated high availability and are an easier and quicker route to market. They bring many advantages and disadvantages. Hashem *et al.* (2015) discussed cloud as a significant shift in tackling complex computing, with many advantages. When deploying databases in the cloud, computation, storage and availability are no longer the concern of the business and in these technical aspects skills are not required. Issues associated with cloud are currently: the potential trust and security of data, loss of critical data, noisy neighbours (other applications from other tenants in multi-tenant environments that consume a majority of resources) and longevity of the companies supplying these services. Cloud database services were used for a variety of environments and in many forms (Figure 4.45 and 4.43).

Cloud databases have only become available for usage in recent years, and adoption was currently fairly low (Figure 4.17). Armbrust (2009) discussed the opportunities and obstacles of cloud database adoption. Practices and procedures to manage these cloud databases were currently in place for only a small percentage of respondents (Figure 4.45).

In-house database management skills are not usually required by users of cloud databases. The introduction of cloud databases shifts part of the database administration practices and procedures from the organization to external suppliers (Figure 4.45). Thus it is possible that a different set of best practices is required to manage cloud databases.

## **6.2.6 Database Engines**

Many database engines exist (Maslett 2012) which offer different features. In many of the respondents' organizations multiple database engines were used (Figure 4.15). The newer NoSQL engines were used, in addition to traditional engines. There were various architecture models for transactional, analytical or scale-out architectures. To achieve service availability, it was important to select the correct system for the organizational requirements; and engines from different suppliers needed to interact with each other.

In most organizations, there were no procedures to select different database engines (Figure 4.62). Without a method for choosing database engines or database as a service, the wrong type of system could be selected, causing problems with database management.

There were different management practices for different database products (Figure 4.54) which could mean added complexity when managing multiple systems. An example of this could be the use of different management tools. In an IOUG survey

Mckendrick (2011a) reported that 77% of respondents use different tools for each DBMS platform.

### **6.2.7 Management Approaches**

The culture in an organization affects how certain tasks are carried out (Handy 1985). The management of the organization dictate resourcing levels, what tools and internal systems are available for use and whether time is available for proactive work. With the evolution of database technology and the surrounding hardware technologies more teams of people are involved to ensure the successful operation of the database platform. There are many interconnected technologies which increase the knowledge required to deliver database platforms.

The survey finding demonstrated that improved communication at all levels was required (Figure 4.59). Many respondents stated that communication deteriorated when communicating across multiple teams and that the worst communication was cross boundary communication with the stakeholders. Child (1983, p.113) argued that communication can decrease if conflict between teams occurs due to different goals and criteria. Organization structure could also affect the patterns of behaviour (Child 1983, p.112). Poor communication could lead to problems occurring.

The in-house skill set of people in an organization was not always considered vital in respect of keeping their training up to date. Certification (Figure 4.20) was rarely encouraged by respondents' organizations. Formal training (Figure 4.22) and the opportunity to attend database conferences, workshops or seminars (Figure 4.23) was not given for a quarter of respondents' organizations. Attendance at community events provided free training and an environment where problems could be discussed.

Problems occurring in database operation could be due to a poor in-house skillset, which may influence database product selection (Figure 4.62) so that the best choices to handle the requirements might not be selected.

Control of database choices were reported by the respondents (Figure 4.12) to be in the majority, by Database Administrators and Database Managers. Cloud database software used by the organization was controlled by Head of IT Operations. This control was reported to be with the Head of IT Operations, possibly due to IT budgetary ownership.

Database Management tasks were visible within the database team but further up the management hierarchy visibility reduced. In addition other teams such as development and operations had less visibility (Figure 4.61). Lack of cross team visibility may cause management issues.

### **6.2.8 Summary**

The analysis of the findings from the quantitative survey suggested that there were many and varied practices and procedures throughout the database lifecycle. The complexities discussed in the Claremont and Beckman Reports (Agrawal *et al.* 2009; Abadi *et al.* 2016) have penetrated the everyday workplace for the management of database systems. There were many components involved in managing databases and many stakeholders and best practices and procedures could be affected by these. Best practices were continually changing and many organizations had their own custom best practices.

There were a variety of adoption levels for best practices and procedures. The management of the servers was only a part of managing database systems. The data required management by the respondents. Haas (2015) suggested that the integration of data variety should be context aware, allowing for different usage

patterns of data sets to be combined differently, as challenges arose from data driven complexity. "Documented design patterns" (Figure 4.24) were used, rather than established frameworks. Documented design patterns may be vendor specific or business specific or may be based on practical knowledge and can provide repeatable architectural designs. Operational work was often reactive but the only framework significantly adopted by the respondents was ITIL for service management. Documentation was important to the respondents. Cloud practices and procedures were not well established and it was unclear what skills will be required in future. The database engine selection method was unclear, however software features were an important factor in the choice. Financial budgets had some effect on the version of database platforms selected.

A lack of control (enforcement) of best practice could be a contributing factor as to why best practices were not always followed. Sometimes enforcement was required to ensure conformity. Cross boundary communications with stakeholders required improvement. Formal training on existing systems and keeping up to date on new systems was an important factor in training but was sometimes lacking.

The results of the survey highlighted that there were a vast array of technical areas and technical knowledge required for the management of the database system. Socio-technical issues such as control were also highlighted. There were various areas within the database lifecycle where the survey suggested that there were current gaps in practices and procedures for database management.

## 6.3 Complex Interactions in the Management of Database Systems

After completion of the quantitative survey data collection and analysis, a qualitative phase of research began, in order to address the subsequent research questions.

This section discusses the second research question:

What are the complex interactions that are an integral part of the management of database systems?

The focus group data provided invaluable insight into the research. There was a variety of database roles undertaken by the participants, most of whom had over 10 years' experience. The 10 questions spanned the entire database lifecycle and focused on obtaining an in-depth understanding of best practice, complexity, management and people, including control and communication. The participants' views of database management and best practices were shared. This resulted in considerable variety and depth of the issues raised. The key insights gained from the participants, who shared their experiences of being engaged in the operation of databases, are discussed in this section.

The focus group discussions examined the integral complex interactions within the management of database systems. The findings are discussed in this section, in relation to the second research question (above).

Sections 6.3.1 – 6.3.8 are based on the systems and subsystems groups of codes/components presented in the systems map in Figure 5.13. The systems map consists of four systems: technical; people; business and management. The technical system has three sub-systems: architectural, app dev and operational. The people system has a single knowledge subsystem.



### 6.3.1 Technical System

The technical system contains the “technical”, “platform”, “engine”, “security”, “cloud”, “data” and “engine” components. Security of data, compliance with legislation and establishing procedures could be onerous (Q8 3.2 line 60) yet these are required to protect the database system from attack or data loss.

The many technical layers could cause issues: *“simple fact you have layers - different teams - differing best practice - can end up with teams pulling rather than working together”* (Q5 4.1 line 70). Sanwal (2008, p.52) agrees there are many best practices often with people contradicting each other on what is best practice.

These interactions can impact the support functions and diagnosis of faults (Q5 5.3 line 82). A good database system has *“hardware that is appropriate for the job. The end goal or vision needs to be well articulated, understood and digested.”* (Q3 5.3 line 74). Communication early on helped ensure key functionality was not missing and DBAs were kept informed. *“There is always a complexity between databases and storage. [...] Also there seems to be a massive gap between people who intricately know databases and those who know the underlying storage.”* (Q 5 1.1 line 1). This lack of understanding could cause performance problems as well as the need to understand the data query patterns (Q2 8.1 line 87). The evolution of cloud technology and speed of changing systems and rules made *“it hard to build best practice”* (Q6 8.1 line 100). Gonnering (2011, p.100) argued that best practices are the beginning but adaptation is likely for complex areas.

*“Cloud introduces a variety of complexities and moving parts to new IT projects”* (Q6 7.4 line 98). It introduces uncertainty with often no guarantee as to whether and when service will be interrupted. Agrawal *et al.* (2009, p.63) argued that cloud manageability was complicated by three factors: limited human intervention, high-variance workloads, and shared infrastructures.

The incumbent platform could change when there was dissatisfaction amongst administrators with the incumbent platform provider. However, a participant argued that *“most customers will adopt a preferred platform based on experience of the team, cost and other non-technical factors.”* (Q2 8.2 line 92). It might be that some engine choices were tied to vendors’ requirements and distorted irrelevant requirements were created to match that vendor (Q2 5.1 line 66).

There were challenges working beyond people’s comfort zones (Q5 8.1 line 101) for *“Database engines end up as edge cases for the storage admins, sysadmins, licensing admins, etc. Databases are also growing faster than we’re able to learn how to manage them”* (Q5 8.1 line 100). *“The data tier is often where the complexity lies”* (Q7 5.5 line 93). The number of data curators are increasing in the data driven world; Abadi *et al.* (2014, p.68) argued in the Beckman Report that the challenges lay in building the platforms and with people-centric tasks. In addition Abadi *et al.* (2014, p.64) highlighted the diversity in data management landscape, high volumes, rich data types, shapes and sizes which add challenges.

Data security was hard to maintain and *“rogue operators creating customised interpretations on the data conflict with core data set confidence”* (Q9 5.5 line 63).

The technical system codes and themes are in Figure 6.1.

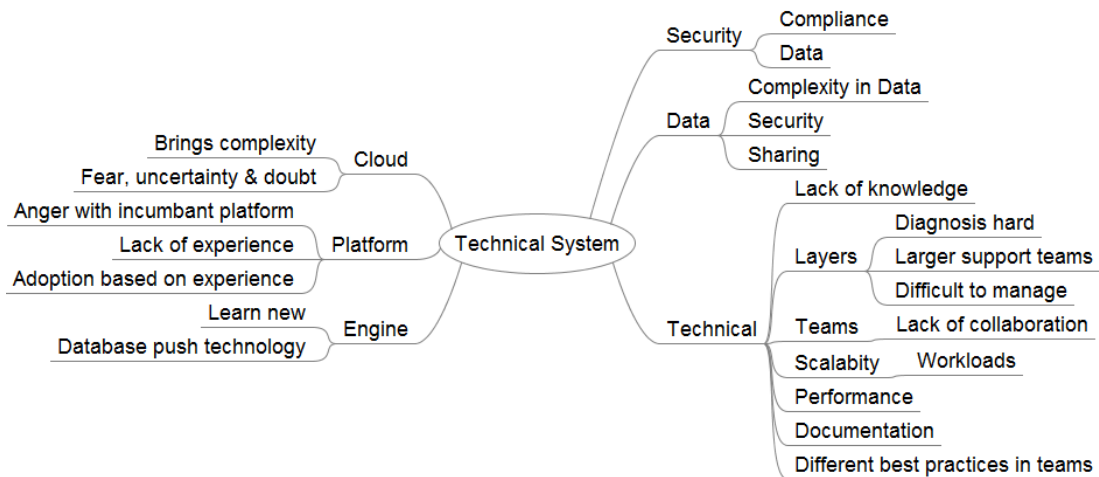


Figure 6.1 The technical system

### 6.3.2 Architectural Subsystem

The architectural subsystem contains “requirements”, “architectural”, “product”, “tools” and “selection” components.

A respondent stated that we don’t architect for modern application patterns:

*“solutions being built today ignore many of the modern advances in hardware and software – people stick with what they are comfortable with”* (Q3 5.5 line 78).

Architectural design and best practise were described by a respondent using the analogy of string theory where stability lies on the surface and chaos underneath. (Q3 2.2 line 44). The current frameworks were sometimes not known, too complex or time consuming to implement. Complexity increased when layers were added (Q5 1.5 line 7). Some people just guessed, using instinct as a guide for hardware design (Q3 6.1 line 81). Kralj (2008, p.17) also argued that the common dilemmas of high complexity, high interdependency, and low transparency projects were overwhelming and that analysis, decomposition and abstraction was important for IT architecture design. Agrawal *et al.* (2009) discussed the changes involved with having many layers.

Requirements do change and needed to be realistic, with a trade-off of cost versus complexity (Q3 8.2 line 90). Sometimes cost and time conflicted (Q7 8.2 line 111). The changing requirements needed to keep up with the changes in the business (Q10 3.2 line 48). Agrawal *et al.* (2009) discussed the architectural shifts in computing which raised the price and performance metric for large systems which expanded beyond the typical DBMS.

Selection criteria for engines could be affected by available skills. An *“important factor will always be whether skills to support a particular engine already exist on-site”* (Q2 7.1 line 74). This could be related to strategic reasons or the organization could be following their defined enterprise architecture. However *“procurement should be controlled to avoid unnecessary proliferation of disparate technologies”* (Q2 1.2 line 2). The selection should fit with *“the ecosystem of your existing processes and infrastructure”* (Q2 3.4 line 39). Davenport (1997, p.98) argued that stakeholders did not fully participate when information architectures were developed, which inhibited change; and due to a lack of understanding there was a lack of commitment when implemented. The selection process could be distorted to satisfy one vendor (Q2 5.1 line 6).

The product, once deployed, could continue until the end of its lifecycle and the *“shininess of the version not necessarily the best thing to do”* (Q10 6.1 line 94). On the contrary, new database technology deployments should be in the pipeline as they could have some advantages to improve availability (Q10 2.2 line 37). The introduction of new technologies require DBAs to learn about the technology. The survey undertaken by King (2015) found that learning new technologies was a key challenge for DBAs.

There were many tools to help manage databases and it was best to use the right tool for the right job (Q9 7.1 line 77), even though it could mean the business was locked into one vendor, which could be uncomfortable.

The architectural system codes and themes are in Figure 6.2.

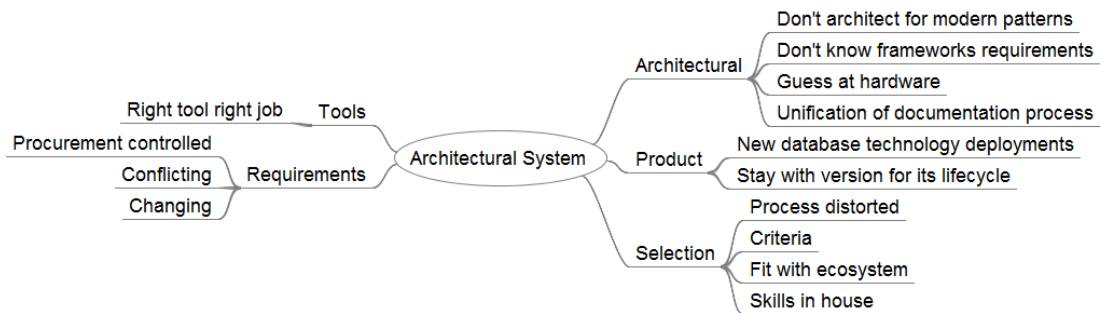


Figure 6.2 The architectural subsystem

### 6.3.3 App Dev Subsystem

The App Dev subsystem contains the “design”, “development” and “application” components.

Different sizes of businesses had different design processes; with some larger organizations designing their requirements at the start of a project, whereas smaller organizations had smaller budgets which prevented that. “A hybrid of layers using differing application architectures is most difficult to deal with” (Q5 5.1 line 84).

Design obscurity caused issues such as slowing analysis down and some aspects (such as naming of objects) were hard to change once the design was in place (Q7 7.2 Line 104). An approach to increasing functionality and performance of the database design and its schema could be obtained using tools (Ioannidis *et al.* 1992).

Lack of cross-team communication in development, and complex interactions occurring with different teams’ communication methods, could cause problems with

producing a supportable database system. *“It’s good enough’ is subjective and does not fit with the Operations side. Sometimes this just shows a lack of experience, understanding, or willingness to talk to the Production support teams.”* (Q3 5.3 Line 75). Wettinger *et al.* (2015) argued that fast and frequent releases of software require changes to be pushed quickly onto production environments, and that this can be blocked by the goals of the operation teams, who are trying to keep the platform stable. Communication was key between developers and DBAs, and there was a need for someone who understood both sides (Q8 3.1 line 61).

Abstracting in development could give productivity gain but cause worse performance *“there are hundreds of “best practice” in development platforms and techniques (Entity framework etc.) which can bring productivity gains at the expense of clarity and performance.”* (Q5 7.1 Line 93). Some products compromised the ability to follow best practice (Q7 2.2 line 17) with bad practice forced by the application.

The App Dev System codes and themes are in Figure 6.3.

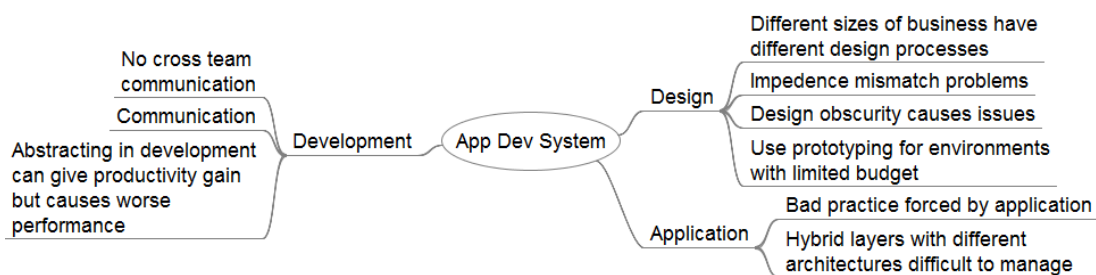


Figure 6.3 The app dev subsystem

### 6.3.4 Operational Subsystem

The operational subsystem contains the “process”, “support”, “implementation”, “change” and “documentation” components.

Troubleshooting issues could be hard with more layers. *“The more layers there are to manage the larger a support team is usually required and the harder it is to quickly diagnose issues and resolve them.”* (Q5 5.3 line 82). False alerts were distracting and could be caused by people leaving the servers cluttered with old files and folders. *“You need that kind of rigor and you need order from chaos. Not chaos from chaos which is what you get if you don’t have your quality gates as you go through”* (Q4 2.2 line 16). Mckendrick (2015) found that complex issues were being engaged with when management and monitoring tools were implemented. New databases that were added to a database system by the operations team may not meet the defined standards to ensure backups, security and performance were unaffected and the production DBAs were rarely happy with that (Q3 5.3 line 74). For some issues it was necessary to have full access to all the layers and collaboration between people was required (Q5 2.1 line 16).

Documentation usage could be low (Q1 3.2 line 75) and sometimes there was a lack of documentation. *“I hope that best practices will evolve faster in the same way that Wikipedia rapidly iterates over human knowledge documentation”* (Q4 8.1 line 79). Also some people who wrote documentation lacked skills to understand exactly what to write (Q8 8.1 line 102). Control of documentation could prove hard in many cases, some being written by outsourced companies or different teams’ naming conventions having a mismatch (Q9 5.5 line 63).

The implementation of business guidelines could be by DBA teams but sometimes this was out of the control of the DBA when purchased products had poor implementation of database configuration. There might be no processes or *“the people do not know the process”* (Q5 8.2 line 106). The change of one component in the business process could affect another due to technological connections. Following processes could help reduce personal bias (Q3 8.2 line 107). There could

be no process visibility. *“Too often silo systems are built because an organization had no visibility where the data’s actually used”* Q3 5.1 line 77. Silo systems are where an insular mind set occurs and people do not work or share resources with other people or teams. Silos are organizational phenomena that can occur through a lack of communication, structural impediments or poor integration between teams.

Many changes carry risks and best practice should not be implemented blindly (Q7 8.1 line 107). The rate of change was very quick and could be affected by external factors. *“I think the best practices are going to guide everything that happens within it (the organization) and all you are going to do is design what your best practices are as part of it”* (Q10 2.1 line 34). Changes to layers of the database system could have an effect. *“As more layers get added (DBA, Network, SAN, Virtualisation, Cloud, and BI) to database projects, the more teams or individuals need to be consulted for any proposed change.”* (Q5 5.4 line 83).

The Operational System codes and themes are in Figure 6.4.

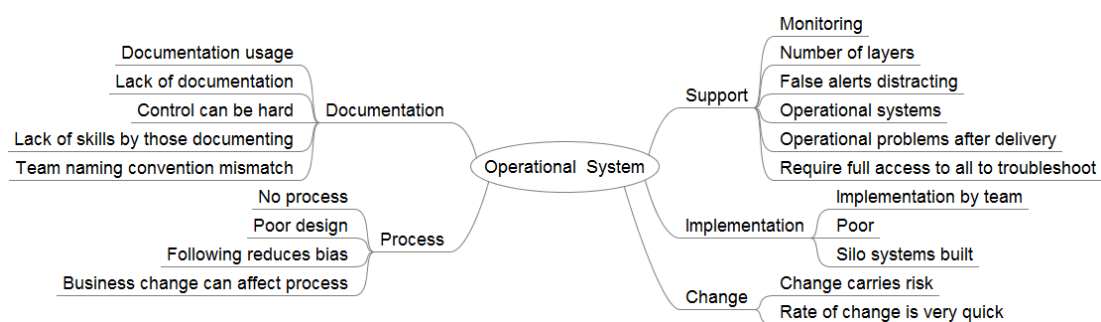


Figure 6.4 The operational subsystem

### 6.3.5 People System

The people system contains the “people”, “stakeholders”, “vendors”, “teams”, “group dynamic”, “control”, “culture”, “conflict” and “communication” components.



Office politics could be a big problem: *“the biggest thing that comes to mind it's not a technology problem but a people problem. People's egos and office politics are usually the biggest problem that have to be dealt with”* (Q5 8.3 line 109). Even today the ‘class struggle’ has not been entirely eliminated and could be in evidence between teams and stakeholders with different viewpoints. Drucker (2007, p.52) argued that a challenge for management was that the social universe changed continually, and assumptions could quickly become misleading.

There was a lack of knowledge and the complexity of database systems compromised ability to implement best practices and procedures. *“Lack of documentation and don't know what some databases are for”* (Q7 5 line 85). There might be a mismatch of goals, which required understanding and communication. *“Stakeholders have their own agenda, have their own requirements, their own job description to achieve”* (Q9 3.1 line 43). Child (1983) stated that the implications of change needed to be examined by the managers. People could be stuck with one way of thinking (Q10 4.1 line 67).

The answer to complexity was to implement better procedures and practices (Q7 7.4 line 105) *“impediments are almost always human beings and their political interests in a major IT budgetary decision”* (Q7 7.4 line 105). Szulanski (1996, p.27) argued that knowledge transfer was caused by factors such as: lack of absorptive capacity, causal ambiguity and an arduous relationship between the source and the recipient.

There were challenges which were different for different use cases. *“Always look through a critical lens at ‘Best practices’, some may not be very applicable to your particular use case whilst others may be very relevant”* (Q1 1.5 line 5). Best practices could cause conflict between roles. *“There may also be a conflict between*

*'best practice' for database specialists and those for Admins: Little point in having a system that is perfect from a database perspective but cannot be maintained.*" (Q1 1.5 line 5). When systems were inherited it was sometimes difficult to get them to link together (Q5 3.1 line 55). Certain things could mean different things to different people, for example developers often had different viewpoints to DBAs.

Vendor selection could be a concern for a business wanting to avoid lock in and the politics of vendor selection (Q3 7.2 line 83). Also vendor recommendation could introduce a whole new set of complexity for the business.

Stakeholders' own agendas and priorities brought challenges *"the challenge here is that one stakeholder may have a best practice he wants to adhere to, but another stakeholder (or even worse non stakeholder) is responsible for the resources that are required."* (Q9 8.2 line 86). Stakeholders could think the DBA is in total control; however they did not really understand why certain best practice was followed.

*"If some stakeholders aren't aware of why certain best practices are being enforced then it can often be misinterpreted as 'DBA never lets us do anything', or seen as foot dragging from IT whenever a business change is required."* (Q9 5.4 line 61)

Teams might not work together well, resulting in conflicting best practices (Q5 4.1 line 70). Dani *et al.* (2006, p.1725) stated that best practices could be related to context. Cross team requirements might be obstructive and stop other teams completing their work (Q9 8.2 line 89). Schein (1980, pp.142–145) argued some managers believe in team work and others in individual work. Schein believed that groups affect both the organization and other groups. Schein argued that not all the people in departments or organizations interact and work together, so they may think they are in a group but they are not.

Group buy-in and group lack of business knowledge could prevent best practice adoption. (Q8 3.1 line 57). Team conflict regarding what was the best technical decision could result in applications not performing well (9 8.2 line 89). Best practice and process could be affected: *“best practice for SAN or network team may be in conflict with database.”* (Q9 4.1 line 50).

Some items were controlled and others not: *“procurement should be controlled to avoid unnecessary proliferation of disparate technologies”* (Q2 1.2 line 2). Lack of control compromised ability to implement best practice and could lead to duplication of work. *“Some users have more access than they should have, so can result in more than one person working on a problem”* (Q7 4.1 line 71). The people in control are not always IT: *“IT lost out on control to the business.”* (Q2 5.4 line 58). Mistakes can occur without control: *“without a level of control people make mistakes and things can get missed”* (Q8 2.1 line 30). There was a loss of some control when using the cloud (Q6 7.4 line 98).

The implementation of best practice could be affected by cross boundary communication and the company deciding to implement a different technology stack where staff have no experience (Q9 2.2 line 16). There might be a communication mismatch between teams *“easily distorts any well intended efforts to apply any sort of practice – best or otherwise. This is often magnified by the different perspectives and (spoken and computer) language used by different individuals and organisations”* (Q9 5.1 line 62). Lack of communication could sometimes mean *“Supportability operations are the last people to know [...] it is hard for the IT department and reaches them too late”* (Q2 5.4 line 58). Child (1983, p.114) argued that there was an

“inadequate coordination between the different departments [...] it is exacerbated by the variation in outlook between people trained in different functional disciplines and by the conflict between specific criteria of performance which are attached to separate departments”

which often manifests in poor coordination between sales and production.

Sometimes the limitations of technologies were not shared (Q9 8.1 line 84). Dani *et al.* (2006, p.1725) concluded that lack of communication and lack of understanding prevented common sense usability.

The culture in some organizations was disconnected from practical issues, and current frameworks were too complex to use. *“It is ivory towers, you can’t ignore the real world and all the components and people. Why is there no time to implement best practices, don’t know current frameworks, current frameworks are too complex to use or too time consuming to implement”* (Q7 1.5 line 7).

Insulated management systems and a lack of appreciation of other teams’ caused problems. *“When departments are siloed - implementing sound architecture across the database and database access code is nearly impossible. The result is often organisational fights that add no business value”* (Q9 7.2 line 79). Cultural values of the teams affected business decisions for best practices and procedures (Q9 7.4 line 80). When new technology or features were added to a system, support was considered afterwards (Q4 5 line 56) and this could result in retrofitting being required.

The People System codes and themes are in Figure 6.5.

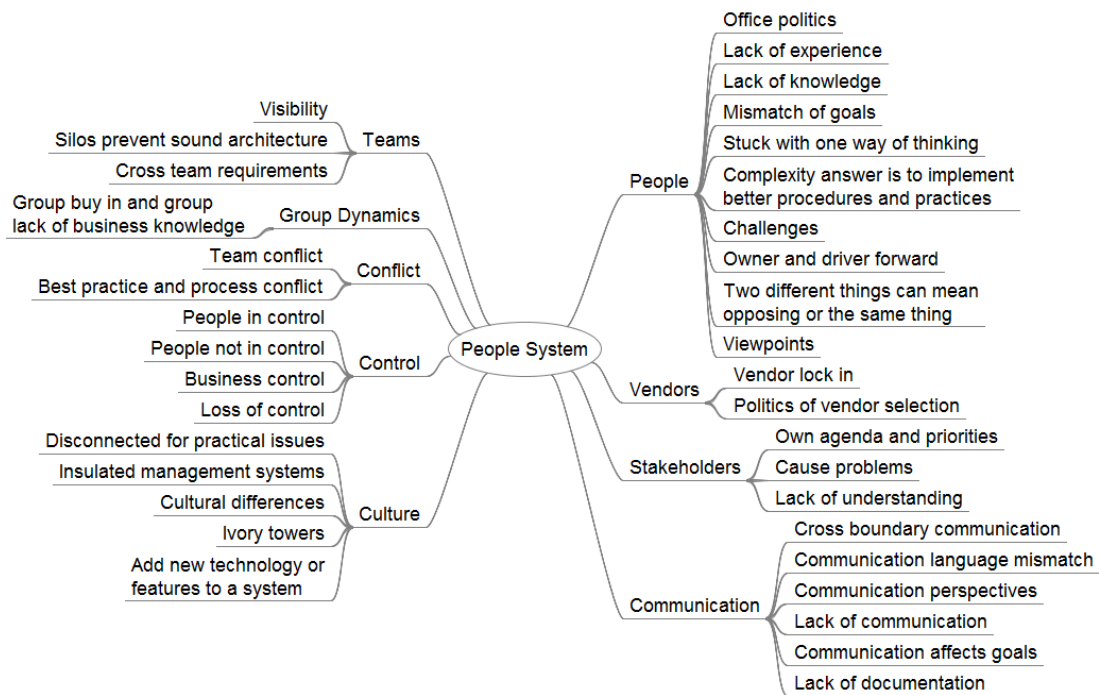


Figure 6.5 The people system

### 6.3.6 Knowledge Subsystem

The knowledge subsystem consisted of “training”, “understanding” and “learning” components.

Participants shared views about the lack of understanding. “*The lack of access or knowledge about the infrastructure under the database*” (Q6 8.3 line 104). Technical understanding was required for all components in the DBMS. “*I think understanding your environment will dictate what best practices are*” (Q1 3.1 line 104). Holze and Ritter (2011) noted the lack of knowledge when trying to automate management of technical components, which was due to the lack of knowledge of the system-wide affects and relationships between components. There are differing levels of understanding between people. “*Different people want different things and those who don't fully understand how things work don't understand why things need to be done in a specific way*” (Q9 7.3 line 78). The understanding of other roles in other teams was important: “*understand the current staff's strengths and weaknesses - for*

example, their comfort level with existing database engines, and their ability to learn new ones” (Q2 8.1 line 89). Denning (2014) argued that knowledge, learning and practices are the best defence for a rapidly changing world.

Learning was too slow, “databases are also growing faster than we're able to learn how to manage them” (Q5 8.1 Line 100). Learning could be a big problem: “the staffing challenges and learning difficulties are the biggest problem facing databases” (Q5 8.1 line 101). When contractors were employed there was sometimes no knowledge transfer (Q4 4.1 line 44). Tucker *et al.* (2007) argued that knowledge transfer of practices was easy when it could be combined as written communication (“Know What”), but if it included context dependent understanding and had tacit knowledge (“Know How”) the transfer was problematic, and could result in poor practice. Staff might not be adequately trained (Q5 8.2 line 105) and cross team training was crucial (Q9 7.2 line 79).

The Knowledge System codes and themes are in Figure 6.6.

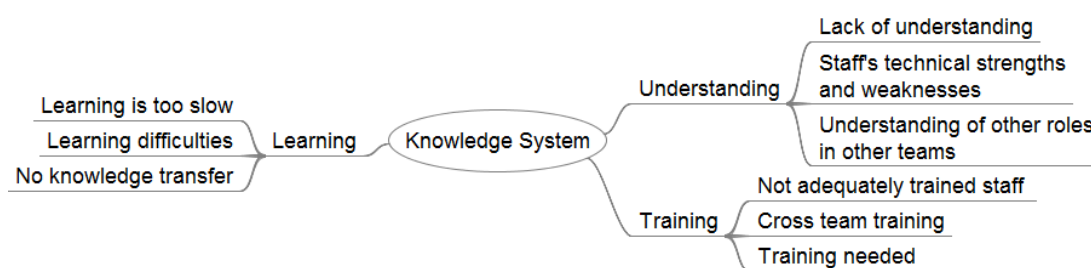


Figure 6.6 The knowledge subsystem

### 6.3.7 Business System

The business system consists of “vision”, “goals”, “cost”, “business”, “strategic” and “plan” components.

Processes could be affected by a rapidly changing business model “The business model may be changing too rapidly for the processes and people that are in place to

support them.” (Q5 8.2 line 108). Kurtz & Snowden (2003, p.475) argued that the Cynefin model enabled more sophisticated decision-making when framework changes were understood during periods of rapid change. The business drove what was selected (Q2 5.4 line 58). Customers might have conflicts between cost and time (Q6 8.2 line 103) and budgeting, and projects could be easily over spent. (Q6 5 line 83).

The business set the vision and management followed it but “there has to be a consistency and clarity to the vision that enables others to follow it” (Q10 5.5 Line 86). The strategic plans could be followed by technical staff if they had clear guidance and budget (Q10 5.5 line 86).

Flexible plans were useful as they might need to change (Q10 6.1 line 94). Some Roadmaps just did not work as everything was moving too fast. Disruptive technologies affected planning (Q10 5 Line 75) and Denning (2014) argued this could happen at any time. Plans might need to be reworked to make sure the goals met the strategy (Q10 7.4 line 104).

The Business System codes and themes are in Figure 6.7.

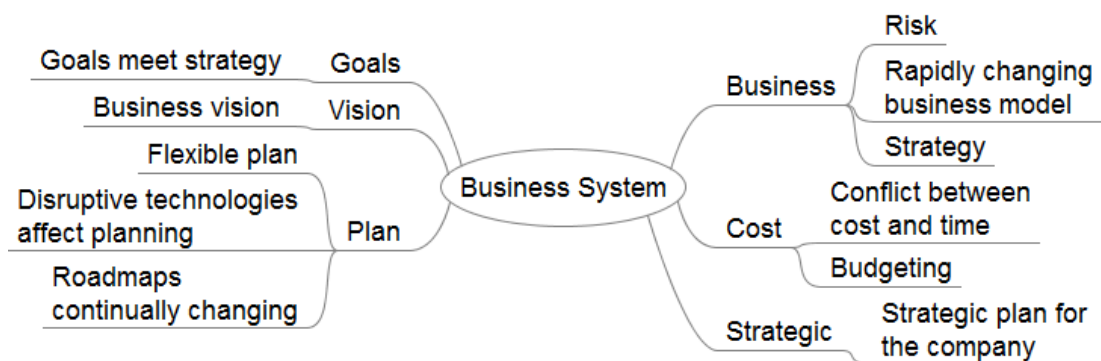


Figure 6.7 The business system

### 6.3.8 Management System

The management system contains the following components: “lifecycle”, “flexible”, “complexity”, “simplicity”, “best practice”, “management” and “standardization”.

Operational and strategic plans varied. One respondent commented “*we have a general strategic IT plan of which database management is a part as it should be. But such plans tend to be fairly static if not abstract; change happens quickly and new requirements constantly arise, so operational or tactical factors are of much more concern*” (Q10 7.1 line 101). Best practices needed discussions and agreement. “*If a best practice doesn't fit, then eventually someone will go around it or it'll get overridden by senior management.*” (Q8 5.4 line 85). With differing management techniques in use in different teams, for requirements gathering, hardware and agile development, discrepancies could occur (Q3 2.2 line 42). Realistic expectations and keeping things simple were important (Q8 3.2 line 30). If best practice is too onerous it won't be followed (Q1 3.2 line 72).

There was a lack of open standards (Q5 5.5 line 85) but some standards changed and were forced upon the organizations by the vendors “*closed standards that are pushed by a vendor and then abruptly discontinued which means applications and databases have to go through a lengthy development process in order to be upgraded.*” (Q5 5.3 line 82).

A consideration for any database system was “*the lifecycle of the data, which often long outlasts the lifetime of any one particular database or application.*” (Q4 5.1 Line 67). Silberschatz *et al.* (1991) argued that persistence of long term database maintenance was a key area for database systems.

Flexible management of database systems could be affected by external factors. “*Sometimes the plan may need to be modified due to numerous reasons (such as*



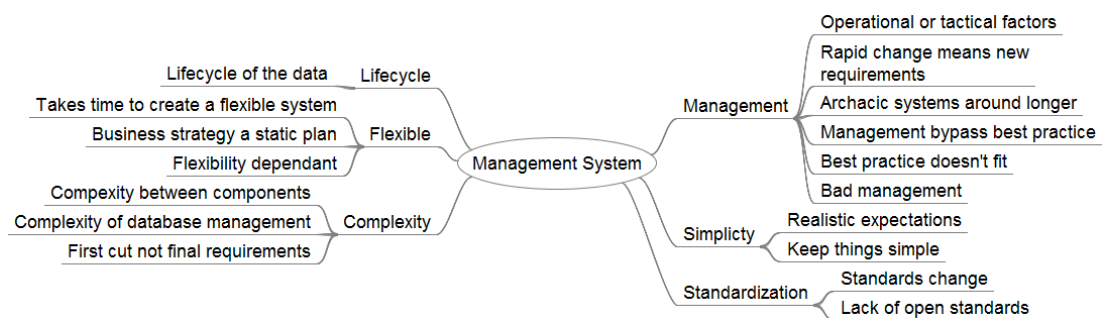
database compatibility, third-party support, etc.) and this may result in archaic systems being around longer than the plan suggests that they should. The plan requires a degree of flexibility in it to be effective.” (Q10 1.1 line 1)

Complexity of external data and source systems could be addressed through certain designs (Q7 3.2 line 63). Mix and match of environments could increase complexity (Q5 3.1 line 55). Best practice was not always appropriate for every scenario and articulating when and why not to use it was key (Q6 8.2 line 101). There could be complexity between layers: “Any boundaries between technology layers (i.e. non-integrated components) introduce complexity in terms of interfaces and data flow.”(Q5 1.2 line 2).

“The complexity of the database system is that of management or the complexity of the system that is implemented; the infrastructure that is there” (Q7 6.1 line 94). The requirements at the beginning often need revision “It’s a rookie mistake to treat this first cut as final requirements the process is one of negotiation and trade off v cost and complexity.”(Q3 8.2 line 90).

Mckendrick (2013) argued that as data became more important in businesses, more segregated environments were introduced and multiplication of data sources increased the complexity of management

The Management System codes and themes are in Figure 6.8.



**Figure 6.8 The management system**

### **6.3.9 Summary of the Complex Interaction Findings**

The complexity of the management of database systems, and the interactions, were exemplified by the participants. The findings highlight that there were differing requirements and deployment models, and that these were affected by many factors.

Changing technology and the increased depth of technical layers add complexity. It was reported that the more layers there are the harder it is to diagnose faults. Rapid business model change and rapid technology change and disruptive technologies affected planning. Disruptive technologies like the cloud added complexity. Changes in any part of the business carried risk for database management.

Architectural design could be using old designs and not modern platform design but sometimes the “shininess” of a new version encouraged its usage before it was in a state ready for operational delivery. Respondents stated there was the need for the right engine for the right job. Frameworks were too complex and data outlasted the databases. Archaic systems could have a longer shelf life than expected. Data security could vary depending on how users use the data, and new requirements were constantly appearing. Documentation was hard to control, although it was important to have, and keep up to date, to ensure standards were maintained for delivery. There was a need for extracting order from chaos, and quality gates to enforce that. It is not advisable to blindly follow best practice.

There was a lack of understanding of the technical set up. Limits in technical experience of current teams could be influential in decisions made. Learning was too slow and the businesses had a lack of experience and skills. There was a lack

of control and consultation with technical teams, who had the database knowledge.

Where there were silos systems there was no understanding of data usage.

Political views, office politics, differing viewpoints or agendas of stakeholders could affect management. Teams had conflicting best practice, which could affect productivity. There was a communication mismatch between individuals and organizations whereby organizations did not provide clear guidance.

Communication between teams was poor. Teams did not work together and communication was not early enough in the process. There could be conflict between cost and time for the business and customers. Procurement often controlled budgets and costs. Also respondents reported conflict between teams, and 'ivory towers' where teams were disconnected from the practicalities of managing database systems. Silos could prevent the implementation of sound architectures; sometimes departments did not wish to share information, resulting in reduced efficiency in the operation of database management. Management were reported as bypassing best practice that did not work, which could be good if the 'best practice' was actually reducing the quality of the management. Communication was an important factor between all stakeholders.

The complexities of current database systems were such that they affected what was regarded as best practice. Indeed it would seem that different individuals in different teams had conflicting views of what was best practice in particular circumstances. The many details of particular conflicts have been revealed.

## **6.4 Adoption of Best Practice Affected by Complex Interactions**

This section discusses the third research question:

Is the adoption of best practices and procedures affected by the complex interactions that are an integral part of the management of database systems?

This question builds on the quantitative survey findings which address Question 1 (Section 6.2) and the qualitative focus group findings which address Question 2 (Section 6.3). The complex interactions shown in the influence diagrams in Figures 5.9 - 5.14, when mapped together, produce Figure 6.9. Figure 6.9 illustrates the complexity of large database systems. McKendrick (2016) showed in his survey that complexity had increased in the last 5 years, and that this was driven by data growth, business growth, more interconnectedness between data environments, increased security requirements, compliance or regulatory requirements, variety of data environments and greater connection to cloud or external environments. McKendrick (2016) asked respondents what steps their teams had taken in the past five years to reduce complexity. The top three answers selected were: migrated databases to virtualized environments; greater automation; and applied management tools and configuration tools.

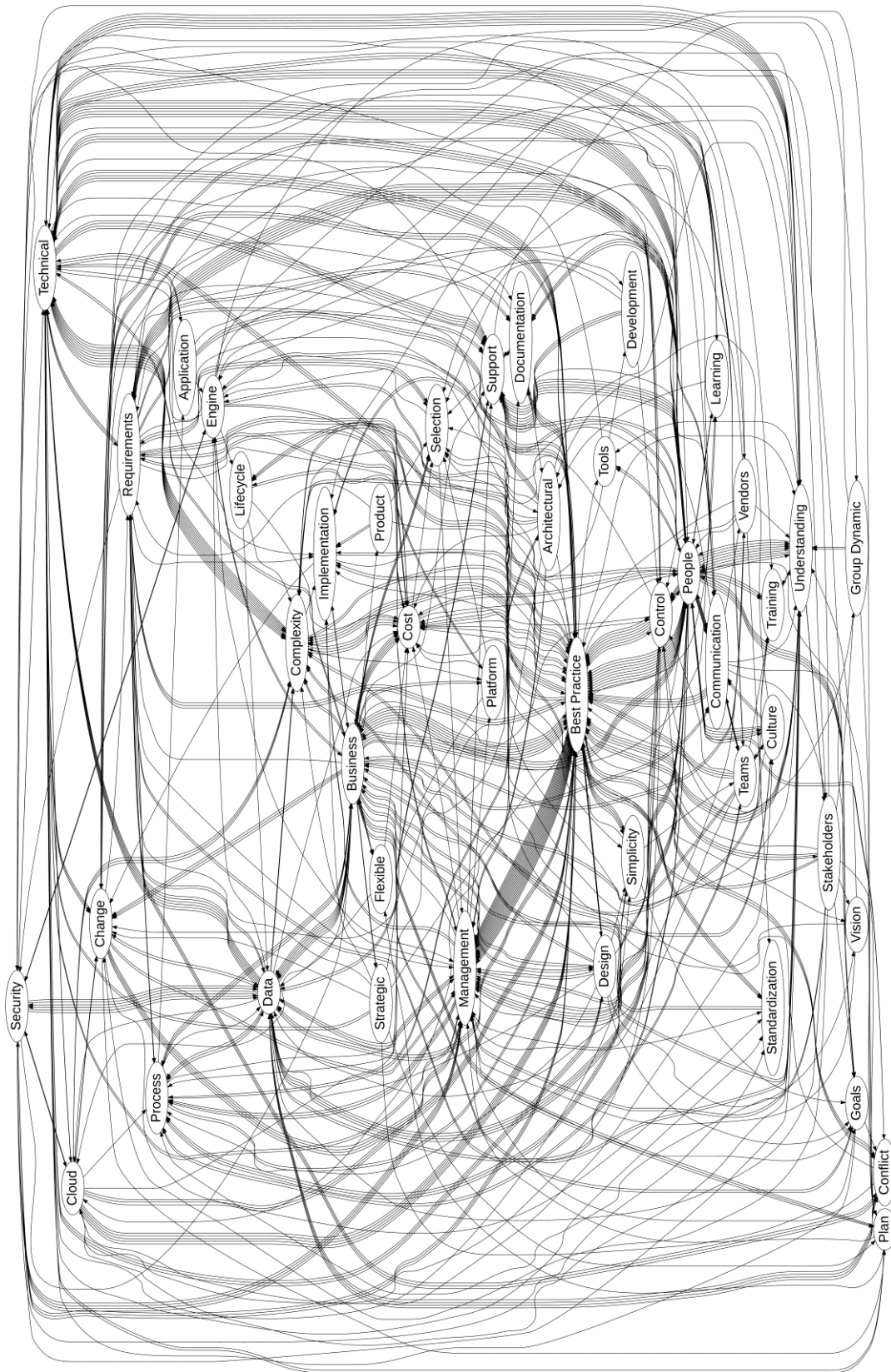


Figure 6.9 The whole database system influence diagram

The influence diagram Figure 6.9 shows the main components of the database system and the relationships between the components. Figure 6.9 shows the whole database system to be highly complex, with many interconnected components. The greater the number of influence lines depicted, the greater the number of interconnections between components that influence each other in the system. The influences between the many components shape the management of database systems. It was not just the technical side of management that affected database systems. Seddon and Caulkin (2007) argued that systems thinking is about interconnectedness, and that leads to looking at the whole set of parts. The influence diagram not only shows how complex the system is for managing databases but it also identifies what the complexities are and which components have the greatest influence or are influenced by the other components the most.

To be able to determine whether the adoption of best practice and procedures was affected by complex interactions it was necessary to identify what the components of the system were. Complex interactions were defined in Chapter 1 Section 2 as interactions involving three or more components. Johnson (2009, pp.13–15) highlighted some key components of a complex system. The complex system could be adaptive, receive feedback, have emergent properties, have order and disorder. Johnson (2009, p.76) argued that feedback presence was central to complex systems and complexity, even if it was not explicit, for example memory of previous events, could bias the discussions as a form of feedback.

The following quote was particularly relevant.

*“All the time we must treat best practices as a general guideline, but the reality is that complex systems (people, process, technology or*

*business) sometimes have needs that hamper conventional best practice.” (Q6 8.2 line 101)*

### 6.4.1 Adoption of Best Practices

The adoption of best practices and procedures was affected by the complex interactions. The focus group discussions highlighted a number of complex interactions which could affect adoption of best practices and procedures. In some cases people within the business had to ensure that the quickest and cheapest options were selected for hardware, software and resources. Resources could be limited so attempting to make best use of the available resources could result in unwanted emergent behaviour. This emergent behaviour could be that best practices and procedures were not followed and compromises were made to facilitate business goals, software selection, price, time and manpower.

*“Most of my work involves finding the fastest, cheapest, easiest compromise to implement in order to accomplish the business goals. The business has to be able to make money and avoid loss - and unfortunately, a lot of best practices and procedures ignore costs. If we all had unlimited time, manpower, and money, we'd all build systems according to best practices, but like Steve Jobs said, artists ship.” (Q7 8.1 line 108)*

On the other hand emergent properties might contribute to survival in the rapidly changing environment. Checkland & Scholes (1999, p.19) argued that using communication and control would enable adaptive behaviour to environmental changes.

There were complex interactions arising through the technical layers. They interact with their surroundings. People could operate in ‘ivory towers’, which results in their

own department's needs only being considered. A lack of time and complex frameworks affected the adoption of best practices. Simplicity of design could help reduce complexity.

*"It all links into the requirements and how they are met. Basic complexity is increased just by adding a layer. [...] adds complexity to implementation. You need requirements analysis for implementations. [...] Easy to add constraints of the other layers at the beginning, harder at the end."* (Q7 1.5 line 7)

The culture between the teams could have a significant impact on the adoption of best practices. Different teams might have best practices which matched in some cases, but in others might be diametrically opposed. There were layers within all the technical and non-technical components and regardless of whether these were simply configured or simple non-technical layers, these layers, and teams and ways of working produced complex interactions (Q5 4.1 line 70).

Another respondent also raised the issue that database specialists' best practices might conflict with admins' best practices. Management of database systems was not solely about providing the best database system in isolation. The system had to interact with the environment around it and adapt to particular use cases which provide the most relevant best practices. The system has to be manageable. Best practices in themselves need to be reviewed and feed back into the process (Q1 1.5 line 5).

The adoption of best practices might be affected by limited resources such as cost and time. The complex interactions of these constraints may affect the quality of database management. The database consultants and managers might need to take on the mantle of assessing the risks and covering their own business for issues



the customer may have later in the project. These complex interactions would be unknown at this time.

*“Often we need to cover ourselves as the customer may have conflicting requirements around cost and time and we need to make sure that they are aware that one constraint may affect other system qualities such as best practices and quality risks.” (Q7 8.2 line 111)*

The adoption of best practice was affected by complex interactions related to support issues. The change of any configuration settings, hardware or software could affect services by either improving them or reducing performance if the wrong decisions were made. Worst still the adoption of best practice might affect the reliability of the system, the speed, or make it harder to manage the system.

*“I’m a consultant who gets called in when the server’s on fire. It’s not reliable enough or fast enough. Because of my job, I can’t blindly implement best practices. Any change inherently carries risk.” (Q7 8.1 line 115)*

Watzlawick *et al.* (1974, p.41) argued that changes could cause complexity as more and more exceptions and inconsistencies were included in the overall premise. A second order change, changes to governing rules or internal order, might be required to simplify the situation.

Understanding of the environment, the data and support requirements could affect how best practice was configured. These complex interactions could affect best practice, dictating the requirements necessary to maintain the type of data to the required standards.

*“I think understanding your environment will dictate what best practices are if your financial and accuracy is important. I hold a lot of data including paediatrics medical records and availability and if that went down for 2 days it is not the end of the world but if we lost it that part of it, it is a massive disaster.” (Q1 3.1 line 86)*

Best practices could be misinterpreted by stakeholders as universal truths which could affect the adoption. This adoption could result in reduced quality if the best practices were not understood. Best practices were often portrayed as the gold standard but in fact they were guidelines to follow, not the only configuration options available. They could be a rule of the business on how they would like certain components within their business conducted.

*“Poorly understood best practices make it all worse. Best practices should be treated with care. They are not universal truths but guidelines (think: maxims)” (Q9 7.5 line 81)*

Gonnering (2011) argued that adoption of best practice might be problematic, with a lack of ability to contextualise the practice.

There were complex interactions between stakeholders and the business. Poor communication could lead to misunderstandings between the stakeholders. This could be misinterpreted as tight control of the system or stalling the coming changes when best practices were enforced (Q9 5.4 line 61).

The adoption of best practices could be connected to people and their communication skills. If communication was not taking place best practices might not even exist.

*“If you don’t have any communication you tend not to have any best practices or procedures” (Q9 3.4 line 40)*

Business risk could be reduced through best practice adoption. An ingrained culture of firefighting, fixing issues as they occur rather than providing a proactive supportable service, did not work well when trying to adopt best practice. Trauth (1989, p.266) argued that traditional firefighting for information management must change to incorporate planning and feedback. Communicating best practice configurations could ensure database systems could be recovered.

*“In general, best practices reduce exposure to risk, minimize firefighting, and ensure strong performance and easier recoverability / business continuity.” (Q4 7.4 line 75)*

Falconer (2010) argued to the contrary that there was often insufficient information to de-risk the use of best practice. There was a lack of situational or contextual information and measure of success recorded after application.

## **6.4.2 Changing Best Practice**

Best practice could be a control mechanism to prevent previous errors from recurring. With the rapid changes in hardware and software, agility in best practice was required to survive in this fast changing environment. Best practice was quite often set and remained the same throughout the lifetime of the product, but in the context of the continual changes and new products. Technical people were having to continually learn and adapt to deal with changes. The understanding of the current set up, of how different sets of new hardware interacted, became clearer over time, and best practices evolved with this newly gained understanding. The feedback from lessons learned in database systems ensured the system was adaptive and evolved over time.

*“In mature databases, best practices and procedures avoid relearning mistakes. Unfortunately, the best practices and procedures have to remain agile because hardware and software continues to change.” (Q4 8.1 line 78)*

Processes could be put in place by the business which adhered to best practice. People added further complex interactions by making mistakes in their work if it was not controlled (Q8 2.1 line 30). A business could control the standards through the establishment of best practices. A business needed to prevent things being missed, which could cause possibly catastrophic results and loss of business.

Adoption of best practices might allow people to focus on the areas in which they work rather than becoming swept away with the rapid change to operational and tactical factors. The support required for database systems during the lifecycle could change rapidly, with the growth of data, new technologies and new environmental changes (such as the cloud model). Rapid change could affect the adoption of best practice as database management could be a part of these plans.

*“Personally speaking, we have a general strategic IT plan of which database management is a part as it should be. But such plans tend to fairly static if not abstract; change happens quickly and new requirements constantly arise, so operational or tactical factors are of much more concern. [...] I find it best to concentrate, place the focus of best practices and procedures on those factors.” (Q10 7.1 line 101)*

Gonnering (2011) argued best practice should be the starting point of a complex adaptive system. The emergent outcomes formed from structure and process and the adoption of best practice depended on the transfer of explicit knowledge.

Falconer (2011, pp.173–174) argued that best practice has led to people taking

refuge in it, and it masking the laziness of people. Falconer also discussed the issue of strategic thinking only happening at the higher levels of organizations and that best practice usage supplanted strategy. Identifying and transferring best practice could be problematic “Management consultants do not possess the alchemy for identifying and transferring best practices” (Wellstein & Kieser 2011).

The cloud offerings could be considered an environmental change. The variety of cloud models added a level of complexity, whilst the technical configuration and components behind the cloud product continuously change. It was difficult to create best practices whilst not having any say in what product changes were available or and control of what database or database servers shared the same physical environment. The changes to the cloud infrastructure or configuration were often not communicated to the customers who use it. There was no control, for example, when services were migrated for maintenance. This could make planning hard as the rules were always changing. The cloud configurations could break existing best practice.

*““Cloud” usually means “someone else’s black box.” You’re playing by someone else’s rules, and since they keep changing their systems, they’re changing the rules, too. It’s hard to build best practices when the underlying mechanisms are evolving so rapidly and you’re not privy to the changes.” (Q6 8.1 line 100)*

The loss of control within IT to the business could result in best practice not being defined correctly, if at all.

*“Supportability operations are the last people to know. The business leads and drives what is selected. IT lost out on control to the business.”*

*It is hard for the IT department and reaches them too late.” (Q2 5.4 line 58)*

Davenport (1997, p.181) stated that organizational change could lead to IT changes or IT influence organizational changes. Davenport (1997, p.31) also argued that there needed to be a recognition of evolutionary change and managers did not understand how evolving information needed to be dealt with.

Bretschneider (2004, p.309) highlighted three important characteristics of best practice: a comparative process; an action; and a linkage between action and some outcome or goal. Checkland & Scholes (1999, p.277) discussed a systems approach that is not engineering or optimisation but a process of enquiry and learning.

Don & Priess (2008, p.22) recognised the complexity problem and in their view, as described by Brooks (1986), there were two types of complexity. ‘Fundamental complexity’ was described as stemming from business problems; and ‘accidental complexity’ as springing from the technological and architectural choices made. Both could be resolved by a skilled architect who could fully understand the business problems and could make the best decisions on manageability, to balance often conflicting requirements.

### **6.4.3 Summary**

Many examples have been given that showed that the adoption of best practices and procedures were affected by the complex interactions. The complex interactions were varied and changed depending on the precise situation, the people involved, the technology, understanding of the situation and technology and business. Some people follow best practice and ignore the situation while others were affected by their specific situation in the organisation. Sometimes it was not

possible to take a best practice route because of the interactions. The results of adoption of best practice could be positive or negative. The adoption was not only affected by the internal organization but also by the environment, cloud technologies being an example.

This research has shown that there is a need to deal with the complexity of management of database systems. An architect, alone, is rarely able to perform this task satisfactorily, at least not in the current world of big data and fast paced technological change. A consultation process with team leaders who fully understand the process in their respective teams seems to be the best way forward, with full collaboration between the teams during operation.

## 6.5 Improvement and Innovation

This section discusses the fourth and final research question:

How can a better understanding of the complex interactions contribute to improvement and innovation?

Improvement methods for database management were not followed by the majority. In the quantitative survey a free text survey question (Q80) reported how the respondents could improve practices and procedure. The top nine selected items are in Table 6.1.

**Table 6.1 Most important areas to add improvement for database management**  
*Q80 free text question*

<b>Business</b>	<b>Management</b>	<b>Understanding – Learning - Skills</b>	<b>People</b>
Better understanding of the business requirements	Database lifecycle management	Up to date training	Better communication at all levels

Business	Management	Understanding – Learning - Skills	People
Increased time	Organizational database roadmap		More employees
Increased budget	Improved documentation		

A respondent discussed “*employee turnover*” (Q80) which raised the issue of how to transfer best practices and procedures. Communicating why practices and procedures were important, could minimise the risk of people following them blindly. Respondents stated that “*all the various groups understanding why it’s important*” (Q80) and “*more visibility with other silos*” (Q80) are important areas to be considered. “*Consultation with subject matter experts*” to set up the best practices and procedures should include “*more rigorous testing*” (Q80). Formalization and standardisation of practices and procedures were raised as items which could also improve practices and procedures for database management.

O'Donovan (2014, p.5) discusses Seddon’s (2003) work as a way to improve service organizations using systems thinking.

The focus group research discussed improvement through the consolidation of resources (Q4 5.5 line 68) and the need to be “*given the time to understand the system and requirements*” (Q7 7.6 line 106)

Some participants used a continual service improvement method with ITIL.

*“I am a big fan of evolving best practices using ITIL style “Continual service improvement”. In this way a new problem would become a case for system improvement and maybe the addition of a new best practice.”*  
(Q8 8.2 line 104)



Gonnering (2011) argued that learning to use a continuous improvement method would reap benefits and that the outcomes would be emergent.

Database systems consist of many components which are interconnected. Although the ownership of the whole is often considered to be by one team, it was suggested that “*each ‘owns’ a piece of the database infrastructure.*” (Q8 8.3 line 105). This could encourage all the parties to work together. Communication was important with all parties, including the non-technical ones.

*“I think you do have to communicate to non-technical people who own the data, get their thoughts on how it is controlled.”* (Q8 3.4 line 55)

The data from both the quantitative survey and qualitative focus groups covered aspects of business requirements, from gaining a better understanding of them to finding a balance with management.

*“it has to be simple to be adhered to, so you have to get the actual on the ground developers to give a realistic expectation of what it is what is possible, without doubling the size of the team or making projects take twice as long and find a balance between the two but it is both. Managing solves and satisfies the business requirements, but it is actually the management from the day to day perspective and that is a very difficult balance to find.”* (Q8 3.2 line 60)

Prahalad (2010) suggested that best practice could be viewed as a means to catch up with market leaders and ‘next practices’ were about innovation and utilizing opportunities. He claimed that the development of next practices was constrained by the imagination of executives, and not by resources.

The business vision of database management in 10 years (Q81) provided a view of what the respondents thought was likely future based on their experience. The comments provided by the respondents were speculative. There were several main codes and themes within them, summarised in the following spray diagram (Figure 6.10).

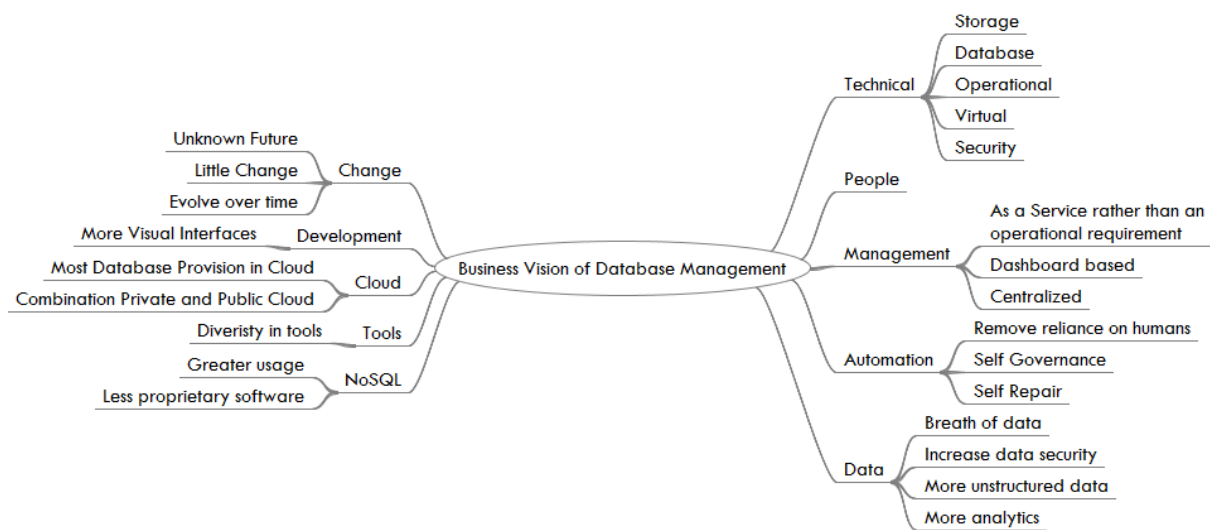


Figure 6.10 Spray diagram of a business vision of database management

### 6.5.1 Application of Lessons from the Research

This section discusses a way of addressing these areas, to help to improve the management of database systems. Based on the research findings, a blueprint for agile best practice is proposed: the CODEX (Control of Data EXpediently). In the CODEX acronym, C is for Control; O for control of Operations; D for data; E for EXpediently; and X for unpredictable events. The paragraphs that follow discuss how these five elements of the CODEX were drawn from the research findings. The research endeavours to propose a blueprint that can be used in the management of database systems. The CODEX is based on an evaluation of the influences of each data component each time there is a change. The value of this is that when the output of components is changed, this will be stored, rather than relying on the

memory of one person. Also, a set of use cases could be created. It is a way of potentially improving practices in the management of each database which could then be adjusted to a new set of circumstances.

The qualitative research enabled further detailed investigation of areas highlighted in the initial quantitative questions. The resulting critical information pointed to possible ways of improving the operation of large database systems. Database systems were evolving, transforming with various possibilities arising due to technological change (Abadi *et al.* 2016). To improve the management of database systems requires a shift in perspective to that of systems thinking (Ackoff 1981a; Checkland 1999) in order to view the database system in a holistic manner. Systems thinking is the belief that the behaviour of the whole can be understood in its entirety from the parts in a complex system (Capra & Luisi 2014). With that in mind the parts of the complex system are: organizational management, technology, operational management and data tasks. This innovative way of looking at database systems management could enable improvement within the system.

For the data collected and analysed there were various salient findings. The research has highlighted a number of issues relating to **control**. The principal results of control moving up the organization hierarchy is the loss of control by others. The purpose of control was not only the control of people, but the essential control of valuable data. There is a multileveled order between the teams involved in database management. Capra and Luisi (2014) continued to state different laws governed different levels of complexity and the lower levels did not have the same properties that could be seen in the higher levels of complexity. This concerned the enforcement of best practice and the control of preventative measures to avoid failure. Communication between teams and stakeholder communities to enable the choice of best practice, based on feedback, seemed the best way forward. The

discussions pointed to further complications related to constraints of cost and time. Capra and Luisi (2014, p.308) argued that communication networks have both material structure, such as technologies, and social networks, and this perspective needs to be considered.

The results showed that the **operational** management of database systems contained many tasks. Best practices and procedures were undertaken by many where they existed. Newer technology models and cloud services for database management had not yet been addressed. Sometimes operations were the last to know about changes. Technology components were fundamental to providing database management. Technological components were always a part of operations management, whether they were on premises or in the cloud. To take into account these issues it is the **control of operations** that should be managed. Different technologies added complexity. Mourad and Hussain (2014) suggested that adopting cloud based systems required a review of the best practices ITIL provided, to ensure the business was not exposed to unnecessary complexity.

**Data** was at the heart of any database management system and respondents linked the data requirements to managing database systems. Data life cycle management was not generally followed. The security and sharing of data, the diversity in data types and size of databases were where the complexity lay. Kandel *et al.* (2012) argued that the data sets collected each year were ever increasing in size and complexity.

**Expediency** is paramount in this fast changing world. The research has shown that agile was the most used development method. The 9<sup>th</sup> Annual state of Agile Survey (VersionOne 2015, p.2) reported that larger companies and more companies were

using agile to “*deliver software faster ,easier and smarter*”. The use of architecture frameworks was low and documented design patterns were not typically used. Having patterns or frameworks could potentially help improvement. For any future systems, keeping everything simple and easy to use would aid expediency. The research has shown that learning difficulty and lack of knowledge and skills influenced decisions for the chosen technology. This factor presented challenges to the management of database systems.

A final area that the research highlighted with unprecedented clarity was the **rate of change**. The rate of change of technology and changes to product selection and disruptive technologies can often govern what choices are available. Requirements gathering was only partially carried out, requirements continually changed and sometimes there were conflicting requirements. Other unpredictable events in the environment, competition from other businesses or business models may be affected by rapid change. There was a need to be flexible in the plan or road map. Amongst these changes lay the need for best practice and documentation. The largest area that respondents thought would provide improvement was documentation. There could be many forms of this.

A proposed way of addressing these areas, to help to improve the management of database systems, is the CODEX (Control of Data EXpediently) blueprint.

### **6.5.2 CODEX Blueprint**

The CODEX is a blueprint for database management which is described in this section. The acronym CODEX has been selected by analogy with the revolutionary introduction of the Codex (Netz & Noel 2007, pp.69–85) in the first century AD which changed the storage medium from a roll to a Codex (book format). This brought challenges migrating the data, but significant benefits of increased speed of

data access, reference and durability (of the parchment). Not all texts were migrated from rolls to Codex and those that were not migrated became defunct. Text case was changed from capitals to lowercase and minuscule copies made, resulting in further change; original majuscule manuscripts have not survived. This scholarly activity led to a revival in reading classic documents and a development of a centre of culture.

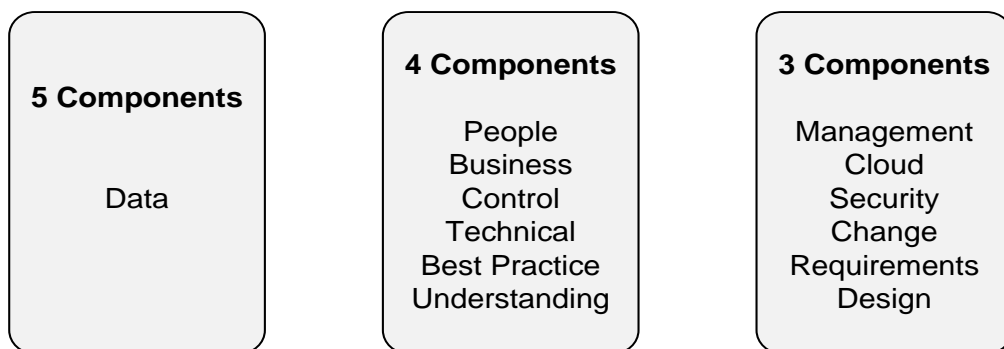
The CODEX is an output from the research, based on interpretation of the data. The main findings from the quantitative survey and qualitative focus groups are displayed in Table 6.2. Table 6.2 is a revised version of Table 5.13 also including a summary of the key quantitative findings. The table includes findings from the distribution of codes (Figure 5.5), code landscaping (Figure 5.8), data map (Table 5.5), code relations (Figure 5.9) and the findings from the quantitative data in Section 4.4. The left hand column shows how the results are collated.

Business	Management	Knowledge	People	Operations	Architectural	App Dev	Technical	
Business Cost	Best Practice Management		People Culture Control	Documentation		Development	Data Technical	Distribution of Codes (in 5 or more questions)
Business Company Plan Time	Best Practice	Know Think	People	Procedures Change	Systems Requirements	Design	Cloud Security Technology Data Storage	Code Landscaping (Data Corpus word counts above 42)
Business Cost	Best Practice Management	Understanding Complexity	People Control Communications	Change Support	Requirements	Design	Technical Data Security Cloud	Data Map (Inter-connections above 4)
Business	Management Best Practice	Understanding	People Stakeholders Control	Change	Architectural Requirements	Application Design	Technical Cloud Engine Data Security	Code Relations (Inter-connections 14 and above)
Plan Unpredictable	Management Efficiently Effectiveness	Knowledge Training	Communication Control				Technical Data	Quantitative (Key findings)

Table 6.2 Summary from Table 5.13 with the quantitative results.

These findings have been consolidated to show the most prevalent components. As an example, 'data' appears 5 times in the table whereas 'control' appears 4 times. In the 'Knowledge' column of Table 6.2, 'Knowledge', 'Know' and 'Understanding' all appear (once, once and twice, respectively). These have been consolidated together and labelled 'Understanding' with a count of 4.

The main components displayed in Table 6.2 that occur at least three times are shown in Figure 6.11.

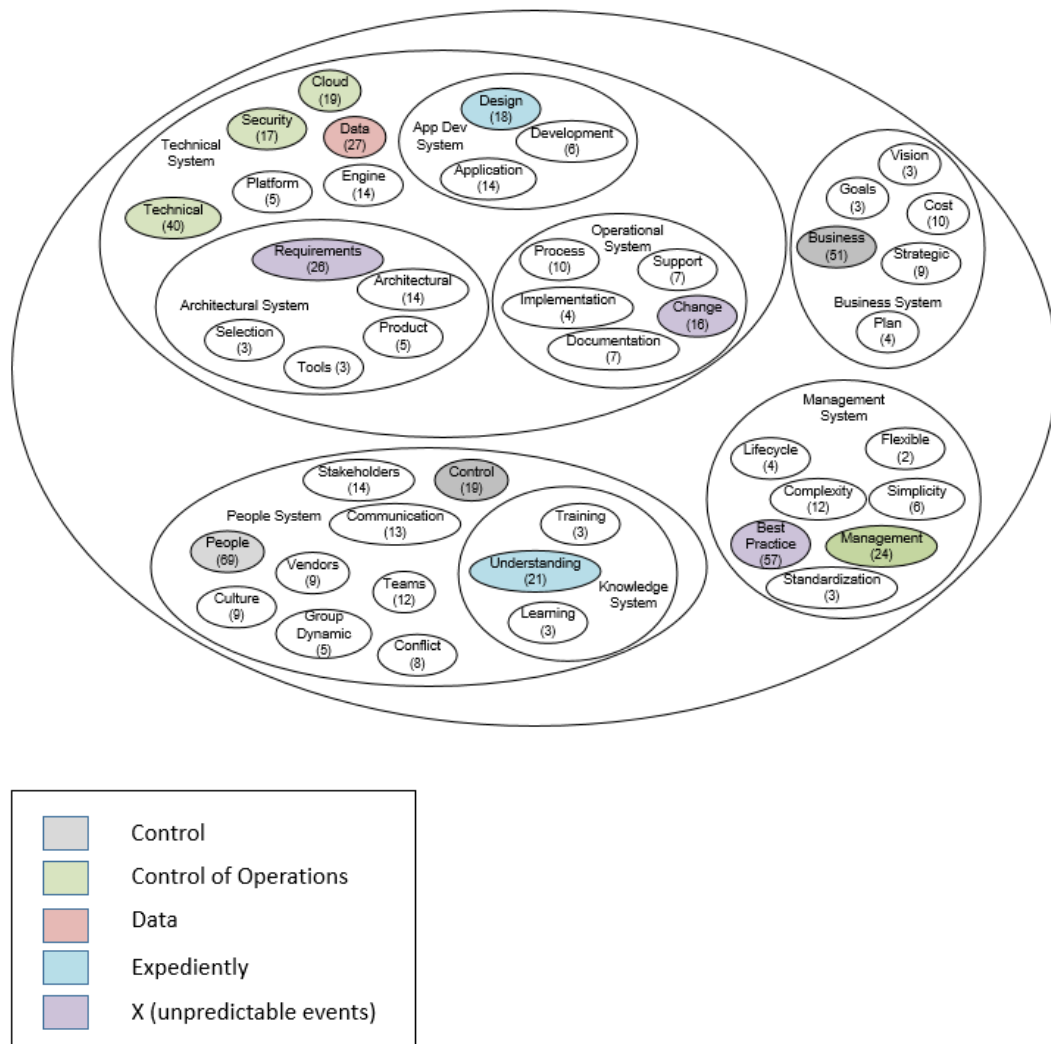


**Figure 6.11 Most prevalent components**

These findings (Figure 6.11) were then reviewed in a holistic manner with reference to the data and sorted and labelled as shown in Table 6.3. The components were allocated into the new groups as follows.

These groups were based on three findings: the systems map (Figure 5.12) which incorporates groups of components into systems and subsystems; the code relations chart (Figure 5.9) which is based on the total counts of influences presented in the data map (Table 5.5); and the operational model diagram (Figure 5.10) which showed how the components of the system work together.





**Figure 6.12 The systems map incorporating data from the code relations chart and showing (by colour) the CODEX groupings**

Figure 6.12 shows the systems map (Figure 5.12); with superimposed data from the code relations chart (Figure 5.9), giving the total numbers of interconnections. The shaded (coloured) components are those with the most interconnections (Figure 6.11). The shaded colour coding provides a key to the final CODEX groupings, described below.

In the systems map (Figure 5.12) *people* (4) and *business* (4) have a system of their own and were placed in the **control** new group. Control was in the people system in the systems map (Figure 5.12). In the code relations chart (Figure 5.9) the two

highest set of influences were *people* with a count of 69 influences and *control* with 19 influences. Thus *control* was placed in the group **control**.

*Technical* (4) has its own system and was placed on **control of operations**. *Cloud* (3) and *security* (3) are all in the Technical system in the systems map (Figure 5.12) so were grouped together under **Control of Operations**. *Management* (3), in the context of the research, is about managing the database system and has its own management system in (Figure 5.12). With best practice having been classified into another group, *management* influences from the code relations chart (Figure 5.9) totalled 24 was reclassified to the same group (**Control of Operations**) as technical, cloud and security.

*Data* is a core component and had a count of (5). As this was the only count of 5 it was allocated its own group: **Data**

*Understanding* (4) and *design* (3) are in two separate subsystems. *Understanding* is in the Knowledge subsystem (part of the People system) and *design* in the App Dev subsystem (part of the Technical system). In the operational model diagram (Figure 5.10) the App Dev system has no connection to the Knowledge system. The App Dev and Knowledge systems appear to have no communication with other parts so *understanding* and *design* are grouped together and labelled **Expediently**.

The research data talks about *best practice* (4) needing to *change* (3) continuously. In the code relations chart (Figure 5.9), based on the data map (Appendix F), *best practice* had a count of 57 influences and *change* had a count of 16 influences.

From the systems map (Figure 5.12) *best practice* is in the Management system and *change* in the Operational system. Thus these two items were placed together in a group labelled **X** (unpredictable events). *Requirements* (3) is in the architectural system and they are unpredictable. *Requirements* had the highest count of 26

influences in that system. In summary Table 6.3 the shows how the components are grouped and the reason for the grouping. The counts of connections to other components shown are for the most prevalent components in Figure 6.11.

**Table 6.3 Creation of the CODEX**

<b>Control</b> <i>The interconnections are evident from the influence diagram linking people, business and control.</i>	<b>Control Of Operations</b> <i>This grouping relates to technical components and management of those technical components.</i>	<b>Data</b> <i>Data is the most prevalent component.</i>	<b>Expediently</b> <i>Expediency can only be achieved by understanding and suitable design of the operational model.</i>	<b>X (unpredictable Events)</b> <i>Changing circumstances, changing input and output requirements, and hence changing best practices, are unpredictable as a database ages, so cannot be known when the database is set up.</i>
People (4)	Technical (4)	Data (5)	Understanding (4)	Best Practice (4)
Business (4)	Management (3)		Design (3)	Change (3)
Control (4)	Cloud (3)			Requirements (3)
	Security (3)			

The CODEX (Control of Data EXpediently) blueprint was then created. The acronym is constructed as: C for control; O for control of Operations; D for data; E for expediently; and X for unpredictable events. It is an acronym for a system or way of controlling operations and data in a rapid, efficient and accurate manner.

The CODEX is a blueprint for database management improvement and innovation which is given in Figure 6.13.

CODEX (Control of Data EXpediently)

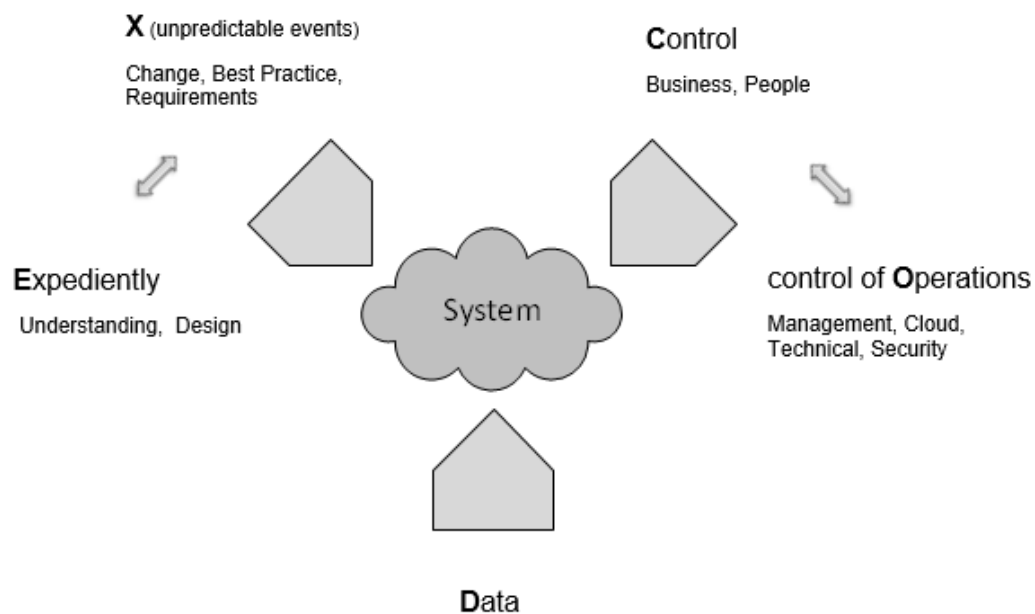
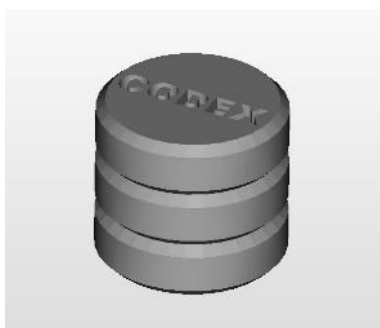


Figure 6.13 The database system CODEX

This suggested CODEX (Control of Data EXpediently) is a pattern to help improvement in the management of database systems. An important part of the CODEX is that over time the complex interactions of components needs to change. Between the diverse components and time, continuous feedback is required from all elements of the system.



The introduction of a CODEX to help in the management of database systems consists of multiple inputs and was based, in the core, on best practice and complex component interactions. The CODEX is the usage of knowledge through control, operations,

data, expediency and diverse environment variable factors. There is no starting reference point other than the objective of considering database systems management.

When a component is changed there will be a set of components that are influenced by the change. Using the CODEX will enable a record of the changed components to be kept. The CODEX reference pattern should enable improvement within the system.

The CODEX has two parts. The continual data collection that records which components were affected by a change for an actual completed task, to create a lookup table (data map). The second part looks up the component in the data map to give you the list of connected components for the management of that part of the database system. Being able to change and quickly modify best practice could be advantageous. The data collection takes the known inputs and known outcomes of the influences which are recorded over time. Once the new components are determined, the CODEX could use the outputs to identify the complex components that would be affected for this system. When a change occurs you would be able to arrive at the new set of complex components required after consideration of the interactions. The five inputs into the CODEX are all required for every piece of database management work. These five inputs are described in the paragraphs below.

**C.** An important step in any database system is the **control** system: defining the business needs, budget, controlling the people and time factors. People are important in the management of the database system. It involves the stakeholders and the teams working together to achieve a single goal. The culture driving this collaborative venture forward will undoubtedly raise conflict, but this should be integrated with a high level of communication with all levels in management, the stakeholders, the teams and data and database staff. Also the governance related to data and data quality should be controlled.

**O.** Control of **operations** of the database system is the core day to day running of management tasks, the processes and the performance of the system, orchestrating management through automated and self-managing systems where possible. Technical management needs to understand how internal and external technologies integrate. This is vital when using cloud technologies because internal managers have no control of the details. All of these operations require security to be considered to protect the data.

**D.** The increasing volume of **data** acquired today requires storage in computer based systems. Thus databases have developed to satisfy the current demand not only for storage but also to provide information quickly and accurately. The variety of data, big or small, requires governance and has a purpose. The reporting and visualization of data is key to enhance business ability to grow, adapt and understand the complexity. Data is continually changing and more of it needs to be stored to meet the demands of society. Being able to understand the data for it to be available and useful, is a core requirement to improve and innovate.

**E. Expediency** is driven from the need to have efficient control over costs, speed of delivery and change. Designing database systems that are easy to manage, simple and agile utilising reference architectures and blueprints is key to performing expediently. To be able to proceed the critical factors are knowledge, skills, learning, leading to understanding and allowing planning to unfold unhindered. Development can lead to fast performing applications and efficient management through automation.

**X.** With any system and particularly in a diverse and ubiquitous database, systems change is always happening, be it with the number and type of database platforms,

the new technologies, global business or environment change. Change is rapid and diverse. Using patterns and always establishing best practice will help in the management of database systems. These best practices need to be able to rapidly change as requirements change or are not known at the outset. Producing documentation that can be automatically created is key for accuracy and ensuring documentation is available. Also unpredictable events can occur and any changes to the components must be documented and changes to all respective components made. There is continuous feedback over time.

Outputs of the CODEX are adaptive, have emergent properties, can be complex or chaotic. Best practices and procedures used in the management of database systems require a continuous feedback loop. Thus the CODEX could produce a system of checks that is useful at the start of each new task, based on comprehensive reports from the people involved in the design and operation.

The effectiveness of the CODEX depends on the data collected and the pattern of the blueprint. The data provided in each of the elements of the CODEX is likely to evolve and grow over time as the environments change and due to feedback from the database system. The feedback creates new inputs to the system. Following the input of the components the output might suggest best practices to be refined.

This model or blueprint could continue to grow, adapt to new inputs and change over time. The blueprint could be used to eventually map the connected components involved in the management of database system. The continually evolving system will be mapped, thus the complexity could be reduced as the system becomes understood.

### 6.5.3 The Working CODEX

The CODEX suggested provides an overview of the database system, which has become more complicated as new developments have led to frequent changes. The CODEX is a simplified way of understanding the processes needed in order to produce satisfactory outputs without the assistance of more automatic processes.

The CODEX takes the five inputs and already collected outputs of connected components in the data map to predict the components that need to be looked at, in relation to a particular set of inputs.

Once the requirements for a new database are known, a first run scenario can be proposed using the CODEX, after considering the details suggested and how each component can fit together. At this point it may become obvious that either (a) the first run scenario is acceptable, or (b) not all the requirements can be met, possibly due to conflicting factors such as cost and staff availability, or (c) the initial requirements have changed. The first run would have:

Initial practice  $C_1 O_1 D_1 E_1 X_1$

Thus a second scenario would have to be drawn up if (b) or (c) occurred, until the best compromise could be reached. However it is vitally important that changing any part of the planned CODEX necessitates the reconstruction of how the whole system links together. Subsequent changes in any factor, would require re-examination of the CODEX due to the complexity of the interactions that take place between the components. The codex would produce:

Revised practice  $C_2 O_2 D_2 E_2 X_2$

An example scenario for the purpose of explaining how the CODEX works is given below, based on the data collected within this research.



Scenario Requirements

The business is looking to migrate away from old physical servers and create a virtual database server. The data that is to be stored is relational but needs fast access and column storage data methods and the use of In-memory technology would meet the requirements. This requirement means the business has to take advantage of the fast moving technology change and use a new database engine. The new database engine needs to be learnt by the employees.

Breaking this scenario down into five parts:

The stakeholder is the business which controls the situation: “The business is looking to migrate away”. The technical hardware component: “create a virtual database server”. The data type: “The data that is to be stored is relational”. Learn the new technology: “The new database engine needs to be learnt by the employees”. X is changing technology: “business has to take advantage of the fast moving technology change and use a new database engine”

These five inputs described above can be summarised in Table 6.4. The items in italics are the key CODEX components from Table 6.3.

**Table 6.4 Data input features**

<b>Control</b>	<b>Control Of operations</b>	<b>Data</b>	<b>Expediently</b>	<b>X (unpredictable events)</b>
<i>business, people</i>	<i>management, cloud, technical, security</i>	<i>data</i>	<i>understanding, design</i>	<i>change, best practice</i>
Stakeholder: business	Technical: virtual	Data: relational data	Learn: engine	Change: fast moving technology

These are then taken as the ‘component A’ inputs in the Data Map (Table 5.4).

Component A components are the components that influence component B

components. A summary of the influences for this scenario is shown in Table 6.5.

This table shows the influences from the inputs but does not show whether the influences are two way. This can be determined from looking at the Data Map in Appendix F.

**Table 6.5 Component influences from the scenario**

Stakeholder	Technical	Data	Learn	Change
Data	Application	Cloud	People	Data
Requirements	Design	Engine	Technical	Technical
Change	Data	Security	Architectural	Requirements
Goals	Cloud	Technical		Support
Plan	Requirements	Requirements		Plan
Vision	Support	Process		People
Control	Training	Support		Communication
Group dynamics	Understanding	Change		Standardization
Best practice	Learning	Cost		Best practice
Management	Standardization	Business		Management
Implementation	Implementation	Teams		
	Documentation	Conflict		
	Change	Complexity		
	Cost	Best practice		
	Plan	Management		
	Business			
	People			
	Teams			
	Communication			
	Complexity			
	Best practice			

Stakeholder	Technical	Data	Learn	Change
	Management			
	Simplicity			

This is a holistic approach which provides a list of the components that are affected by the situation in the scenario. Knowing the list of affected components could help improve the quality of the management by ensuring the connected components are reviewed. Every time there is another scenario, there will be a new set of five inputs, and this will produce a new version of Table 6.5. These five inputs and output tables are stored. After collecting these inputs and outputs, duplication may start to appear and cases where the input is shared by multiple scenarios. This could eventually result in Figure 6.14. The components are shown in block capitals with the scenario input. An example of how different inputs into the system may link together is also shown.

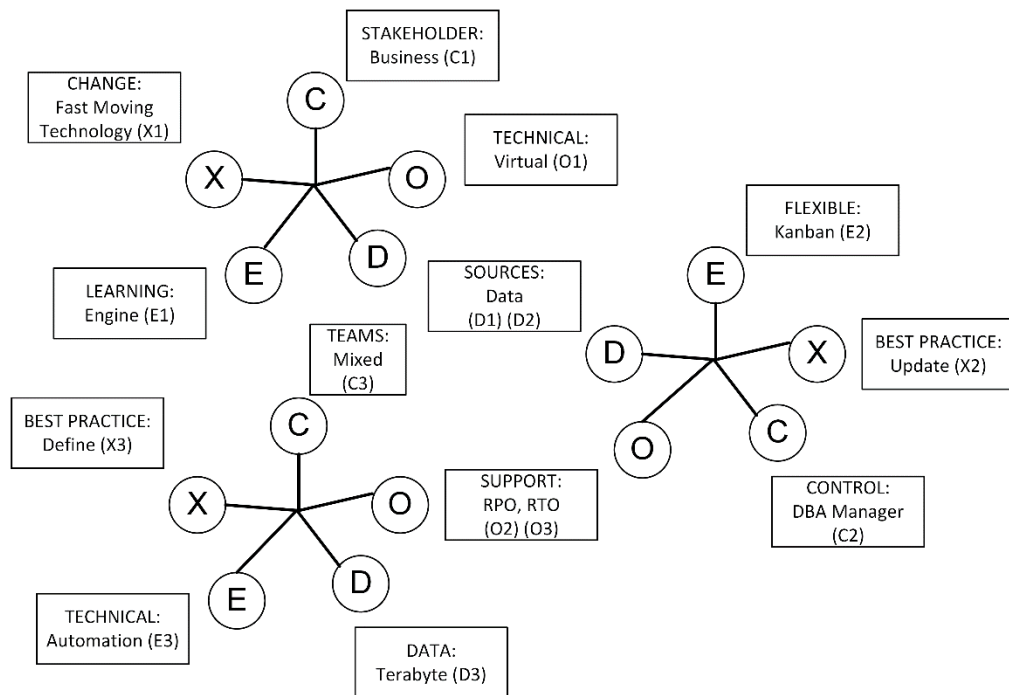


Figure 6.14 Codex component linkage example of three inputs scenarios

Figure 6.14 shows an example of three sets of component inputs scenario 1 ( $C_1 O_1 D_1 E_1 X_1$ ); scenario 2 ( $C_2 O_2 D_2 E_2 X_2$ ); scenario 3 ( $C_3 O_3 D_3 E_3 X_3$ ). There will be many possible inputs but this example explains, using three scenarios, how this might work in practice. After scenario 1 is drawn up from the inputs in Table 6.4, there may be a change in the type of data ( $D_1$  changed to  $D_2$ ) to be input. This change means that in order to draw up scenario 2, consideration needs to be given to each component of the CODEX. One change could well influence other components of the system giving a complete revision for scenario 2. At a later date, the client may request a change in availability target ( $O_2$  changed to  $O_3$ ). Scenario 3 will then need to be drawn up again looking at all the influences affecting each part of the CODEX, until a new compromise is reached.

When real data is added it would include items to finer technical granularity and the list of influences would vary in length. The data is continually collected and the number of influences will continuously change. This will be represented in the data map in Appendix F.

The best working solution is to refer to the complete data map to take into account all of the influences and interactions between the components that can occur. This can only be done successfully by computer as it would be too time consuming otherwise. There could potentially be at least 1936 entries from the 44 components in Appendix F. Hence the need for automation tools such as machine learning or graph theory.

Analysing the system with all of the very large number of components (each of which influences at least three other components and is influenced by at least three other components) may be impossible in one step. If further research is to succeed in processing algorithms that will define and solve the entire system it will probably

be necessary to take a multistep approach. A start could be made by considering a small number (say 5) of the most important components. Including only the influences that each of the components have on the other five, it should be possible to produce an algorithm that would solve this limited system. The algorithm would need to be extended to take account of more components for a complete solution.

## 6.6 Summary

This chapter brought together the quantitative and qualitative research to discuss the overall findings. The discussion considered all four research questions and answered each in turn: Question 1 discussing best practice usage; Question 2 considering complex interactions; Question 3 examining further the adoption of best practices; and Question 4 exploring improvement and innovation.

The chapter then used an analogy of a codex, a quire of manuscript pages attached together in the form of book. The CODEX (Control of Data EXpediently) is a system or blueprint for controlling database management in a rapid, efficient and accurate manner. The CODEX blueprint was explained and an example was given of how this might work.

The research has clearly identified that database management is not a singular activity as it once was, and many complex interactions need to be managed to improve and innovate and to understand the continually changing system. Currently there are checklists, guidelines, best practices and standards. These all help to ensure database management is effective and meets its goals. However, the lack of framework usage identified in the research suggests that checklists may only be a part of the solution. If an organization is too process driven it could lead to a lack of spontaneity and innovation. The research also identified communication issues

between teams. The complexity of current technology means it is vital for subject matter experts in different teams to work together effectively.

The final chapter summarises the key findings and the contribution to knowledge from the research. It also discusses possible future work and how improvement and innovation could be achieved in the field of database management.

## Chapter 7: Research Conclusions

### 7.1 Introduction

The development of information technology has meant that data can be processed quickly and accurately. Database systems today are used for many purposes and have progressed to offer a great variety of information for business, for science, for governments and for society in general. The explosion of data and the speed of technological improvements has meant that currently database systems face continual change.

For these reasons the complexity of database systems has increased to such an extent that best practices in operation require management to be aware of up to date software and hardware and to understand the objectives of the people involved. To enable managers to make informed choices for specific requirements for each individual database system, understanding of best practices is required. This research studied the uses of best practices in database systems, revealing the complexity of these systems. Socio-technical aspects have emerged as a key factor.

'Best practice' can be problematic and is a confused and contested concept. The perception of best practice which emerged from this research is different from the way the concept is used in industry, and this is backed by the empirical evidence collected. Recommendations for best practice are made by various groups such as vendors, different organizational groups or industry experts. This diverse set of stakeholders who own and control the best practice can cause conflict as technology changes, cross boundary goals change and people disagree as to

whose best practice should take priority. Best practices are embedded into every part of the end-to-end management of database systems, and perhaps best practices are among hurdles that people face when trying to improve their own technical systems. An example of internal team conflict that can occur in relation to best practice is given in the following quote from one of the interviewees:

*“A new project wants specific storage requirements for their project such as a “fast track” design. They design the logical solution on this basis, but the storage team does not agree with fast track as a principle and force the design to fit on a shared SAN. The solution goes live and the solution brings the SAN to its knees. The storage team panics and throttles the solution so it performs really slowly.” (Q9 8.2 line 88)*

The dilemma is: who has the right to define these best practice recommendations that affect the business? Best practice can lead to more complexity and can stop people thinking for themselves, which can reduce their effectiveness and application of judgement on what is, or is not, beneficial to the organization.

This chapter discusses the contribution to knowledge and summarises the key findings for the four research questions. The main outputs from the research, including the CODEX blueprint, are discussed, along with the methodological contributions this work has introduced. In terms of future work, the chapter proposes future data science work, using either machine learning or graph theory, to automate predication of the CODEX blueprint.

## 7.2 Contribution to Knowledge

The contributions to knowledge are as follows.



- An innovative method for qualitative data analysis was developed, combining thematic analysis with systems diagramming. This method was found valuable in analysing the complex field of database management, and proposing suggestions for improvements. The thesis demonstrates in detail how this new approach can be used to gain insights.
- An in-depth data map has been created detailing the multiplicity of interconnected complex components. The data map can be 'mined' to give detailed information about how the many different aspects of the database management field relate to each other, and how best practice fits into this complex picture.
- The CODEX is a proposed approach for deciphering the complexity of interconnections, and has the potential to create an autonomous way to deliver management of database systems. The CODEX could be the building block of a multi-layered system, one that enables organisations to reflect upon their practices in database management in the light of their past experiences (and those of others) and improve them.

The principal findings of the research are:

- Plurality of sources of materials, technology, skills, teams of people, locations, and the related interconnectedness means that what were once straightforward management techniques need careful control.
- Constant, rapid change is one of the most difficult challenges in the management of database systems.
- Complexity makes it difficult to define and use best practices for the management of database systems.
- The satisfactory control of all components is a sociotechnical problem i.e. the people aspect is as important as the technology aspect.

## 7.3 Summary of Key Findings

This research was a study into the best practices and procedures used in the management of database systems. The purpose of the research was to improve the management of database systems to avoid failures experienced in the past; and to be effective, efficient and perform well. It examined four questions:

- To what extent are best practices and procedures utilised by the database community?
- What are the complex interactions that are an integral part of the management of database systems?
- Is the adoption of best practices and procedures affected by the complex interactions that are an integral part of the management of database systems?
- How can a better understanding of the complex interactions contribute to improvement and innovation?

As part of the research it identified the complexities and socio-technical factors that exist within the database system and proposed a CODEX (Control of Data EXpediently).

### 7.3.1 Best Practices and Procedures Utilised by the Database Community

The first research question is discussed in this section. The research investigated which best practices were used in the management of database systems. Best practices are central to the research undertaken. A quantitative survey was undertaken to examine the current best practices used by the database community. The survey was a key part of the findings as this highlighted the current usage of best practices across the entire management lifecycle.

The findings covered a number of issues. Best practices could be found in diverse locations from within organizations and externally. There was no clear definition of best practices, and it could be seen as a guideline. Best practices were controlled in different areas to different levels, and could cause issues that it might have been trying to prevent. Additionally there were some areas where best practices were not used by the respondents.

There was no consistent method of database training across organisations. Access to technical database training materials was available from different sources and locations. There was diversity within the industry and within organizations for application products, type of database engine used, and platforms. The size of the data management techniques used by multiple teams to ensure the data was available, recoverable and of acceptable quality.

Database administrators or database managers controlled database management choices. However the on-premises database software was controlled by database administrators or database managers, while cloud database software adoption was controlled by the Head of IT Operations. This could be partially to do with the adoption of cloud technology being seen as a business wide financial decision rather than one of database management. The choice of practices and procedures determined the performance operation of a database.

Producing a framework seems unlikely to be the way forward to improve the situation; there are many frameworks that already exist. The survey showed that while the service management framework ITIL was used for the management of database systems, other frameworks that cover parts of the database system were not used, or were maybe outside the sphere of the respondents who participated in the data collection. Architectural frameworks were reported as mostly not used by

the respondents. It is clear that the actual time spent managing the database was only a small part of the workload.

Database management could often be undertaken by specialised teams of DBAs. Some of the more complex management tasks required the integration of technical components to be carried out by subject matter experts in the current technical area. To ensure all the parts of the database were managed, many teams needed to interact and collaborate. Communication decreased as database management became less visible, and was often too late to contribute to the process.

Agility in the changing environment, flexibility and simplicity of any designs were important to address the speed of learning required to keep up with technical advances.

### **7.3.2 Complex Interactions of the Management of Database Systems**

The second research question is discussed in this section. An in-depth analysis of the qualitative data collected from focus group respondents provided greater insight. The focus group data was analysed using thematic analysis and through the use of a transitional section shifted to use systems approaches to present overarching findings on complex interactions.

The systems map in Figure 5.15 presents the database as four main systems with four subsystems. These were Technical System (Architectural subsystem, App Dev subsystem and Operational subsystem), People System (Knowledge subsystem), Business System and Management System. The systems map presents an overview of the whole database system and offers an understanding of how the separate systems need to work together.

The influence diagram was developed from the systems map and shows the database system complexity in Figure 6.10. The influence diagrams are all strongly grounded in direct quotes from participants. This derivation of those influences is at a higher level of abstraction. The influence diagram presented the highly complex database system with best practices at the centre. The most prevalent influences to and from best practices were management, business and people. Various other two way influences were clear: people and understanding, best practices and control, technical and complexity, security and data. The most numerous influences show the strength of the relationships.

### **7.3.3 Adoption of best practices and procedures affected by the complex interactions**

The third research question is discussed in this section. The research has demonstrated that the database management system is diverse, multifaceted, continually changing and complex. There were many technical layers which introduced management complexity where different teams interacted together. Differing levels of knowledge and best practices were combined with new technologies. New technologies were introduced continually, sometimes due to cost or other non-technical factors such as staff skills.

Managing and securing the data in this diverse landscape brought challenges. Architectural frameworks might not be followed in the design process, due to being considered too complex; or modern designs might not be used due to the preferences for well understood products which reduce the risk to the business and require no new skills to be learnt. Requirements could conflict, change and be restricted due to cost. Skills of the staff and their commitment and ability to adapt to change was an important factor in the management of database systems.

Development of systems that use databases could benefit from understanding both development and operations. Sometimes there was unwillingness of developers to talk to production teams. The operational systems were a core to keeping the everyday database systems up and running. Operational teams require things such as quality gates to ensure chaos does not ensue. Often a part of this was comprehensive documentation and processes, evolving knowledge base articles but needing some control to be maintained over this. Operational teams needed visibility, otherwise 'silo' systems and teams might develop. Best practices might guide everything and needed to be embedded in the systems that were put in place.

People are involved in the entire systems from beginning to end. They include stakeholders who could have their own agenda, organizational teams with cross team requirements that might be obstructive, vendors and administrators. The best practices that database administrators and other administrators in other areas used might be counterproductive and not work as a whole. Conflicting resources managed by budget holders and best practices that administrators wanted to deploy could cause challenges. There could be a lack of understanding between people of how the database environment needed to be configured to work well. The culture in some organizations was that of an "ivory tower". Practical everyday considerations were dealt with in isolation. It was the people who have to implement best practice who needed knowledge of how to use current frameworks in a timely manner.

The business system included many facets. The business political interests, budget and internal culture could cause problems with management and influences whether best practices were adopted. The business model may be adapted to market needs and changed rapidly if disruptive technology caused uncertainty. Supportability of systems was quite often an afterthought. The organization's lack of appreciation of their teams was counterproductive to providing sound architectures. The

proliferation of disparate technologies should be controlled to restrict diversity to help ensure that mistakes were prevented and the best quality of database management was maintained.

Managers might not understand their staff's weaknesses in learning or that they were working at the edge of their technical ability and these challenges might affect database management. The knowledge transfer of how to manage these systems could be problematic. Management of these factors was complex and if best practice was not easy to use people would find a way round it. Keeping this simple was important to achieving success. An important factor to consider was that the data that was being managed had a longer life expectancy than a particular management system. Best practice was not always appropriate for all scenarios and it was the overlap between technology and all layers that added complexity.

In summary, database systems are heavily reliant on the people (database staff, teams and stakeholders) with control and communication highlighted. The study showed that many best practices and procedures were used, but the implementation of best practices can have good or bad implications for practice. For mature databases it was found that best practices and procedures avoided relearning mistakes. For newer database engines, best practices did not exist and they might take some time to be developed and improved to suit the individual cases. Lack of understanding by staff was a factor shown to create problems due to fast moving technical changes.

The complex interactions involving the themes in the whole database system affected the outcome: "*the biggest thing that comes to mind is it is not a technology problem but a people problem*" (Q5 8.3 line 109). Complexity has had a huge effect on the management of database systems. It has meant that although best practices

were useful, it was not possible to predict the outcome. It required careful communication in the changing world to try to achieve a level of certainty of outcomes. Managing between teams required continual communication. The visibility of database management tasks between business and the relevant management teams was important. Although the usage of cloud technologies was controlled by the IT director, it was important that the IT director had a full understanding of the technical and management complexity, as these decisions could significantly affect what management could be undertaken in the database system. Documentation for processes and procedures was poor and required improvement but was only part of the learning required. A knowledge of the business intricacies was also required.

#### **7.3.4 Contribution to Improvement and Innovation**

The fourth research question is discussed in this section. The concept of making small changes to existing systems for improvement could be achieved within the management of database systems. The continual changes to technology and the introduction of new technology forces change on business and database management. Some of these changes are improvements, others not. The database management epoch began with relatively small changes over a long period of time and is now escalating, and the level of disorganization and lack of order is increasing rapidly. The vision that there is a 'data culture' which can drive business decisions added more layers to managing these database systems. The equilibrium where database teams only look after the database and people consume the data is long gone. The shift in this pattern of behaviour is causing an emergence of new requirements. Change was required to either methods or ideas to produce innovation within the management of database systems.



The quantitative data analysis has shown that pressures of cost and time was a problem that businesses are struggling with. Managers needed to establish a culture of communication and collaboration between staff who were encouraged to take regular training. These actions could help improve the management quality. The quantitative data confirmed these problems to be largely down to management control. Best practice is controlled to some extent for on-premises database maintenance, security and resilience. The question arises whether 'control' of the management of database systems can be achieved, in particular now that businesses are choosing cloud suppliers. With the shift in control of decision making from database administrators to the Head of IT Operations, the technical understanding of how database management works needs to be considered to improve the system. The survey found that senior IT managers could become removed from the detailed operations of a database, and this suggests poor communication with the people involved. As more databases contain big data, a controlling manager needs to have up to date knowledge of database software to enable the best choice of engine and tools for the business. These issues need to be addressed to improve management.

It is clear that database problems have been recognized for some time but it seems that progress to a completely satisfactory outcome has not been achieved.

Continual improvement in the technology and software products has not proved to be the solution. Yet the training given to database staff has not moved forward. This has been shown to be an area which needs improvement.

Question 4 of the qualitative focus groups asked "In what ways do you think that best practices and procedures could assist management of the database lifecycle?" and question 80 of the survey asked "Which of the following could improve practices and procedures for database management in your organization?" The following

improvements were suggested relating to the following areas: business requirements including road maps; time; type of business; budget; a communicative and understood vision; group dynamics of cross teams; control; decision making; simplicity; speed of change; data, communication; best practice; understanding; knowledge; skills and database lifecycle management including technical hardware; software; security and tools. For improvement and innovation to take place, an understanding of all the complex interactions in the system is required. It was however not possible for learning to be a one off event. The technology, the interactions and global economics are continually changing at speed so a new way of looking at the situation is required.

## 7.4 The CODEX

The database industry and academia has invested time to research the technical aspects of databases, but rather less time on the systemic management of databases. With this in mind this research proposed the CODEX (Control of Data EXpediently), a blueprint to help improve the management of database systems. The CODEX provides a mechanism for effective efficient control. Successive operations of a database system are likely to involve one or more changes. Applying the logic of the CODEX blueprint ensures that any essential change is carefully assessed on its effect on other components in the system. Thus these can be adapted accordingly to give an agile best practice solution. How the CODEX works is detailed in Section 6.5.3 and 6.5.4.

The following aspects need to be considered: people; business; knowledge and skills; technology; operational practices; architecture and development. Collaboration industry wide would be worthwhile although competition may prevent this. It is with great difficulty that managing databases successfully with all these

interconnections can be achieved. It is recognised that, due to the sheer complexity of most databases systems, it may take considerable time to set up this method. However, the research has produced valuable information through the code book and data map and this information could be used quickly, using a computer application for this purpose. The data map is a clear reminder of which data components affects others. This complexity is currently best managed by very experienced personnel i.e. those who remember much of the information of causal connections and relationships from previous database performance issues. The CODEX would reduce the reliance on memory.

A brief heuristic version of the CODEX requires consideration of the questions:

- Have the most suitable technical choices been made?
- Does the workforce have the skills required?
- Have the recommended practices and procedures been decided following communication between those involved and regular evaluation of documentation?
- Can the demands of the total environment be met by business management due to the complexity of the system?
- How are the best practices and knowledge transferred between new and existing employees?
- Should this be a process of studying patterns, learning the structure and application?

The CODEX proposes a way to record the connections so that the management burden is reduced and reliable results can be produced and reproduced. The increased communication required for the CODEX could be seen to increase, rather than decrease, the management burden because the initial increase in

communication is time consuming. But it will hopefully reduce the overall time needed for operational management by reducing duplication of tasks, errors or having to rework tasks. Moreover, increased communication has been shown, in agile methods such as Kanban, to bring additional benefits of insights and cultural cohesion between and across teams.

There are various limitations that could arise from using the CODEX. The use of the CODEX is for the management of database systems. Other teams will have their own management practices which may not fit in. The control of those practices outside the database system is unlikely to be possible. A good implementation strategy would be necessary to communicate how the CODEX works. Collecting all the variables and documenting them is a significant amount of work and still has the risk of not identifying all connected variables and changing variables. More transparency through communication would be required between database systems teams and other team.

The core principles that need to be adopted are a regular review of the strategy for collecting the data for best practice from interconnected components. A review of the effectiveness of the interconnected outputs should be undertaken to assess whether these are helping to improve the management in the field. A method of communicating the rapidly changing technologies needs to be created, to ensure the CODEX remained valid and sustainable as a working blueprint. Using the CODEX may have implications for other teams, to ensure best practices work well. The outcomes are novel because they are achieved by being able to clarify what is needed for improvement with reference to each component when managing database systems. This could clarify what tasks were required, what processes need to be followed, any regulation requirements, and an idea of the business areas that may be affected.

## 7.5 Implications for Method

The methodology used in this research drew on a number of methods and enhanced those presented. The research design using mixed methods gave a better understanding of the situation than a single method alone. Traditional quantitative research was followed by a connecting phase to the qualitative research. The qualitative research used thematic analysis as described by Braun and Clarke (2006) and an overarching design adapted from the work of Saldana (2013). Then the methodology diverged and added a synthesis aspect through systems methods. The transitional process (between thematic analysis and systems thinking) included: code landscaping, code relations chart and an operational model diagram. The final systems thinking stage included a systems map, influence diagrams to show the interactions between components, and an overarching influence diagram. Applying systems thinking, a way of describing the world (Checkland 1983, p.671), is a crucial change to allow improvement and innovation within the system. Management of database systems is a large area connected to many components so using systems thinking and in particular using system diagramming enabled a holistic view to be taken. This research demonstrated that applying systems thinking enabled the complex interactions to be identified. Identifying the holistic view can help improve the entire end to end database system and uncover the various interconnections.

## 7.6 Future Work

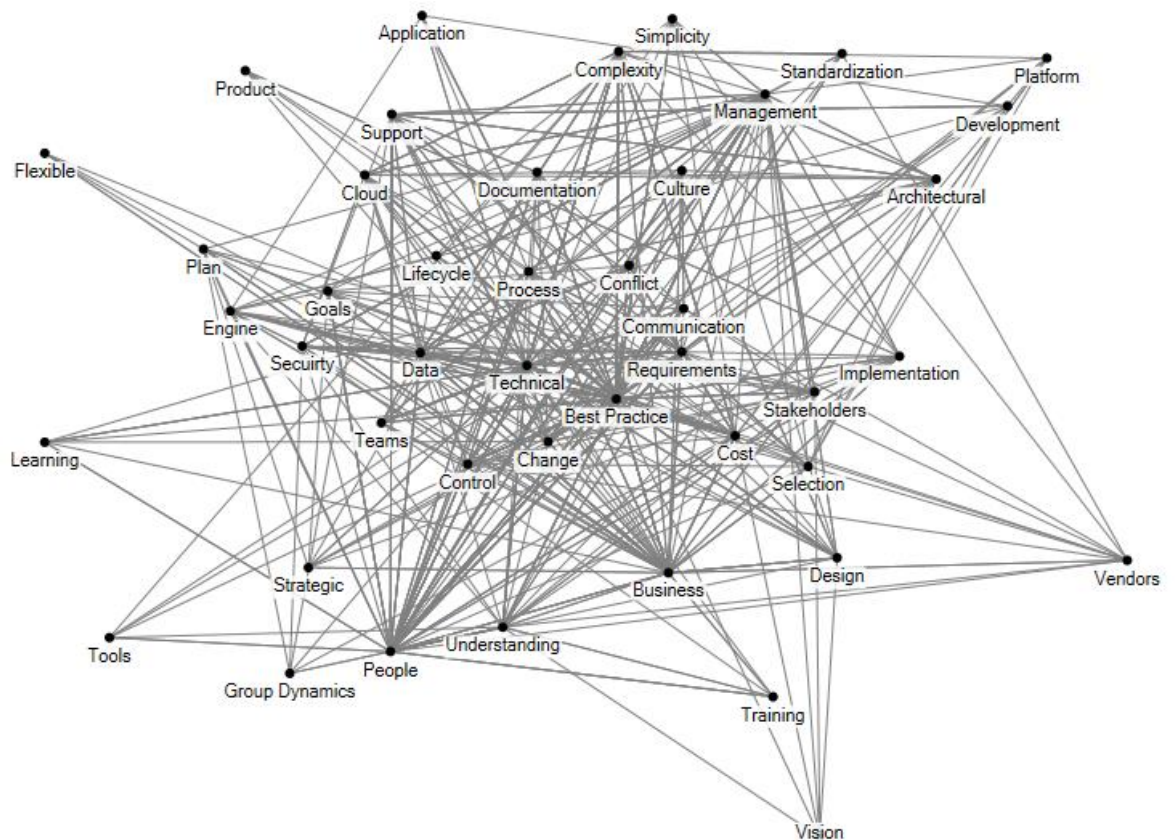
To assist with the improvement of the management of database systems, there is future work that could be undertaken. To facilitate this improvement, the database system needs to be reviewed in a holistic manner. The research investigation has

shown that before the details of the requirements and the processes involved were fully understood at the start (even before they were agreed), it was advisable to discuss key individual parts to be included. Only a very knowledgeable manager could perform this function today due to the increased complexity of the system.

Once a system is set up, this is only the beginning. The numerous interactions in the system might produce unpredictable complexity and chaos. Documentation reports should not be in the form of tick boxes but require opinions and ideas for improvement, preferably with face to face discussions. Diverse inputs into the database management system could lead to outcomes that can be both predictable and unpredictable. Indeed, apart from technical training (which is vitally important), it might be that control by managers with no database managerial training be examined in more depth.

From the data analysis and synthesis using systems diagramming, the CODEX was designed. There could be two possible routes of investigation that could result in improvement of the management of database systems: continuing to evolve the understanding of complex components; or automatically predicting the components that are connected from an agile best practice.

From the data collected and analysed earlier, the connections between the components, also the codes in this research, that lead to complexity are presented in a graph visualization network diagram (Figure 7.1).



**Figure 7.1 Graph of component complexity**

The representation of the data in this manner introduces possible future methods. Moreno (1953), father of sociometry, explained the use of these graph networks as a “method of exploration” and Prell (2012, p.83) argued this type of visualization of directed graph (digraph) provided an initial way of describing the network of interconnections. This could be used to define and predict future complexity.

### **7.6.1 Prediction using Machine Learning**

A possible way to predict which components will be affected, when a change to the inputs is made, is the use of a machine learning algorithm. A machine learning model could be built and retrained as technological advancements are made and new outcomes could be revealed. Machine learning has a huge potential to be able to automate this complex world of database systems management.

When a new technical component, business change or environment change occurs there is currently no pattern to follow. There could be a method established whereby all connections are continually recorded. Utilising machine learning algorithms introduces the possibility for a programmatic process to make better output predictions.

*“Machine learning can be described as computing systems that improve with experience. It can also be described as a method of turning data into software”* (Barnes 2015, p.13)

Machine learning offers many advantages and Shoham (2015, p.49) argues that this method can be applied in a broader AI (artificial intelligence) approach. Applying an AI to database management could bring about significant improvement.

## **7.6.2 Networks through Graph Theory**

An alternative method or combined method could be the use of digraphs and graph theory (Prell 2012). These can provide visual representation of the ubiquitous complexity. Digraphs can be used for pairing data and ties to show relationships between the objects. There is some overlap with systems diagrams.

The mathematician Euler began the study of graphs with his study of the ‘Bridges of Königsberg’ in the 18<sup>th</sup> century. Graph theory can look at complete networks where relationships may or may not be reciprocal (symmetric or asymmetric) using matrixes to record this. Graphs are used in social media to find friend recommendations, by Amazon to find patterns of purchases, and by haulage companies to calculate economical fuel routes. Different kinds of graphs are explored to investigate complex data patterns (Lenharth *et al.* 2016).



Graph theory algorithms can be useful when the data is sparse and the user needs to predict properties. This research problem to predict as accurately as possible what the list of complex components are for each change or new system setup, is a data science problem and needs to follow the data science process. Finding the best way forward could be through either a deterministic model or a probabilistic model. A deterministic model could describe exact outcomes from an experiment. In deterministic models every event has a cause. A probabilistic model could give a distribution of outcomes and the likeliness of an outcome occurring. The probabilistic models do not have all the information for a specific event. Probabilistic graphical models can be used with machine learning.

### **7.6.3 Interdisciplinary Data Science Complexity Prediction**

This interdisciplinary research highlights the key areas of interaction between business, technology and people. This research has shown that themes within each of these areas added to the complexity. A range of outcomes can result from predictability to chaos. This emergent phenomena is possible due to the many interacting aspects of databases. Indeed the remarks of Gonnering (2011, p.100) arguing that adaptation may be necessary to best practice, have been shown by this research to be true. The research has shown that the complexity of the systems affected the management of database systems. There have been a number of studies (McCririck & Goldstein 1980; Gillenson 1982; Gillenson 1985; Gillenson 1991; Mckendrick 2013; Abadi *et al.* 2016; Agrawal *et al.* 2009; Abiteboul *et al.* 2005; Bernstein *et al.* 1989; Bernstein *et al.* 1998; Silberschatz *et al.* 1995) (Abadi *et al.* 2016; Agrawal *et al.* 2009; Abiteboul *et al.* 2005; Bernstein *et al.* 1989; Bernstein *et al.* 1998; Silberschatz *et al.* 1995) which discussed database management problems and research areas to improve the management of database systems. This research has shown there is a need to create a new way of visualizing the

complexity of managing database systems, with an alternate blueprint for improvement.

As a data science problem the automation of the CODEX to predict or state complex components requires experimentation to learn from the situation and depends precisely upon the type of data input into the model.

The CODEX is not just a blueprint to help improve the dynamic system. It also could be considered as a way of documenting the changing components. The development of the CODEX blueprint needs further work to determine how the inputs are connected, how the new inputs are added and how the CODEX blueprint learns and adapts to find new influences.

Stonebraker (2016, p.79) recommends putting ideas into industry to make a difference to the DBMS. As such industry should be involved in providing inputs into the CODEX for further research.

## 7.7 Conclusions

The story of databases is one of rapid expansion. The speed of database change has meant that database management challenges constantly need to be adapted. These changes affect all aspects of the database from the type of data, to the internal structure of the database itself, to the way the data is stored. Today progress is so fast that the people employed in the operation of these systems have the immensely difficult task of keeping up to date. In addition to this, the complex interactions that have been shown to exist, mean that one single error by one person (perhaps due to lack of understanding), or the introduction of a new version of software that does not work with the existing software, or requirements not fully explained at the start, can all lead to a chaotic situation. Thus staff training and

communication becomes more and more important. A key finding of the research was that this is a people problem rather than a technical one.

Practices and procedures need to be regularly evaluated and modified in order to achieve continual improvement, as the digital world is never static and new developments need to be considered. The managers need to be experts at bringing together the combined knowledge of the teams of people now needed to deal with these large databases.

The importance of databases today is a major factor that underpins everything in modern society. Successful control of the vast amount of data is a challenge which will offer great rewards to all who manage them well. Innovation is no longer down to the brilliance of one person, it can only be achieved by the bringing together of individual ideas for improvement, and managers will need considerable expertise to achieve the best results. Communications and discussion with other managers in other organizations across the computing and database world would be essential if an organization is to keep pace with the most successful businesses.

There is a plurality of sources of materials, technology, skills, more teams of people, more locations and a worldwide market for sales of products. The more the division of labour is applied the more interconnections there are in a system. Hence, the more complex a system becomes and many businesses may find that what were once straightforward management techniques now need careful control to avoid the system becoming chaotic and unpredictable, resulting in financial loss. This plurality and the related interconnectedness is a main finding of the research. A major conclusion is that the issues relate more to people than to technology, i.e. socio-technical aspects are of key importance.

The CODEX blueprint, a major output from this research, is a way of utilising technology to handle rapid changes which affect, prevent or allow new adoption of best practices in an agile way. Allowing global collection of variants, mapping complexity and predicting the components that are affected by change when managing database systems, should enable effective and efficient management. Following this investigation into complexity in the operations of databases, the suggested CODEX system for management may well, with some adaptation, have further applications to business in general.

## Appendix A: Quantitative Questions

This is a full list of all questions asked in the quantitative survey.

Question number	Question
1	Which option best summarises your current job role? Please select one only.
2	In which country do you work?
3	What industry sector do you work in? Please select one option.
4	What is the approximate size of your organization's total workforce? Please select one.
5	What individual database roles does your organization have? Please select all that apply.
6	How long have you worked in the database field? Please select one.
7	Do you have any vendor professional certifications?
8	What is the approximate size of your largest database (overall size e.g. including data & logs)? Please select one.
9	What is the approximate number of database servers in your organization? Please select one.
10	How many people administer the databases in your organization? Please select one.
	<b>Server Demographics</b>
11	What database applications do you use? Please select all that apply.
12	What type of database engine do you use? Please select all that apply.
13	What percentage of your database servers use commercial software? (e.g. Oracle, SQL Server) Please select one.
14	What percentage of your database servers use open source software (e.g. MySQL)? Please select one.
15	What percentage of your database servers use on-premises database software (run on computers on the premises, in the building) or outsourced database hosting? Please select one.
16	What percentage of your databases use cloud database services (databases which are accessible via public, private or hybrid cloud instantly, on-demand, e.g. SQL Azure)? Please select one.

Appendix A: Quantitative Questions

Question number	Question
17	What percentage of your time is spent managing database servers? Please select one.
	<b>Database Architecture, Design and Development</b>
18	Do you use any of these architecture frameworks for database design? Please select all that apply.
19	Does your organization use the following processes at the architecture stage? For each question below please select the most relevant response.
20	Are any processes followed at the design stage? For each of the processes listed below, please select the most relevant response.
21	What development methodologies are followed? Please select all that apply.
22	Are any processes followed at the development stage? Please select the most relevant response for each row.
	<b>Database Technical Practices</b>
23	What database platforms are used? Please select all that apply.
24	What percentage of database servers in your organization are virtualized? Please select one option.
25	What separate database environments do you have for supporting database applications? Please select all that apply.
26	How do you install and configure your database server? Please select all that apply.
27	What are your current practices and procedures for database management? For each question, please select the most relevant answer.
28	How are the majority of database servers managed? Please select all that apply.
29	What service availability is required for your databases servers? Please select all that apply.
30	Do you need to provide 24x7 support for your databases? Please select one.
31	What are your current practices and procedures to maintain availability of your database servers? Please select the most relevant column.
	<b>Data and Database Security</b>
32	Are the following security policies enforced? Please select the most relevant column.
33	Do you have database audit policies to gather information about actions within the database? Please select one.

Question number	Question
34	Do you have a database software patching policy (e.g. for service packs, security updates, hotfixes, critical patches)? Please select one.
35	Do you have a standard process to follow for security breaches? Please select one.
36	Do you have policies or procedures in place to transfer data between servers? Please select all that apply.
37	Do you use any database engine encryption methods? Please select all that apply.
	<b>Change Management</b>
38	What are your current practices and procedures to manage change in your database servers? Please select the most relevant column.
39	What is the approximate number of database changes that you carry out in a week? Please select one.
40	Do you think some changes are carried out 'under the radar' i.e. by not following policies and procedures? Please select one.
	<b>Data Management</b>
41	What types of data are currently managed? Please select all that apply.
42	What are your current practices and procedures across your systems for data management? Please select the most relevant column.
43	What data warehouse method do you use? Please select all that apply.
44	Do you have a process in place to categorise types of data? Please select all that apply.
45	Are data requirements driving database management procedures in your organization? Please select one.
46	Approximately how much of your data is unstructured? Please select one.
	<b>Frameworks</b>
47	What IT service management frameworks do you use, if any? Please select all that apply.
48	What are your practices and procedures for data management? Please select the most relevant columns.
49	Do you use a problem management method? Please select all that apply.
50	When a problem is found what usually happens to resolve the issue? Please select one.

Appendix A: Quantitative Questions

Question number	Question
51	Is your database management based on an agile database technique? Please select one.
	<b>Storage</b>
52	What storage types do you use? Please select all that apply.
53	Do you optimise your storage layout for database use? Please select one.
54	Do you have an Information Lifecycle Management (ILM) policy for different data storage models for different data access? Please select one.
55	Do you use solid-state drive (SSD) / Flash as a database storage tier? Please select one.
56	Whose practice is followed for database storage configuration? Please select all that apply.
	<b>Cloud</b>
57	Do you use any of the following forms of cloud database options? Please select all that apply.
58	What (if anything) do you use cloud database services for? Please select all that apply.
59	Do you have practices and procedures in place in the following areas to manage cloud databases? Please select the most relevant columns.
	<b>Organizational Culture 1</b>
60	How do you receive database training? Please select all that apply.
61	How often do you have the opportunity to undertake formal training courses? Please select one.
62	Are you a member of any database community associations? Please select all that apply.
63	How often do you have the opportunity to attend database conferences, workshops or seminars? Please select one.
64	How frequently does the internal IT management structure change in your organization? Please select one.
65	How frequently does the database team structure change? Please select one.
	<b>Organizational Culture 2</b>
66	What are your working team practices? For each question below, please select the most relevant response.



Question number	Question
67	What are your communication and business practices? For each question below, please select the most relevant response.
68	What are your database product practices? For each question below, please select the most relevant response.
69	Is database management visible to the following people or teams? Please select all that apply.
70	Who controls database choices? For each question below, please select all that apply.
	<b>Application Centric</b>
71	In relation to the main application that you are involved with, what are your practices and procedures from a database perspective? Please select the most relevant columns.
72	Do you have different database management practices for different database products e.g. SQL Server, Oracle, MySQL, CouchDB etc.? Please select one.
73	What are your practices and procedures in relations to Big Data (A general term used to describe the large volume of unstructured and semi-structured data that cannot be processed using conventional methods)? Please select the most relevant columns.
	<b>Best Practice</b>
74	What do you think is meant by 'best practice' in database management? Please select all that apply and please expand with other options.
75	Where do you personally find database best practice guidelines to follow? Please select the most relevant columns.
76	Does your organization follow any database best practice guidelines? Please select one response.
77	What issues can occur with following best practice? Please select the most relevant columns.
78	Do you think Database Management and Data Management are separate fields? Please select one response.
79	Please select to what extent best practice is currently controlled within your company in each of the following areas? Please select the most relevant column.
80	Which of the following could improve practices and procedures for database management in your organization? Please select all that apply.
	<b>Thank You</b>

Appendix A: Quantitative Questions

Question number	Question
81	What do you think your business vision of database management will look like in 10 years?
82	If you are using the cloud for database services, what is the reason behind this? E.g. Database as a Service (DBaaS) for Self Service database functionality or Self-managed database servers running on Infrastructure as a Service (IaaS) (optional question)
83	If you would be prepared to participate in further research on database management, please provide your email address. The follow up may include a short online free-text-answer questionnaire and/or an online interview. Many thanks

**Multi Part Questions**

Question Numbers	Multi Part Questions
19	Do you have a set process for requirements gathering?
19	Do you create a High Level Design (logical design)?
19	Do you create a Low Level Design (physical design)?
19	Is the database solution documented?
22	Do you follow a defined database development cycle?
22	Do you have a standard database coding practices?
22	Do you have a standard database testing process?
22	Do you use a source control system for storing database development code?
27	Do you have a policy for backup and recovery? No
27	Do you have set Recovery Point Objectives (RPO) for your databases? (The amount of data loss allowable)
27	Do you have set Recovery Time Objectives (RTO) for your databases? (The time it takes to restore the data)
27	Do you carry out day-to-day database maintenance?
27	Do you have processes for monitoring regular maintenance?
27	Do you have a process for managing database server performance?

Question Numbers	Multi Part Questions
27	Do you manage database servers using automated procedures to issue alerts?
27	Do you record a performance baseline?
31	Do you have a documented disaster recovery process?
31	Is your disaster recovery process tested regularly?
31	Does the data reside in multiple geographical locations?
31	Do you have a process for managing data in multiple geographical locations?
31	Are there any processes in place to enable capacity management (CPU, RAM, Disk Space)?
31	Do you have a release management process that encompasses planning, design, build, configuration and testing for hardware and software releases?
31	Do you use a configuration management database (CMDB) as a repository for storing database configuration items?
38	Do you have a change management process to control, manage and implement changes to the live IT infrastructure or IT service?
38	Is there a procedure to mitigate the risk of loss of a database during a change?
38	Do database changes require sign off by business users?
38	Can anyone carry out changes to the database server?
38	Is a change procedure enforced for all database engines?
42	Do you have a master data management (MDM) policy?
42	Do you have any data governance policies?
42	Do you have practices or procedures in place for reporting on data in the databases?
42	Do you have procedures to follow to keep data for legal reasons?
42	Do you have data quality practices or procedures?
42	Is any data quality work undertaken within your databases?
42	Is there a policy in place to keep historical data for a specific number of years?
42	Is there a management policy for archiving data indefinitely for long term preservation?

Appendix A: Quantitative Questions

Question Numbers	Multi Part Questions
42	Do you use/ carry out any data analytical processes?
42	Do you carry out trend and pattern analysis or data mining?
42	Is any crowdsourcing (distributed problem-solving) data used for predictive analysis in your organization?
42	Do you have any processes in place for predictive analysis?
48	Do you follow data lifecycle management policies?
48	Do you have your own data management practices and procedures?
48	Do you use the data management association framework (DAMA-BOK)?
48	Do you use MIKE2.0, the open source standard for information management?
66	Are customer requirements clearly defined at the outset?
66	Do customer requirements change in the midst of projects?
66	Are frequent malfunctions dealt with in a re-active role (e.g. fire fighting)?
66	Do you put long term fixes in place for regularly occurring issues to future proof the database applications?
66	Does the company foster an environment to encourage certification?
66	Are database management decisions based solely on customer requirements?
67	Is communication good between management and database team members?
67	Is within-team communication good?
67	Is cross-team communication good (e.g. between DBA's and service team, development team, storage team etc.)?
67	Do all the stakeholders communicate (cross boundary communication e.g. customers, database software vendors etc.)?
67	Does the business have a clear strategy for the database management team, e.g. virtualizing databases, moving databases to the cloud, database consolidation?
67	Do policy changes to the business model (e.g. disaster recovery plans, volume of disk storage allowed) create risks for database management?
68	Is the database software product selection constrained due to your employee skill set in house?

Question Numbers	Multi Part Questions
68	Is the database software version selected for financial reasons? (e.g. Standard or Enterprise Edition)
68	Does your budget determine what database platform is used? (e.g. SQL Server, Oracle, MySQL)
70	Who makes the decision to upgrade to new versions of database software? E.g. SQL Server 2005 to SQL Server 2012
70	Who has the choice to buy supplementary tools for managing the databases from vendors?
70	Who influences the choice of primary database software product?
70	Who has control of database management?
70	Who controls the introduction of new on-premises database software products into the organization?
70	Who controls what cloud database software the organization uses?
71	Do the included database software features govern the type of management that can be provided?
71	Is database application scalability a requirement?
71	Do you have a procedure to manage scalability?
71	Do you have an improvement method to follow for database management?
71	Do you have a procedure to select different database engines for the task required?
71	Do you have a procedure to review new database engine changes?
71	Is the database management abstracted from the hardware layer?
71	Do you have a procedure in place for managing virtualized databases?
73	Do you manage Big Data at the moment?
73	Do you have procedures for managing Big Data?
73	Do you have different management practices for different sizes of database?
73	Do you manage more unstructured data than this time last year?
73	Is the database management abstracted from data management?
73	Does the organization have any processes in place to state which publically available data sets are acceptable to use? (e.g. government data sets)

*Appendix A: Quantitative Questions*

<b>Question Numbers</b>	<b>Multi Part Questions</b>
77	Do you think it is important to have best practices?
77	Can following best practice be a labour intensive process?
77	Can following processes obstruct best practice?

## Appendix B: Qualitative Questions

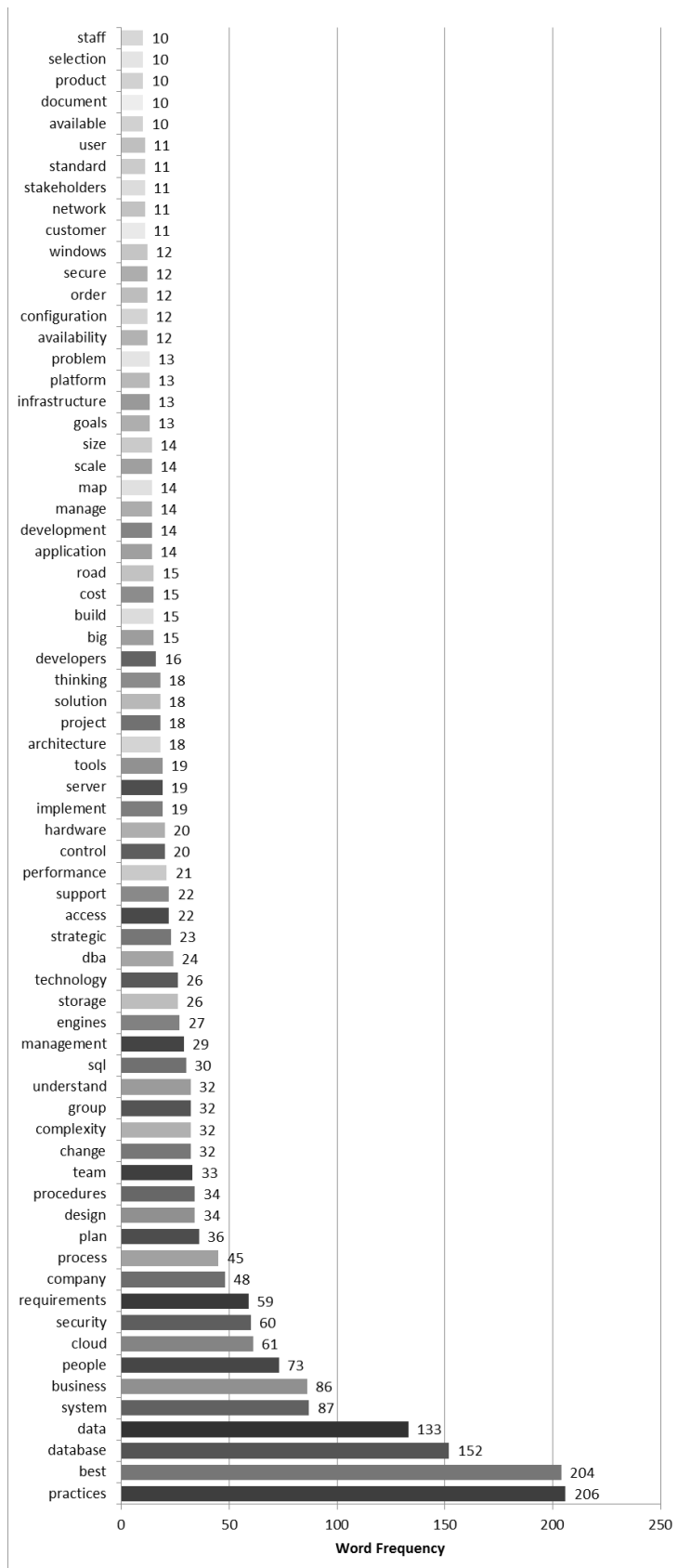
This is a full list of all questions asked in the qualitative survey.

Question number	Question
1	Do you think some best practices and procedures are more important than others for managing database systems? If so, what are the most important ones?
2	What best practices and procedures do you think should be considered when selecting different database engines?
3	What kind of requirements gathering and architectural design processes for the hardware, data and databases do you think are important? Why are these important?
4	In what ways do you think that best practices and procedures could assist management of the database lifecycle?
5	What complexities between technology layers, do you think, affect the operation of databases?
6	Describe any complexities that exist with the adoption of best practices and procedures when managing cloud databases?
7	Was there ever a time when you felt the complexity of database systems compromised your ability to implement best practices and procedures?
8	Who you think should create and control database best practices and procedures?
9	How, if at all, do cross boundary communications among stakeholders affect best practices and procedures?
10	What effect can a database management strategic plan have on best practices and procedures for the management of database systems?

## Appendix C: Word Frequency Count

The word frequency count is a basic list of words which provides the number of times they have been found in the data set. It provides some useful information on the prominence of words, however the results only provide partial meaningful indications. The figure below provides a list of the most frequently occurring words ignoring prepositions, postposition, conjunctive adverbs and linking works.

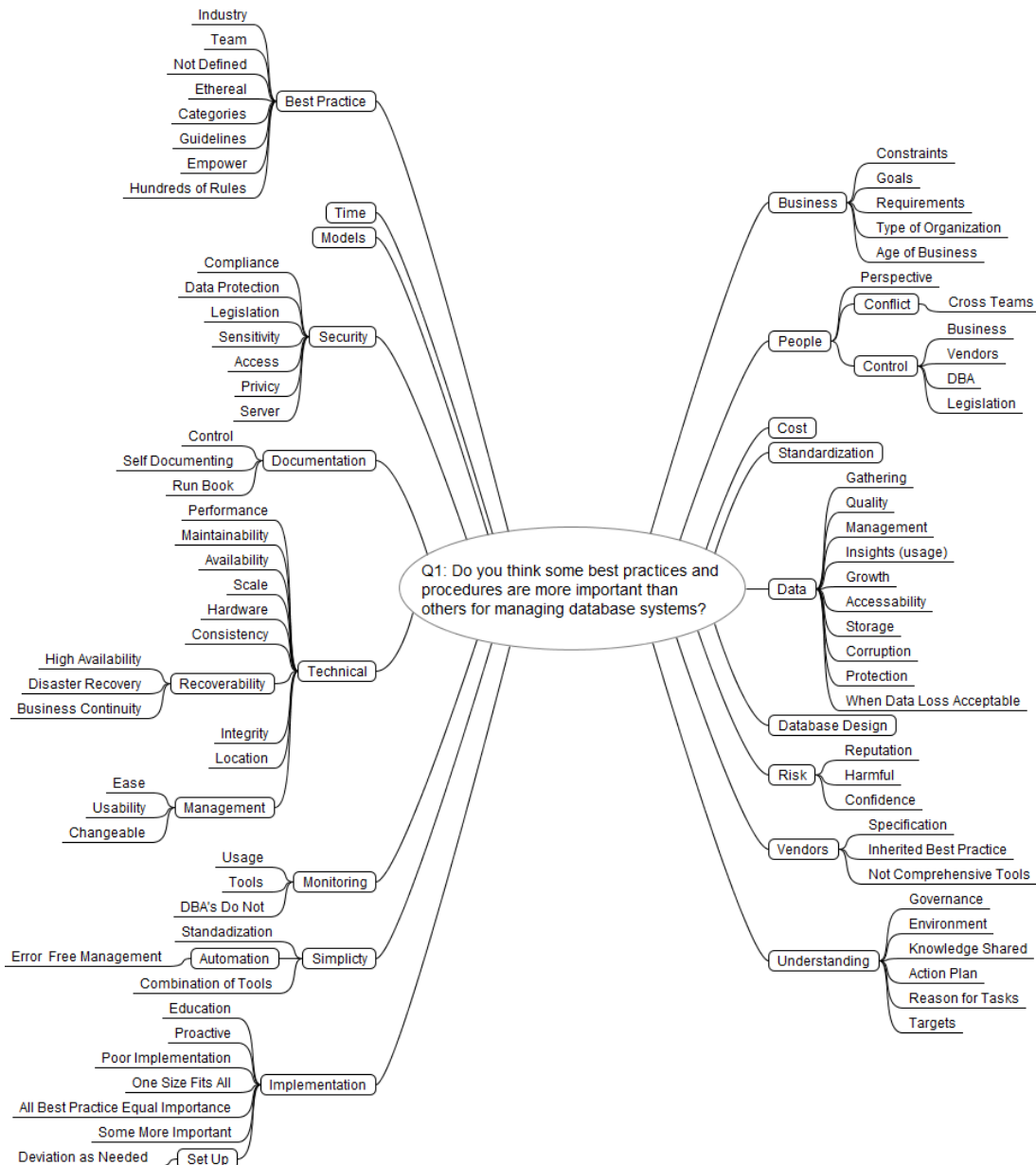




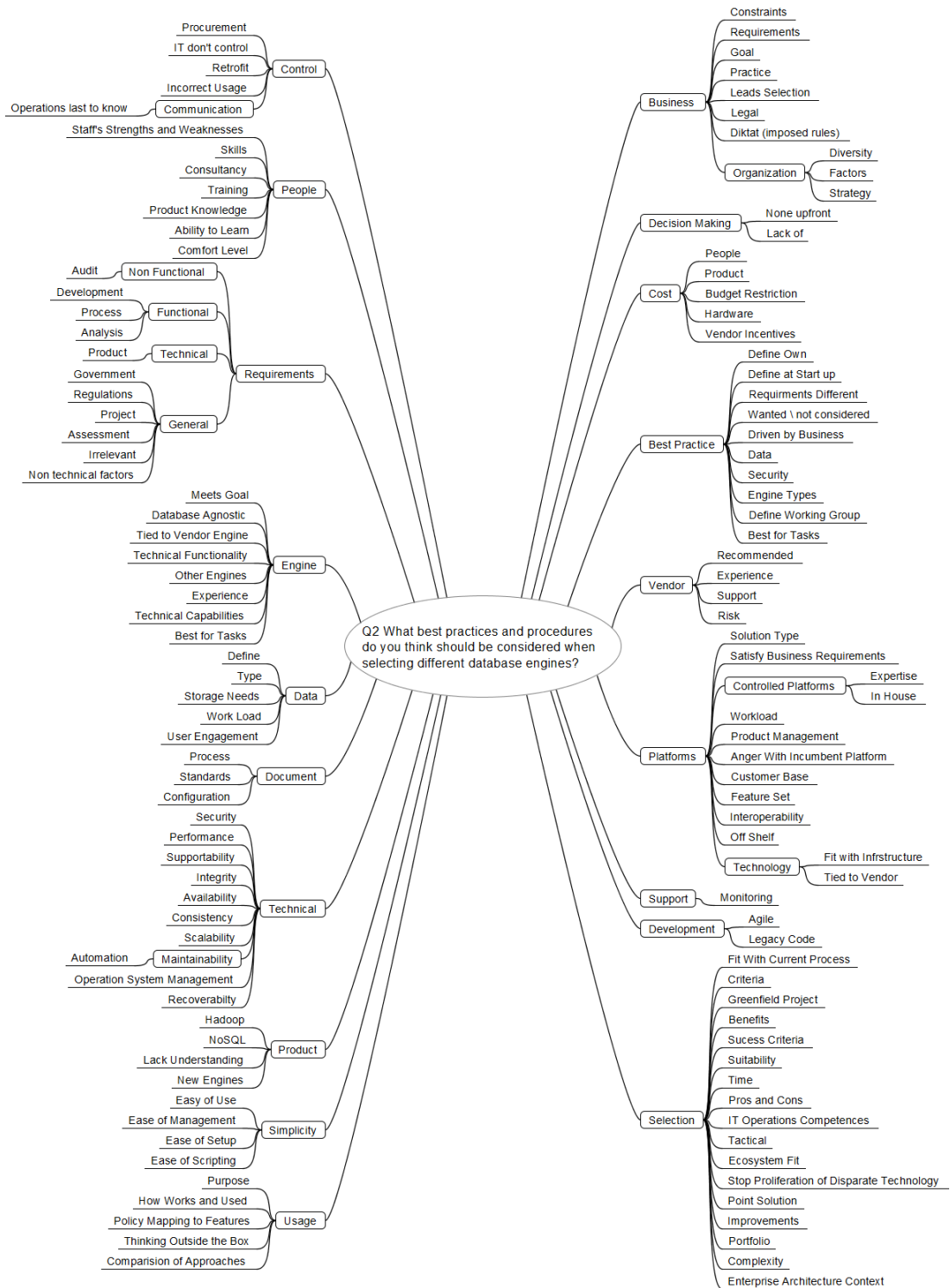
# Appendix D: Qualitative Question Spray Diagrams

The spray diagrams from the 10 questions are included below:

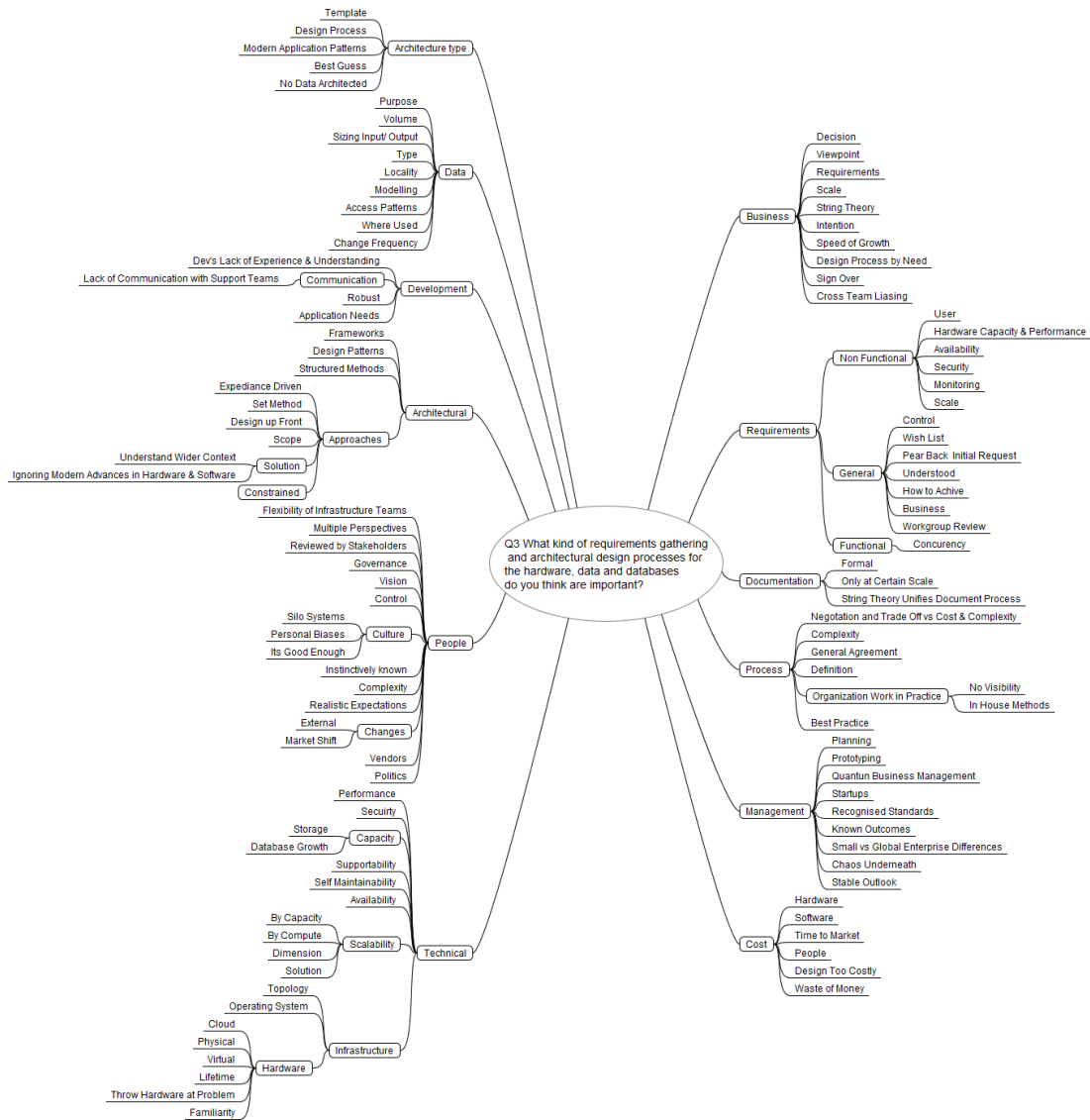
## Question 1



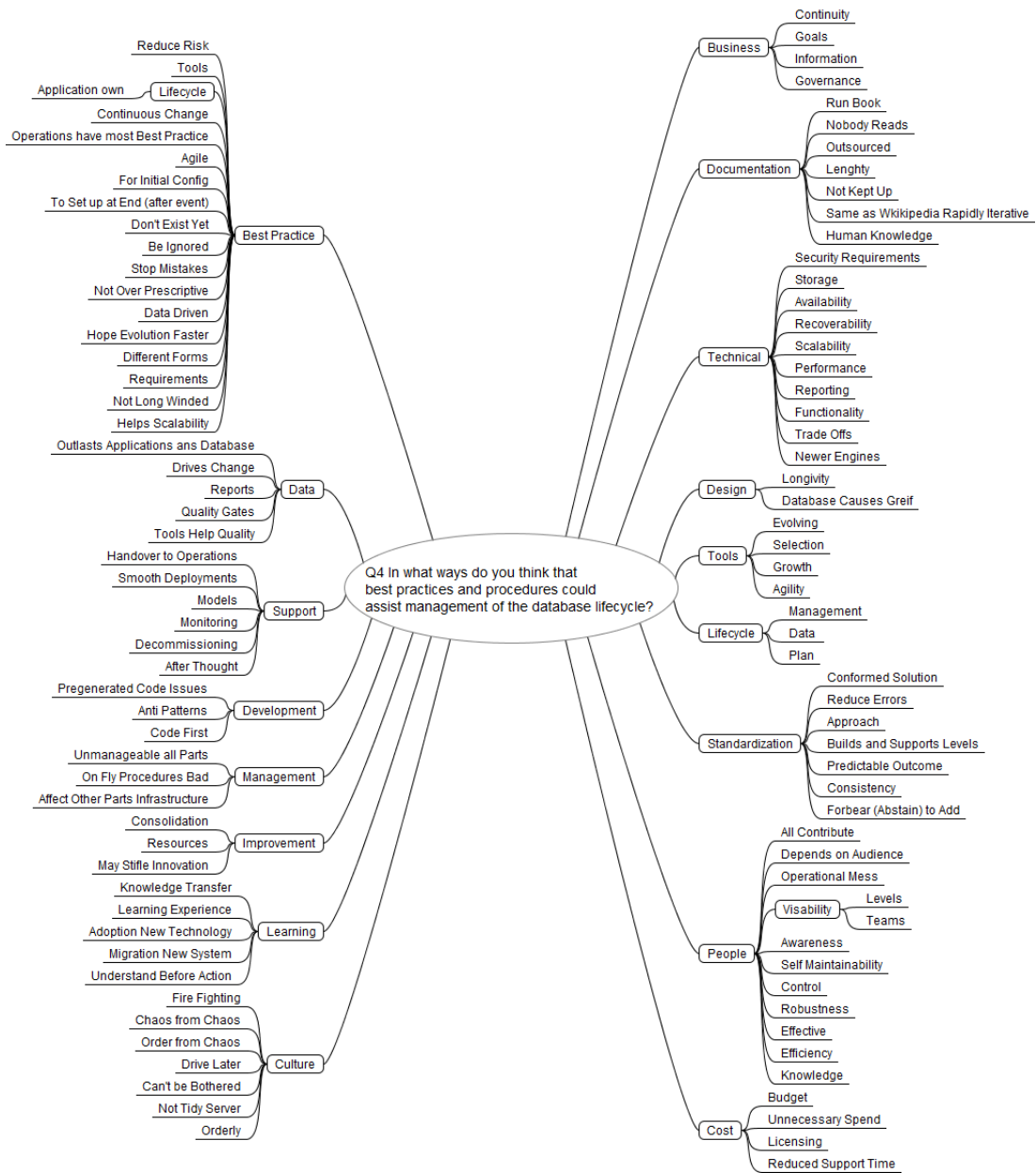
Question 2



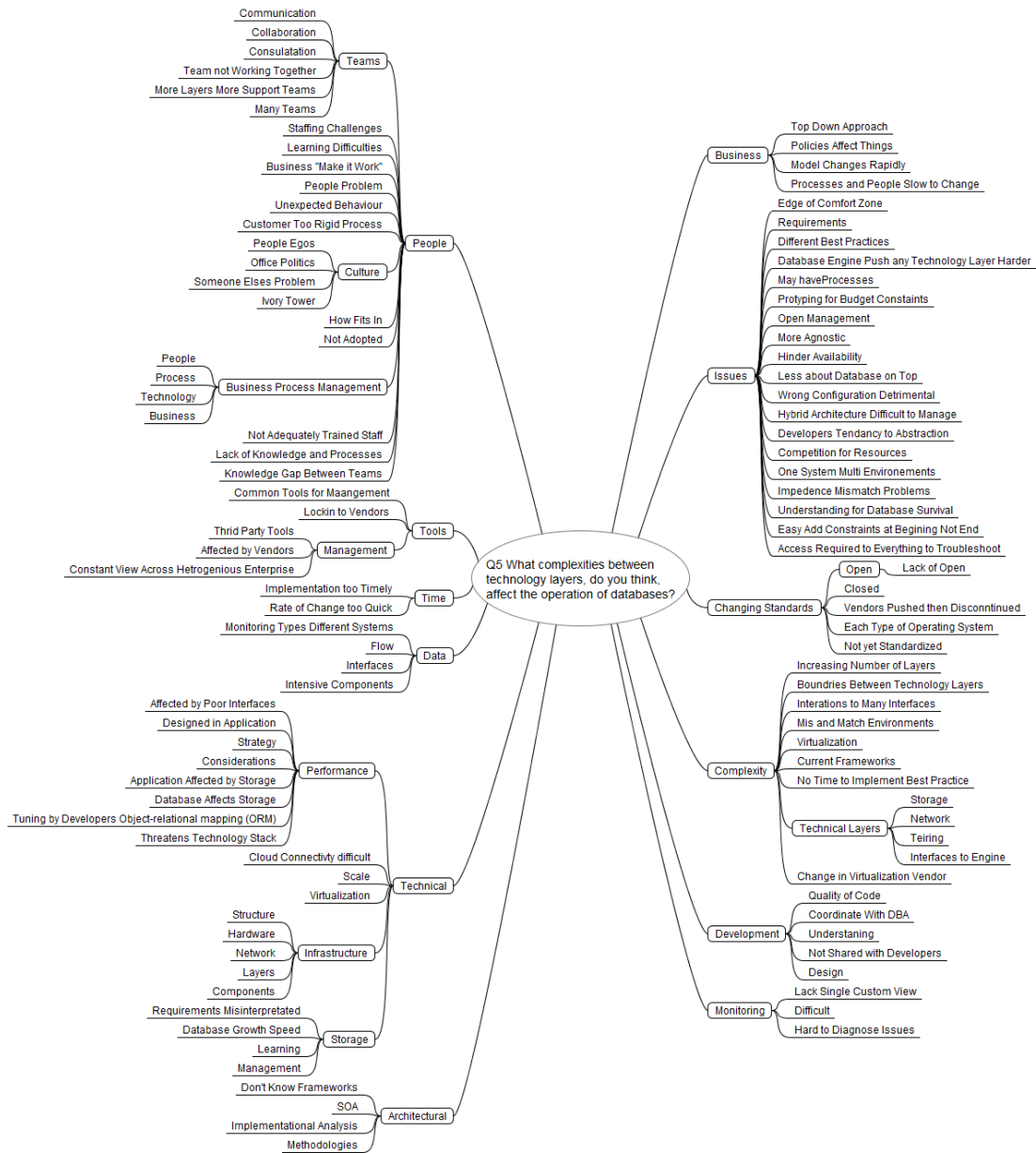
Question 3



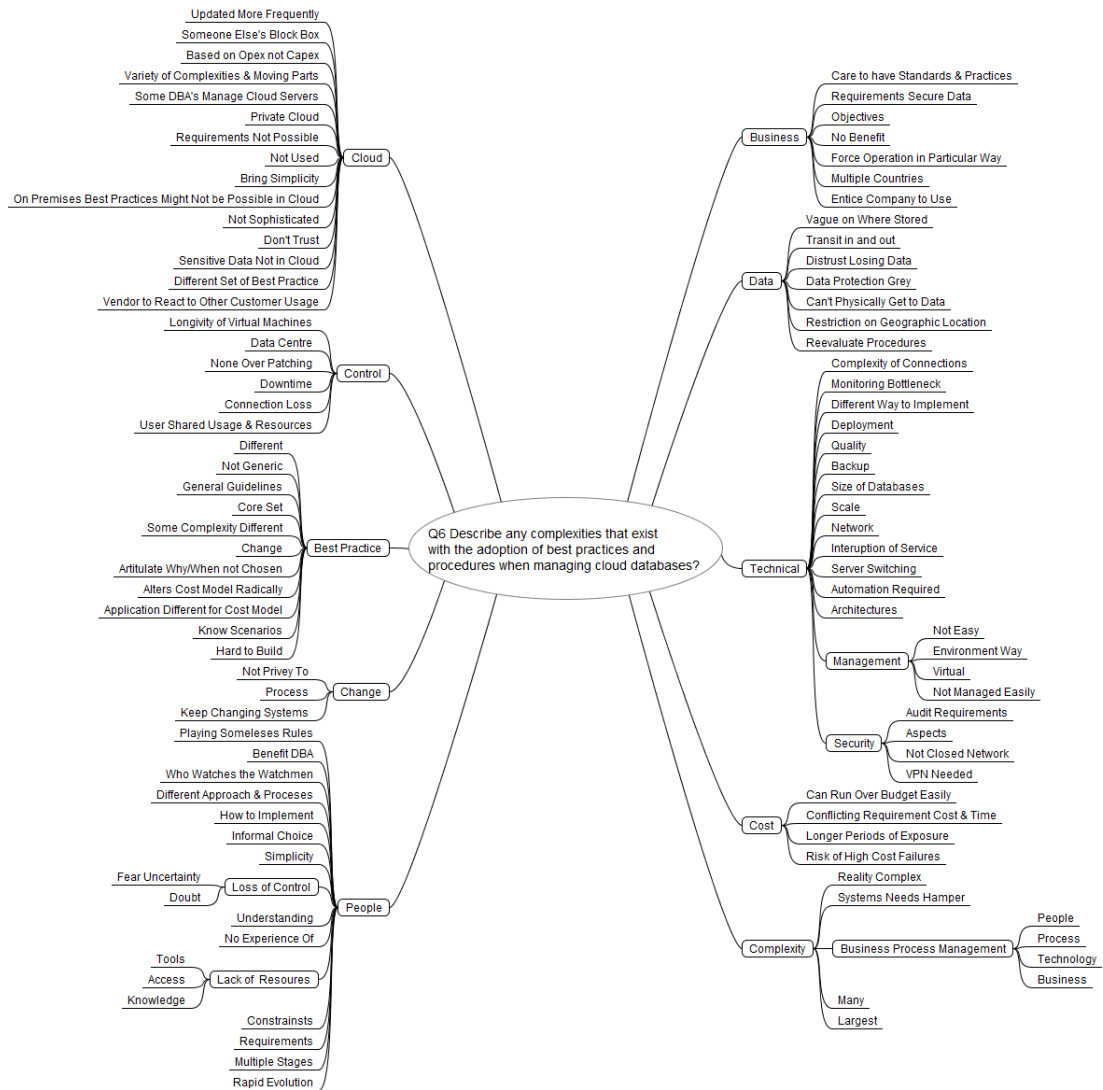
Question 4



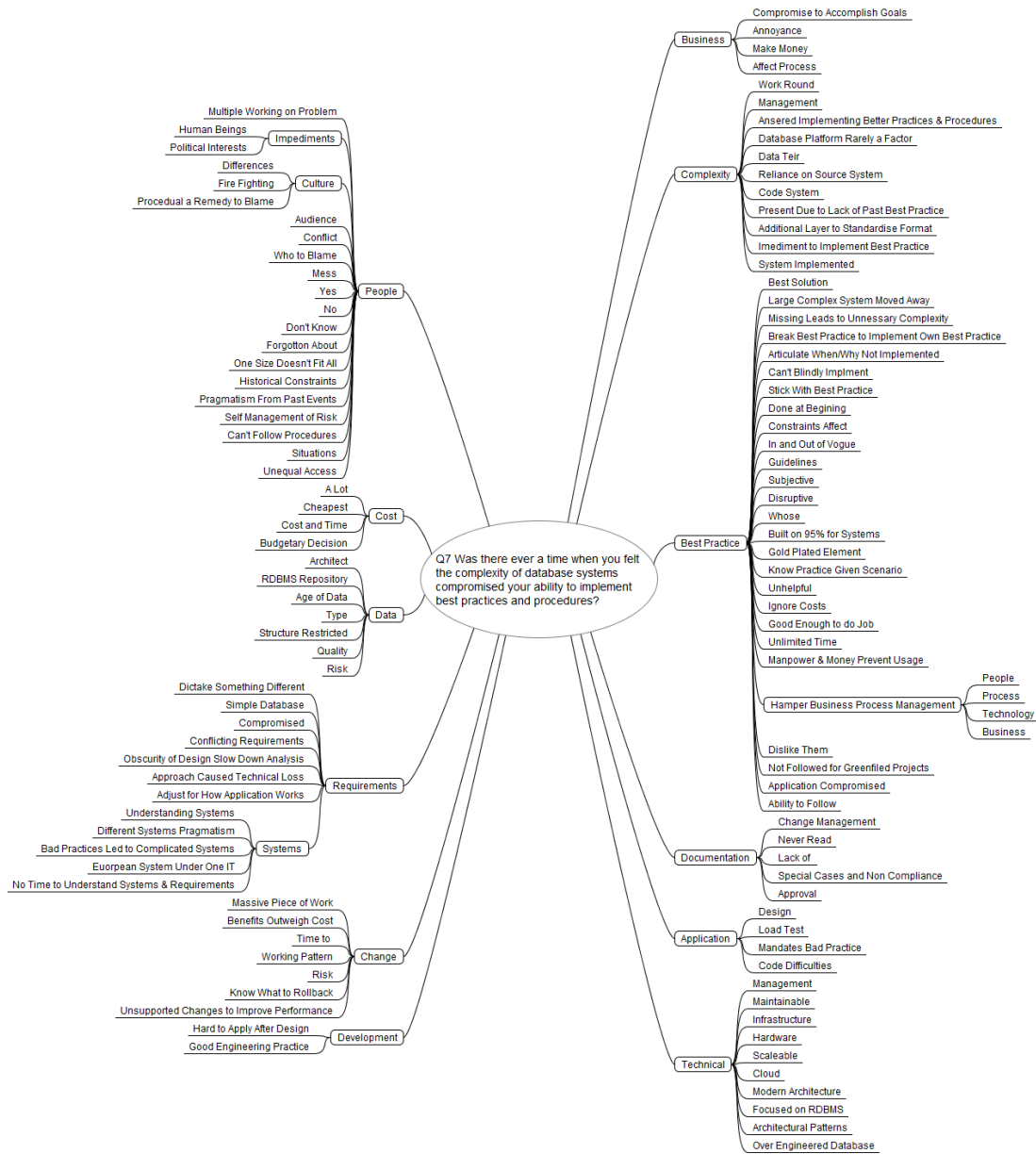
Question 5



Question 6

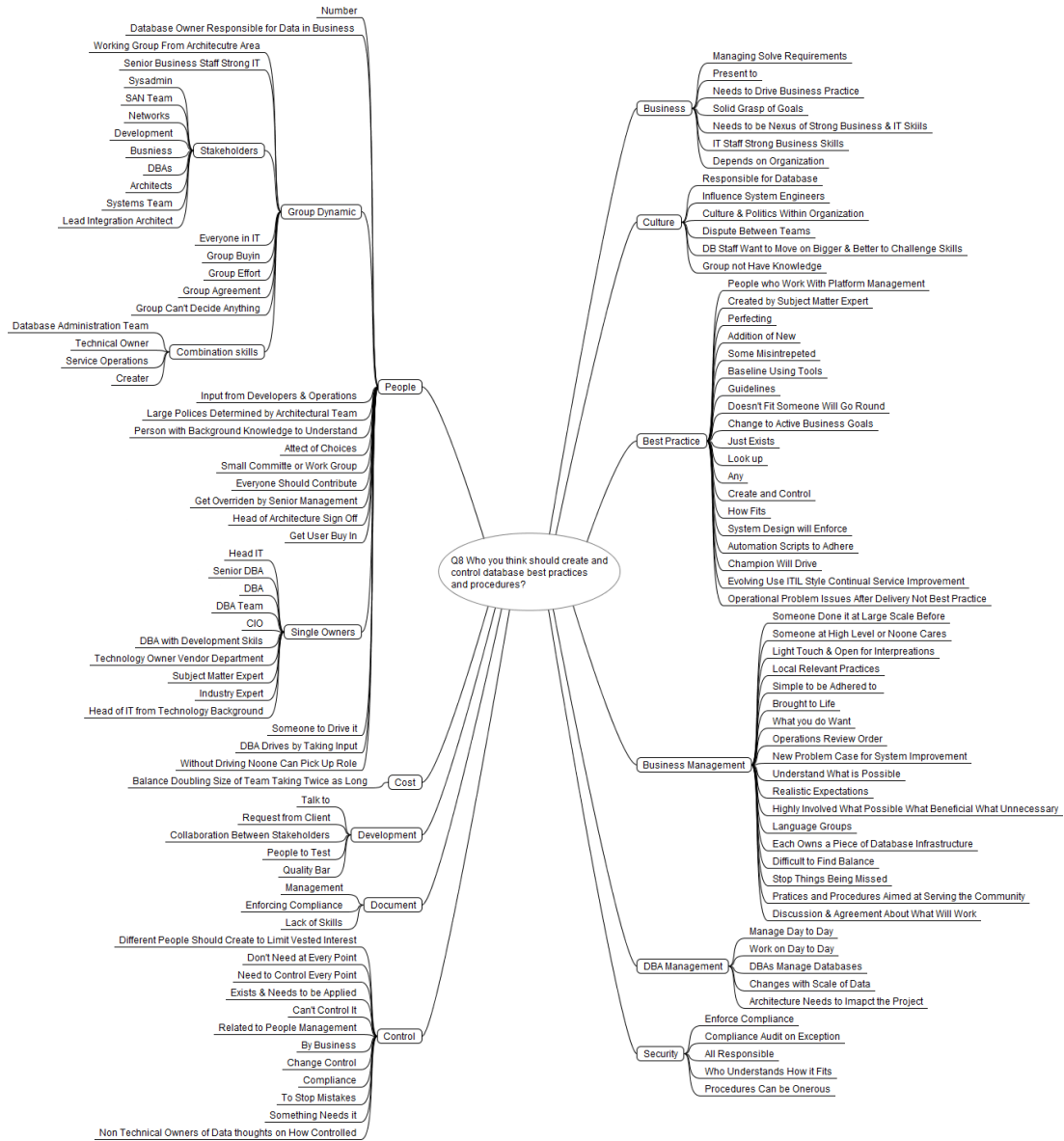


Question 7

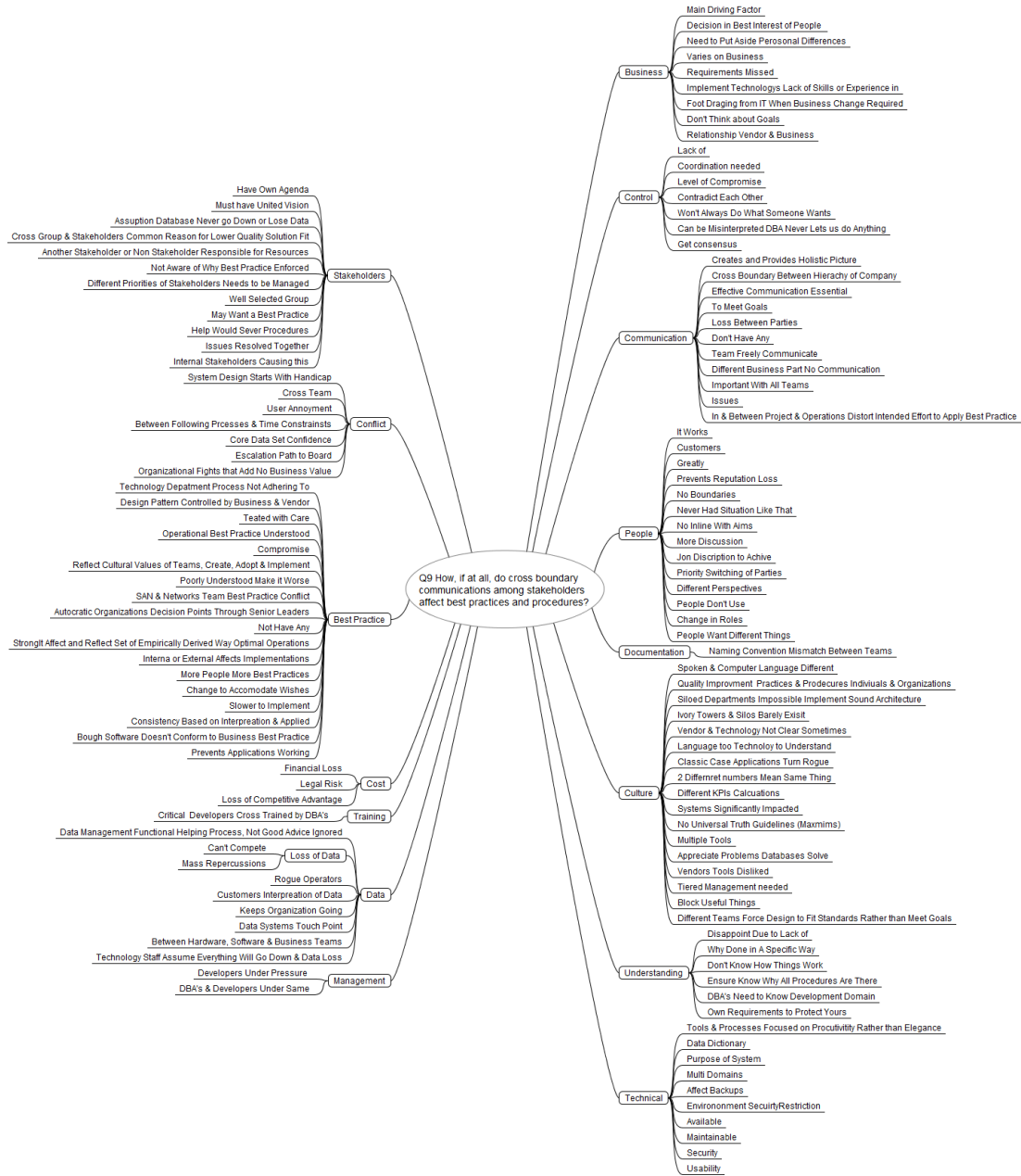




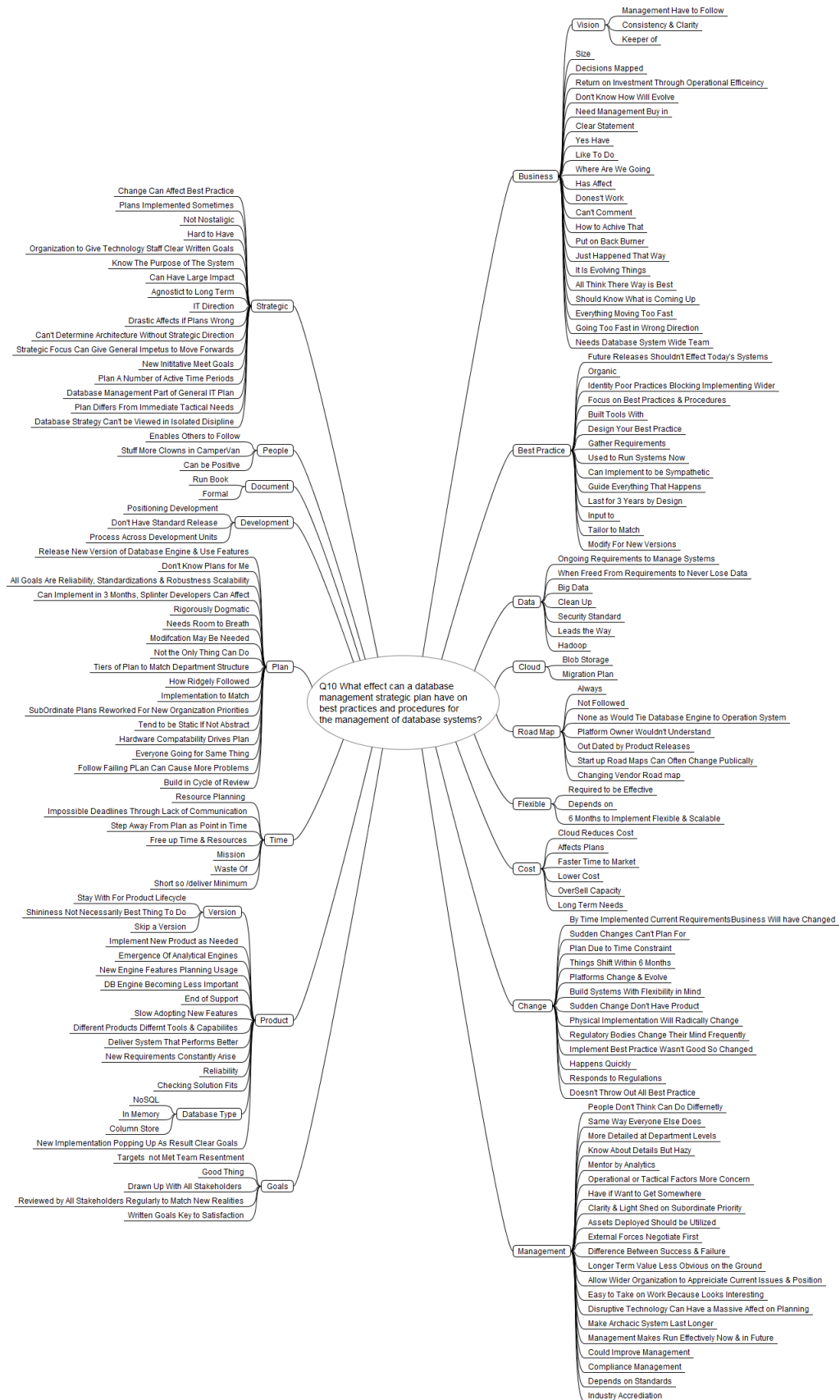
Question 8



Question 9



Question 10



## Appendix E: Code Book Code Summary

This is the key to the code book.

<b>What is your job area (JOB)</b>	<b>Code</b>
DBA	A
Developer	B
BI	C
Architect	D
Other (Accidental DBA)	E
DBA / Architect	F
BI /Architect	G
BI / Developer	H

<b>How long have you worked in the Database Field (TIME)</b>	<b>Code</b>
Less than 5 years	L
5 to 10 Years	I
More than 10 years	M

<b>What size of organization do you work in (SIZE)</b>	<b>Code</b>
Self Employed Consultant	S
SME	M
Large Enterprise	L

<b>In which Country do you Work (PLACE)</b>	<b>Code</b>
UK	U
Non UK	N

**Example**

For myself as transcriber and facilitator

PID	Job	Time	Size	Place
x.0	F	M	M	U

The initial set of codes found and labelled in the working database

Codes	id
Application	1
Approaches	2
Architectural	3
Architecture type	4
Automation	5
Best Practice	88
Business	92
Business Management	6
Business Process Management	60
Capacity	7
Changes	82
Changing Standards	45
Cloud	61
Combination Skills	8
Communication	76
Complexity	77
Conflict	62
Control	83
Controlled Platforms	10
Cost	90
Culture	85
Data	91
Database Design	11
Database Management	12
Database Type	13
Decision Making	14
Design	15
Development	86
Document	78
Documentation	84
Engine	16
Flexible	17
Functional Requirements	63
General Requirements	64
Goals	18
Group Dynamic	19
Hamper Business Process Management	9
Hardware	20
Impediments	21
Implementation	22
Improvement	23
Infrastructure	65
Issues	24
Lack of resources	25
Learning	26
Lifecycle	27
Loss of Control	28
Loss of data	29
Maintainability	30
Management	87
Model	31
Monitoring	66
Non Functional Requirements	67
Open Standards	32
Organization type	68
People	93
Performance	33
Plan	34
Platforms	35
Process	36
Product	69
Recoverability	37
Requirements	79
Risk	38
Road Map	39
Scalability	40
Security	80
Selection	41
Setup	42
Simplicity	70
Single Owners	43
Solution	44
Stakeholders	71
Standardization	72
Storage	46
Strategic	47
Support	73
Systems	48
Teams	49
Technical	89
Technical (Requirements)	50
Technical Layers	51
Technology	52
Time	81
Tools	74
Training	53
Understanding	75
Usage	54
Vendor	55
Version	56
Visibility	57
Vision	58
Work in Practice	59

# Appendix F: Data Map

Component A		Component B											
APPLICATION	DESIGN	DEVELOPMENT	DATA	CLOUD	ENGINE	SECURITY	TECHNICAL	PLATFORM	ARCHITECTURAL	SELECTOR	PRODUCT	TOOLS	REQUIREMENTS
DESIGN			1			1		1					
DEVELOPMENT						1		2					
DATA					1	1	2	5			1		1
CLOUD							1	1					
ENGINE							1	1					
SECURITY			4	1	1	1	2						
TECHNICAL	1	1											1
PLATFORM										1			1
ARCHITECTURAL					1			1					
SELECTOR						1							1
PRODUCT											1		1
TOOLS			1					1					
REQUIREMENTS				1		1	2	1	2	2			
PROCESS	1		2						1				1
SUPPORT			1			1							
IMPLEMENTATION													
DOCUMENTATION									1				
CHANGE			1										1
COST								1					1
GOALS								1					
PLAN								1					
VISION													
STRATEGIC													
BUSINESS	3		3	1	1	1	2	1	1	3	1		1
PEOPLE			4			1		2		1		2	7
STAKEHOLDERS			1										1
TEAMS			1					1					
VENDORS								1			1		
COMMUNICATION													
CONTROL			1	1			4			1			1
CULTURE									1				
GROUP DYNAMIC							1						
CONFLICT			3										1
TRAINING													
UNDERSTANDING													
LEARNING							4				1		2
STANDARDIZATION								1					
COMPLEXITY													
BEST PRACTICE	1		2	1	3	2	4	1					
MANAGEMENT			1							1			2
FLEXIBLE													
SIMPLICITY			1										
LIFECYCLE													

Component B		PROCESS	SUPPORT	IMPLEMENTATION	DOCUMENTATION	CHANGE	COST	GOALS	PLAN	VISION	STRATEGIC	BUSINESS	PEOPLE	STAKEHOLDERS	TEAMS	VENDORS	COMMUNICATION	CONTROL	CULTURE	GROUP DYNAMIC	
Component A	APPLICATION																				
	DESIGN	1					1	1	1	1			1	1							
	DEVELOPMENT		3																	1	
	DATA	2	1			1	1	1				3			1						
	CLOUD	1				2	2	1												1	
	ENGINE								1	1			1	1							
	SECURITY											1	1							1	
	TECHNICAL		4		1	1	1	2	1	1		1	4		2				1		
	PLATFORM						1	1				1									
	ARCHITECTURAL	1	3				1														
	SELECTOR																				
	PRODUCT			2																	
	TOOLS																				
	REQUIREMENTS	2	2		1	2		2							1						
	PROCESS							1				1								1	
	SUPPORT					1							1								1
	IMPLEMENTATION						1					1									
	DOCUMENTATION	1	1				1													1	
	CHANGE		1						1				1	2						2	
	COST											1	2								
	GOALS												1								
	PLAN												1								
	VISION							1													
	STRATEGIC				1	1	1	1	1	1											
	BUSINESS	2	1		1		3	4	2	2		2			1				2	1	1
	PEOPLE	2	5		2	3	1	1	3	1		1							5	4	2
	STAKEHOLDERS				2		1	1	1	1									1	1	1
	TEAMS											2							1	1	1
	VENDORS											1	1								
	COMMUNICATION		1				1	1	1						1	2					
	CONTROL		2					1					1								
	CULTURE											1	3		1				2		
	GROUP DYNAMIC								1											1	
	CONFLICT							1					1								
	TRAINING											1	1								
	UNDERSTANDING	1			1		1					1	4	2							
	LEARNING												1								
	STANDARDIZATION																				
	COMPLEXITY				1			2					1								
	BEST PRACTICE	1			1	2	2	2		1		3	2	1							2
	MANAGEMENT	3	2		2			1	1			1									
	FLEXIBLE							1													
	SIMPLICITY				1				1												
	LIFECYCLE							1													

Appendix F: Data Map

Component A	CONFLICT	TRAINING	UNDERSTANDING	LEARNING	STANDARDIZATION	COMPLEXITY	BEST PRACTICE	MANAGEMENT	FLEXIBLE	SIMPLICITY	LIFECYCLE
APPLICATION							3				
DESIGN						2	4	2			1
DEVELOPMENT											
DATA	1					4	1	2			
CLOUD			1			3	5	3			
ENGINE			1				2	1			1
SECURITY			1		1		4	1			
TECHNICAL		1	2	3	1	5	3	1			1
PLATFORM											
ARCHITECTURAL			1			2	1	2			
SELECTON							1	1			
PRODUCT						1					
TOOLS						1					
REQUIREMENTS	2						4	1			
PROCESS							2				
SUPPORT							1				
IMPLEMENTATION							2				
DOCUMENTATION							2				
CHANGE					1		5	2			
COST			1			2					1
GOALS	1										
PLAN								1			
VISION			1					1			
STRATEGIC							1				
BUSINESS	1	1	1	1	1	2	4				
PEOPLE	1	2	5	1		1	9	4			
STAKEHOLDERS							2	2			
TEAMS		1					3	1			
VENDORS			1		1		1	1			
COMMUNICATION			1				4				
CONTROL							5	2			
CULTURE							1				
GROUP DYNAMIC			2								
CONFLICT							1	1			
TRAINING			1								
UNDERSTANDING		1					3				
LEARNING											
STANDARDIZATION						2	1				
COMPLEXITY							3				
BEST PRACTICE	2		2	1	2	3	12	2			3
MANAGEMENT					1	1	7				1
FLEXIBLE											
SIMPLICITY							2	1			
LIFECYCLE						1		1			



## References

- Abadi, D. et al., 2016. The Beckman Report on Database Research. *Communications of the ACM*, 59(2).
- Abadi, D. et al., 2014. The Beckman Report on Database Research. *ACM SIGMOD Record*, 43(3), pp.61–70.
- Abadi, D.J., 2012. Consistency Tradeoffs in Modern Distributed Database System Design. *IEEE Computer Society*, 2(February), pp.37–42.
- Abiteboul, S. et al., 2005. The Lowell Database Research Self-Assessment. *Communications of the ACM*, 48(5), pp.111–118.
- Ackoff, R.L., 1999. *Ackoff's Best: His Classic Writings on Management*, New York: John Wiley & Sons.
- Ackoff, R.L., 1981a. *Creating the Corporate Future*, New York: John Wiley & Sons.
- Ackoff, R.L., 1981b. The Art and Science of Mess Management. *Interfaces*, 11(1), pp.20–26. Available at: <http://www.jstor.org/stable/25060027>.
- Agrawal et al., 2009. The Claremont Report on Database Research. *Communications of the ACM*, 52(6), pp.56–65.
- Agrawal, D. et al., 2009. Database Management as a Service: Challenges and Opportunities. *2009 IEEE 25th International Conference on Data Engineering*, pp.1709–1716. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4812596> [Accessed November 18, 2010].

## References

- Aiken, P. et al., 2011. Data Management and Data Administration: Assessing 25 Years of Practice. *Journal of Database Management*, 22(3), pp.24–45.
- Aiken, P., 2016. Experience : Succeeding at Data Management — BigCo Attempts to Leverage Data. *ACM Journal of Data and Information Quality*, 7(1–2), pp.1–35.
- Aiken, P. et al., 2007. Measuring Data Management Practice Maturity : A Community's Self-Assessment. *IEEE Computer Society*, 40(4), pp.42–50.  
Available at: [http://www.irmac.ca/0708/Measuring Data Management Practice Maturity.pdf](http://www.irmac.ca/0708/Measuring%20Data%20Management%20Practice%20Maturity.pdf).
- Ailamaki, A., Kantere, V. & Dash, D., 2010. Managing Scientific Data. *Communications of the ACM*, 53(6), p.68.
- Alzain, M.A. & Pardede, E., 2011. Using Multi Shares for Ensuring Privacy in Database-as-a-Service. *Architecture*, pp.1–9.
- Ambler, S.W., 2003. *Agile Database techniques: Effective strategies for the agile software developer*, Indianapolis: Wiley Publishing.
- Ambler, S.W., 2007. Data Architecture and Beyond : Strategies for Improving Your Data Ecosphere Access to the Experts. *October*, 10(10).
- American Productivity and Quality Centre, 1999. What is benchmarking. Available at: [www.Apqc.org](http://www.Apqc.org).
- Anderson, R. et al., 2009. *Database State*,
- Anon, 2010. Data, Data everywhere. *The Economist*, 394(8671), pp.3–5.
- Anon, 2014. Oxford Dictionaries. *Oxford University Press*. Available at:

<http://www.oxforddictionaries.com/>.

Argyris, C. & Schön, D., 1978. *Organizational Learning*, Reading: Addison-Wesley.

Armbrust, M. et al., 2009. Above the clouds: A Berkeley view of cloud computing.

*University of California, Berkeley, Tech. Rep. UCB*, pp.1–23. Available at:

<http://scholar.google.com/scholar?q=intitle:Above+the+clouds:+A+Berkeley+view+of+cloud+computing#0>.

Armour, M., 2015. Talking about a (business continuity) revolution: Why best

practices are wrong and possible solutions for getting them right. *Journal of*

*Business Continuity & Emergency Planning*, 9(2), pp.103–111. Available at:

<http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=112058400&site=ehost-live>.

Armour, P.G., 2015. The chaos machine. *Communications of the ACM*, 59(1),

pp.36–38. Available at: <http://dl.acm.org/citation.cfm?doid=2859829.2846086>.

Artus, H.M., 2008. In Quest of Sustainable Information Databases , Database

Analysis , and Database Research - Exploration of a Research Field. In

*Sociology The Journal Of The British Sociological Association*. pp. 159–172.

Available at: <http://www.eurocris.org/Uploads/Web>

[pages/cris2008/Papers/cris2008\\_Artus.pdf](http://www.eurocris.org/Uploads/Web/pages/cris2008/Papers/cris2008_Artus.pdf).

Ashby, R.W., 1956. *An Introduction to Cybernetics* Chapman &., London. Available

at: <http://pespmc1.vub.ac.be/books/IntroCyb.pdf>.

Aslett, M., 2015. The State of Database as a Service. *451 Group*. Available at:

[http://resources.tesora.com/hubfs/Tesora-ebook-State-of-DBaaS-](http://resources.tesora.com/hubfs/Tesora-ebook-State-of-DBaaS-451Group.pdf)

[451Group.pdf](http://resources.tesora.com/hubfs/Tesora-ebook-State-of-DBaaS-451Group.pdf).

## References

- Auerbach, C.E. & Silverstein, L.B., 2003. *Qualitative data: An Introduction to coding and Analysis*, New York: New York University Press.
- Avgerou, C., 2011. Discourses on innovation and development in information systems in developing countries research. In R. D. Galliers & W. L. Currie, eds. *The Oxford Handbook of Management Information Systems*. Oxford: Oxford University Press, p. 650.
- Avgerou, C. & Land, F., 1992. Examining the appropriateness of information technology. In S. Odedra & M. Bhatnagar, eds. *Social Implications of computers in developing countries*. New Delhi: Tata McGraw-Hill, pp. 26–42.
- Bachman, C.W., 1973. The programmer as navigator. *Communications of the ACM*, 16(11), pp.653–658. Available at:  
<http://portal.acm.org/citation.cfm?doid=355611.362534>.
- Ballew, R.M. et al., 1998. A Physical Map of 30,000 Human Genes. *Science*, 282(5389), pp.744–746. Available at:  
<http://74.220.219.81/~mrflipco/teach/writing/GeomProj/images/PhysicalMapOf30000HumanGenes.pdf>.
- Barnes, J., 2015. *Microsoft Azure Essentials: Azure Machine Learning*, Redmond: Microsoft Press. Available at:  
<https://mva.microsoft.com/ebooks#9780735698178>.
- Basalla, G., 1988. *The Evolution of Technology* Reprint., Cambridge: Cambridge University Press.
- Batini, C., Lenzerini, M. & Navathe, S.B., 1986. A Comparative Analysis of Methodologies for Database Schema Integration. *ACM Computing Surveys*,

18(4).

BBC News, 2014. Barclays customer data “stolen & sold” newspaper says. *BBC News*. Available at: <http://www.bbc.co.uk/news/uk-26106138>.

BBC News, 2007. Brown apologises for records loss. *BBC News*. Available at: <http://news.bbc.co.uk/1/hi/7104945.stm>.

Beacham, J., 2006. *Succeeding through innovation: 60 minute guide to innovation turning*, London.

Beck, K. et al., 2001. Principles behind the agile manifesto. Available at: <http://agilemanifesto.org/principles.html>.

Becker, M.C., 2004. Organizational routines: a review of the literature. *Industrial and Corporate Change*, 13(4), pp.643–678. Available at: <http://icc.oupjournals.org/cgi/doi/10.1093/icc/dth026> [Accessed July 9, 2014].

Bekhet, A.K. & Zauszniewski, A., 2012. Methodological triangulation: an approach to understanding data. *Nurse Researcher*, 20(2), pp.40–43.

Bell, G., Hey, T. & Szalay, A., 2009. Beyond the Data Deluge. *Science*, 323(March), pp.1297–1298. Available at: [sciencemag.org](http://sciencemag.org).

Bell, R.L. & Martin, J.S., 2012. The Relevance of Scientific Management and Equity Theory in Everyday Managerial Communication Situations. *Journal of Management Policy and Practice*, 13(3), pp.106–115. Available at: [http://ezp.waldenulibrary.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=79170665&site=ehost-live&scope=site%5Chttp://www.na-businesspress.com/JMPP/BellRL\\_Web13\\_3\\_.pdf](http://ezp.waldenulibrary.org/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=79170665&site=ehost-live&scope=site%5Chttp://www.na-businesspress.com/JMPP/BellRL_Web13_3_.pdf).

## References

- Bell, S., Collins, K. & Lane, A., 2012. Guide to Diagrams. *The Open University Open Learn*. Available at: <http://systems.open.ac.uk/materials/T552/>.
- Bergin, T.J. & Haigh, T., 2009. The Commercialization of Database Management Systems , 1969 – 1983. *Ieee Annals Of The History Of Computing*, pp.1969–1983.
- Berman, F., 2008. Got Data? A Guide to Data Preservation in the Information Age: a guide to data preservation in the information age. *Communications of the ACM*, 51, pp.50–56. Available at: <http://cacm.acm.org/magazines/2008/12/3360-got-data-a-guide-to-data-preservation-in-the-information-age/abstract>.
- Bernstein, P. et al., 1998. The Asilomar Report on Database Research. *ACM SIGMOD Record*, 27(4), pp.74–80.
- Bernstein, P.A. et al., 1989. Future Directions in DBMS Research - The Laguna Beach Participants. *ACM SIGMOD Record*, 18(1), pp.17–26.
- Berriman, G.B. & Groom, S.L., 2011. How Will Astronomy Archives Survive The Data Tsunami? *Communications of the ACM*, 54(12), pp.52–56.
- Von Bertalanffy, L., 1969. *General Systems Theory: Foundations, Development, Applications*, New York: George Braziller.
- Bertino, E. & Sandhu, R., 2005. Database security - concepts, approaches, and challenges. *IEEE Transactions on Dependable and Secure Computing*, 2(1), pp.2–19. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=1416861>.
- Bhagwan, R. et al., 2005. Time-varying Management of Data Storage. *Proceedings*

of the First conference on Hot topics in system dependability HotDep'05.

Blackburn, S., 2001. *Ethics a very short introduction*, Oxford: Oxford University Press.

Blasis, J.-P. De, 1977. Database Administration as a Team Function: An Analysis from Survey Data. In *Proceeding SIGCPR '77 Proceedings of the fifteenth annual SIGCPR conference ACM*. pp. 227–240.

Bolton, C. et al., 2010. *Professional SQL Server 2008 Internals and Troubleshooting*, Indianapolis: Wiley Publishing.

De Bono, E., 1990. *I am right you are wrong*, London: Penguin Books.

Bradley, N., 1999. Sampling for Internet Surveys. An examination of respondent selection for Internet research. *Journal of the Market Research Society*, 41(4), pp.387–395. Available at:  
<http://www.wmin.ac.uk/marketingresearch/Marketing/sam06.html>.

Braun, V. & Clarke, V., 2012. APA handbook of research methods in psychology, Vol 2: Research designs: Quantitative, qualitative, neuropsychological, and biological. In H. Cooper et al., eds. *APA handbook of research methods in psychology*. Washington: American Psychological Association. Available at:  
<http://content.apa.org/books/13620-000> [Accessed July 31, 2014].

Braun, V. & Clarke, V., 2013. *Successful Qualitative Research: a practical guide for beginners* First., London: SAGE Publications Ltd.

Braun, V. & Clarke, V., 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), pp.77–101. Available at:  
<http://www.tandfonline.com/doi/abs/10.1191/1478088706qp063oa>.

## References

- Bretschneider, S., 2004. "Best Practices" Research: A Methodological Guide for the Perplexed. *Journal of Public Administration Research and Theory*, 15(2), pp.307–323. Available at: <http://jpart.oupjournals.org/cgi/doi/10.1093/jpart/mui017> [Accessed July 31, 2014].
- Brooks, F.P.J., 1986. No silver bullet-essence and accidents of software engineering. *Proceedings of the IFIP Tenth World Computing Conference*, pp.1069–1076. Available at: <http://www.cgl.ucsf.edu/Outreach/pc204/NoSilverBullet>.
- Buckingham, A. & Saunders, P., 2009. *The Survey Methods Workbook* Reprint., Cambridge: Polity Press Ltd.
- Bullock, S. & Cliff, D., 2005. Complexity and emergent behaviour in ICT. *Foresight project web site*, pp.1–35.
- Callahan, R.H. & Ishmael, G.S., 2005. *What Drives Innovation? A heuristic framework for corporate innovation*, Available at: <http://www.decisionanalyst.com/Downloads/WhatDrivesInnovation.pdf>.
- Cameron, R., 2009. A sequential mixed model research design : design , analytical and display issues research design : Design. *International Journal*, 3(2), pp.140–152.
- Cannon, D. et al., 2007. *Service Operation*, London: TSO (The Stationary Office).
- Capra, F., 1997. *The Web of Life*, London: Harper Collins Publishers.
- Capra, F. & Luisi, P.L., 2014. *The Systems View of Life*, Cambridge: Cambridge University Press.



- Carpenter, C. & Suto, M., 2008. *Qualitative research for occupational and physical therapists: A practical guide*, Oxford: Blackwell.
- Cartlidge, A. et al., 2007. *An Introductory Overview of ITIL V3*, Available at: [http://itsmfi.org/files/itSMF\\_ITILV3\\_Intro\\_Overview.pdf](http://itsmfi.org/files/itSMF_ITILV3_Intro_Overview.pdf).
- Charmaz, K., 2006. *Constructing Grounded Theory*, London: SAGE Publications Ltd.
- Checkland, P., 1983. O. R. and the Systems Movement: Mappings and Conflicts. *The Journal of the Operational Research Society*, 34(8), pp.661–675. Available at: <http://www.jstor.org/stable/2581700>.
- Checkland, P.B., 1999. *Systems thinking, systems practice* Reprint 19., Chichester: John Wiley & Sons.
- Checkland, P.B., 2011. Systems thinking and soft systems methodology. In R. D. Galliers & W. L. Currie, eds. *The Oxford Handbook of Management Information Systems*. Oxford: Oxford University Press, p. 94.
- Checkland, P.B. & Holwell, S.E., 1998. *Information, systems and information systems – making sense of the field*, Chichester: John Wiley & Sons.
- Checkland, P.B. & Scholes, J., 1999. *Soft Systems Methodology in Action includes a 30-year retrospective*, Chichester: John Wiley & Sons.
- Chen, Y., 2005. Information valuation for Information Lifecycle Management. *Proceedings of the Second International Conference on Autonomic Computing (ICAC'05)*.
- Child, J., 1983. *Organization: A guide to problems and practice*, London: Harper & Row.

## References

- Choy, M., Leong, H.V. & Wong, M.H., 2000. Disaster Recovery Techniques for Database Systems. *Communications of the ACM*, pp.272–280.
- Christensen, C.M., 1997. *The innovator's Dilemma When New Technologies cause great firms to fail*, Watertown: Harvard Business School Press.
- Clarke, V. & Braun, V., 2013. Teaching thematic analysis. *Psychologist*, 26(2), pp.120–124.
- Codd, E.F., 1985. Is your DBMS really relational? *ComputerWorld*, (14 October).
- Codd, E.F., 1971. *Normalized Data Base Structure: A Brief Tutorial*,
- Connolly, T.M. & Begg, C.E., 1995. *Database Systems: A practical approach to Design, Implementation. And management*, Harlow: Addison-Wesley.
- Cooper, J. & James, A., 2009. Challenges for Database Management in the Internet of Things. *Iete Technical Review*, 26(5).
- Creswell, J.W., 2009. *Research Design, Qualitative, Quantitative and Mixed Methods approaches*, Thousand Oaks: Sage.
- Creswell, J.W. & Plano Clark, V.L., 2011a. Choosing a mixed methods design. In *Designing and Conducting Mixed Methods Research*. Thousand Oaks: SAGE Publications Ltd, pp. 53–106.
- Creswell, J.W. & Plano Clark, V.L., 2011b. *Designing and Conducting Mixed Methods Research*, Thousand Oaks: Sage.
- Cuddeford-Jones, M., 2013. Predicting the Future. *Marketing Week (01419285)*, pp.31–33.
- Dani, S. et al., 2006. A methodology for best practice knowledge management.

- Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture*, 220(10), pp.1717–1728. Available at: <http://pib.sagepub.com/lookup/doi/10.1243/09544054JEM651> [Accessed July 31, 2014].
- Date, C.J., 1981. *An Introduction to Database Systems*, Boston: Addison Wesley Longman Limited.
- Davenport, T., 2014. *Big Data @ Work*, Massachusetts: Harvard Business School Publishing.
- Davenport, T.H., 1997. *Information Ecology Mastering the information and knowledge environment: Why technology is not enough for success in the information age*, New York: Oxford University Press.
- Dembowski, F.L., 2013. The Roles of Benchmarking , Best Practices & Innovation in Organizational Effectiveness. *International Journal of Organizational Innovation*, 5(3), pp.6–20.
- Denning, P.J., 2014. Avalanches are coming. *Communications of the ACM*, 57(6), pp.34–36. Available at: <http://dl.acm.org/citation.cfm?doid=2602695.2602324> [Accessed June 16, 2014].
- Denning, P.J., 2002. *The Invisible Future: the seamless integration of technology into everyday life* P. J. Denning, ed., New York: McGraw-Hill.
- Denscombe, M., 2008. *The good research guide: for small-scale social research projects* 3rd ed., Maidenhead: Open University Press.
- Denzin, N.K., 1978. *The Research Act: A theoretical introduction to sociological methods*. Second., New York: McGraw-Hill.

## References

Deutsch, K.W., 1970. *The Impact of Complex Data Bases on the Social Sciences* R. L. Bisco, ed., New York: Wiley - Interscience.

Dilla, W., Janvrin, D.J. & Raschke, R., 2010. Interactive Data Visualization: New Directions for Accounting Information Systems Research. *Journal of Information Systems*, 24(2), p.1. Available at: <http://link.aip.org/link/JINFE3/v24/i2/p1/s1&Agg=doi> [Accessed March 27, 2011].

Dodgson, M., Gann, D.M. & Salter, A., 2008. *The Management of Technological Innovation: Strategy and Practice: The Strategy and Practice*, Oxford: Oxford University Press.

Don, W. & Priess, P., 2008. The Role of an Architect. *The Architecture Journal*, (15), pp.1–32. Available at: <http://msdn.microsoft.com/en-us/architecture/cc505966>.

Drucker, P.F., 1998. The Discipline of Innovation. *Harvard Business Review*, December, pp.3–8.

Drucker, P.F., 2007. *The Essential Drucker*, Oxford: Butterworth-Heinemann.

Drummond, H., 2014. Escalation of Commitment: When to Stay the Course. *Academy of Management Perspectives*, 28(4), pp.430–446.

Drummond, H., 2002. Living in a fool's paradise: the collapse of Barings' Bank. *Management Decision*, 40, pp.232–238.

Eliot & Associates, 2005. Guidelines for Conducting a Focus Group. , pp.1–13. Available at: [http://assessment.aas.duke.edu/documents/How\\_to\\_Conduct\\_a\\_Focus\\_Group.pdf](http://assessment.aas.duke.edu/documents/How_to_Conduct_a_Focus_Group.pdf).

- Erica Wagner & Newell, S., 2011. Changing the story surrounding enterprise systems to improve our understanding of what makes erp work in organizations. In R. D. Galliers & W. L. Currie, eds. *The Oxford Handbook of Management Information Systems*. Oxford: Oxford University Press, p. 401.
- Falconer, J., 2010. "Best Practice" as Worst Practice : Broken Metaphor , Nude Emperor. *Proceedings of the European Conference on Intellectual Capital*, pp.754–762.
- Falconer, J., 2011. Knowledge as Cheating : A Metaphorical Analysis of the Concept of "Best Practice." *Systems Research and Behavioral Science*, 180, pp.170–181.
- Fowler, K., 2008. *SQL Server Forensic Analysis*, Boston: Pearson Education.
- Fowler, M., 2005. The new methodology. Available at:  
<http://www.martinfowler.com/articles/newMethodology.html>.
- Franklin, R., Watson, J. & Crick, F., 2005. *Towards 2020 Science*, Available at:  
<http://research.microsoft.com/en-us/um/cambridge/projects/towards2020science/downloads.htm>.
- Gantz, J., McArthur, J. & Minton, S., 2007. *The Expanding Digital Universe*, Available at:  
[http://www.tobb.org.tr/BilgiHizmetleri/Documents/Raporlar/Expanding\\_Digital\\_Universe\\_IDC\\_WhitePaper\\_022507.pdf](http://www.tobb.org.tr/BilgiHizmetleri/Documents/Raporlar/Expanding_Digital_Universe_IDC_WhitePaper_022507.pdf).
- Gantz, J. & Reinsel, D., 2010. *The Digital Universe Decade – Are You Ready?*, Available at: <http://idcdocserv.com/925>.
- Gantz, J. & Reinsel, D., 2013. *THE DIGITAL UNIVERSE IN 2020 : Big Data , Bigger*

## References

*Digital Shadows , and Biggest Growth in the Far East — United States*,  
Available at: <https://www.emc.com/collateral/analyst-reports/idc-digital-universe-united-states.pdf>.

Gao, H., Barbier, G. & Goolsby, R., 2011. Harnessing the Crowdsourcing Power of Social Media for Disaster Relief. *IEEE Intelligent Systems*, 26(3), pp.10–14.  
Available at:  
<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5898447>.

Gartner Inc., 2013. No Title. *Gartner IT Glossary*. Available at:  
<http://www.gartner.com/it-glossary>.

Garud and Kotha, S., 1994. Using the brain as a metaphor to model flexible production systems. *Academy of Management Review*, 19(4), pp.671–698.  
Available at: <http://amr.aom.org/content/19/4/671.short>.

Gillenson, M.L., 1991. Database Administration at the Crossroads: The Era of End-User-Oriented, Decentralized Data Processing. *Journal of Database Administration*, 2 (4), pp.1–11. Available at: <http://www.irma-international.org/viewtitle/51094/>.

Gillenson, M.L., 1982. The State of Practice of Data Administration-1981. *Communications of the ACM*, 25(10), pp.699–706. Available at:  
<http://portal.acm.org/citation.cfm?doid=358656.358664>.

Gillenson, M.L., 1985. Trends in Data Administration. *MIS Quarterly*, (December), pp.317–326.

Gleick, J., 1998. *Chaos: The amazing science of the unpredictable*, London: Vintage.

- Goguen, J.A., 1999. *The ethics of databases*, Available at:  
<https://cseweb.ucsd.edu/~goguen/papers/4s/4s.html>.
- Gonnering, R.S., 2011. The Seductive Allure Of “Best Practices”: Improved Outcome Is A Delicate Dance Between Structure And Process. *E-CO*, 13(4), pp.94–101.
- Gordon, S.R. & Gordon, J.R., 1992. Organizational hurdles to distributed database management systems (DDBMS) adoption. *Information & Management*, 22, pp.333–345.
- Gratton, L. & Ghoshal, S., 2005. Beyond Best Practice. *MIT Sloan Management Review*, 46(3).
- Gray, J., 2004. The Revolution in Database Architecture. *ACM SIGMOD Technical Report: MSR-TR-2004-31*.
- Greenwood, R.G., 1981. Management by Objectives: As Developed by Peter Drucker, Assisted by Harold Smiddy. *Academy of Management Review*, 6(2), pp.225–230. Available at: <http://www.jstor.org/stable/257878>.
- Gregory, P. et al., 2015. Agile Challenges in Practice: A Thematic Analysis. In *16th International Conference on Agile Software Development, XP 2015*. Helsinki. Available at: <http://oro.open.ac.uk/id/eprint/42061>.
- Guest, G., MacQueen, K.M. & Namey, E.E., 2012. Introduction to applied thematic analysis. In *Applied Thematic Analysis*. Thousand Oaks: SAGE Publications, Inc, pp. 3–20. Available at: [www.sagepub.com/upm-data/44134\\_1.pdf](http://www.sagepub.com/upm-data/44134_1.pdf).
- Guion, L.A., 2002. *Triangulation: Establishing the Validity of Qualitative Studies*, Available at: <http://www.rayman->

## References

bacchus.net/uploads/documents/Triangulation.pdf.

Haas, L.M., 2015. The Power Behind the Throne: Information Integration in the Age of Data-Driven Discovery. *Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data*, p.661. Available at: <http://doi.acm.org/10.1145/2723372.2723373>.

Haerder, T. & Reuter, A., 1983. Principles of Transaction-Oriented Database Recovery. *ACM Computing Surveys*, 15(4), pp.287–317.

Haigh, T., 2006. “ A Veritable Bucket of Facts ” Origins of the Data Base Management System. *SIGMOD Record*, 35(2), pp.33-49-19–29.

Haigh, T., 2012. Fifty Years of Databases. Available at: <http://wp.sigmod.org/?p=688>.

Haigh, T., 2009. How Data Got its Base: Information Storage Software in the 1950s and 1960s. *IEEE Annals of the History of Computing*, 31(4), pp.6–25. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5370776>.

Hammersley, M., 2008. Troubles with triangulation. In *Advances in Mixed Methods Research*. SAGE Publications Ltd, pp. 22–36.

Handy, C.B., 1989. *The Age of Unreason*, London: Arrow Books Ltd.

Handy, C.B., 1985. *Understanding Organizations*, London: Penguin Books.

Hashem, I.A.T. et al., 2015. The rise of “big data” on cloud computing: Review and open research issues. *Information Systems*, 47, pp.98–115. Available at: <http://dx.doi.org/10.1016/j.is.2014.07.006>.

Hatch, M.J., 1997. *Organization Theory Modern Symbolic and Postmodern*



*Perspectives*, New York: Oxford University Press.

- Hayashi, Y., 1992. A structural analysis of database management system technology with some Japanese experience. , 22, pp.347–362.
- Heer, J., Bostok, M. & Ogievetsky, V., 2010. A tour through the Visualization zoo. *Communications of the ACM*, 53(6), pp.59–68.
- Hellerstein, J.M., Stonebraker, M. & Hamilton, J., 2007. Architecture of a Database System. *Foundations and Trends in Databases*, 1(2), pp.141–259. Available at: <http://db.cs.berkeley.edu/papers/fntdb07-architecture.pdf>.
- Henderson, R.M. & Clark, K.B., 1990. Architectural Innovation: The Reconfiguration of Existing Product Technologies and the Failure of Established Firms. *Administrative Science Quarterly*, 35(1), p.9. Available at: <http://www.jstor.org/stable/2393549?origin=crossref>.
- Hensley, Z., Sanyal, J. & New, J., 2014. Provenance in sensor data management. *Communications of the ACM*, 57(2), pp.55–62. Available at: <http://dl.acm.org/citation.cfm?doid=2556647.2556657>.
- Hey, T., Tansley, S. & Tolle, K., 2009. *The Fourth Paradigm: Data Intensive Scientific Discovery* K. T. Tony Hey, Stewart Tansley, ed., Redmond Washington: Microsoft Research. Available at: [http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th\\_paradigm\\_book\\_complete\\_lr.pdf](http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf).
- Hirschheim, R.A. & Klein, H.K., 2011. Tracing the history of the information systems field. In R. D. Galliers & W. L. Currie, eds. *The Oxford Handbook of Management Information Systems*. Oxford: Oxford University Press, pp. 20–

## References

42.

Holwell, S. & Reynolds, M., 2010. Introducing systems approaches. In *Systems Approaches to Managing Change: A Practical Guide*. London: Springer, pp. 1–23. Available at: [http://oro.open.ac.uk/21298/1/systems-approaches\\_ch1.pdf](http://oro.open.ac.uk/21298/1/systems-approaches_ch1.pdf).

Holze, M. & Ritter, N., 2011. System models for goal-driven self-management in autonomic databases. *Data and Knowledge Engineering*, 70(8), pp.685–701. Available at: <http://dx.doi.org/10.1016/j.datak.2011.03.001>.

Huber, G.P., 1990. A Theory of the Effects of Advanced Information Technologies on Organizational Design , Intelligence , and Decision Making. *The Academy of Management Review*, 15(1), pp.47–71. Available at: <http://www.jstor.org/stable/258105>.

Hull, S., 2013. 20 Obstacles to Scalability. *Communications of the ACM*, 56(9), pp.54–58. Available at: [http://dl.acm.org/ft\\_gateway.cfm?id=2512489&type=html](http://dl.acm.org/ft_gateway.cfm?id=2512489&type=html).

Hussein, A., 2009. The use of triangulation in social sciences research: can qualitative and quantitative methods be combined? *Journal of Comparative Social Work*, 1, pp.1–12.

IBM, 2016. *The 2016 State of DBaaS Report: How managed services are transforming*, Available at: <http://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=CDM12345USEN>.

IBM Corporation, 2014. DevOps: The IBM approach. , pp.1–12. Available at: [https://www.ibm.com/developerworks/.../DevOps\\_TheIBMApproach.pdf](https://www.ibm.com/developerworks/.../DevOps_TheIBMApproach.pdf).

Igo, L.B. et al., 2006. How Should Middle-School Students with LD Approach Online

Note Taking? A Mixed-Methods Study. *Learning Disability Quarterly*, 29(2), p.89. Available at:  
<http://www.jstor.org/stable/10.2307/30035537?origin=crossref>.

Imran, S. & Hyder, I., 2009. Security Issues in Databases. *2009 Second International Conference on Future Information Technology and Management Engineering*, pp.541–545. Available at:  
<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5381046>  
[Accessed March 26, 2011].

Ioannidis, Y.E., Saulys, T. & Whitsitt, A.J., 1992. Conceptual Learning in Database Design. , 10(3), pp.265–293.

IT Revolution, 2015. *2015 State of DevOps Report*, Available at:  
<https://puppet.com/resources/white-paper/2015-state-of-devops-report>.

Ivankova, N. V. & Stick, S.L., 2007. Students' persistence in a distributed doctoral program in educational leadership in higher education: A mixed methods study. *Research in Higher Education*, 48(1), pp.93–135.

Jaaron, A. a. M. & Backhouse, C.J., 2014. Service organisations resilience through the application of the vanguard method of systems thinking: a case study approach. *International Journal of Production Research*, 52(7), pp.2026–2041. Available at:  
<http://www.tandfonline.com/doi/abs/10.1080/00207543.2013.847291>.

Jagadish, H. V. et al., 2014. Big data and its Technical Challenges. *Communications of the ACM*, 57(7), pp.86–94. Available at:  
<http://dl.acm.org/citation.cfm?doid=2622628.2611567>.

## References

- Jarrar, Y.F. & Zairi, M., 2000. Best practice transfer for future competitiveness: A study of best practices. *Total Quality Management*, 11(4–6), pp.734–740.  
Available at:  
<http://www.tandfonline.com/doi/abs/10.1080/09544120050008147>.
- Jewell, M., 2004. Big-time ID theft symptom of database culture. *USA Today*.  
Available at:  
[http://usatoday30.usatoday.com/tech/news/computersecurity/2004-08-10-database-culture\\_x.htm](http://usatoday30.usatoday.com/tech/news/computersecurity/2004-08-10-database-culture_x.htm) [Accessed December 14, 2010].
- Jick, T.D., 1979. Mixing Qualitative and Quantitative Methods: Triangulation in Action. *Administrative Science Quarterly*, 24(4), pp.602–611. Available at:  
[www.jstor.org/stable/2392366](http://www.jstor.org/stable/2392366).
- Joch, A., 2007. Always Available. *Oracle*. Available at:  
<http://www.oracle.com/technetwork/issue-archive/2007/07-jan/o17availability-090161.html>.
- Johnson, N., 2009. *Simply Complexity: a clear guide to complexity theory*, Oxford: Oneworld Publications.
- Jones, C., 2006. The Economics Of Software Maintenance in the Twenty First Century. Available at: <http://www.compaid.com/caiinternet/ezine/capersjones-maintenance.pdf>.
- Juiz, C. & Toomey, M., 2015. To govern IT, or not to govern IT? *Communications of the ACM*, 58(2), pp.58–64. Available at:  
[http://dl.acm.org/ft\\_gateway.cfm?id=2656385&type=html](http://dl.acm.org/ft_gateway.cfm?id=2656385&type=html).
- Kahn, B.K., 1983. Some Realities of Data Administration. *Communications of the*

ACM, 26(10), pp.794–799.

Kahn, B.K. & Garceau, L.R., 1985. A Developmental Model of the Database Administration Function. *Journal of Management Information Systems*, 1(4), pp.87–101.

Kandel, S. et al., 2012. Enterprise data analysis and visualization: An interview study. *Visualization and Computer Graphics, IEEE Transactions on*, 18(October), pp.2917–2926.

Kanter, R.M., 2006. + Innovation - The Classic Traps. *Harvard Business Review*, (611), pp.1–14.

Kauffman, S., 1995. *At home in the universe: The search for the laws of self-organization and complexity*, London: Penguin Books.

Kayworth, T. & Whitten, D., 2010. Effective Information Security Requires a Balance of Social and Technology Factors. *MIS Quarterly Executive*, 9(3), pp.303–315.

Kepner, C. & Tregoe, B., 1981. *The New Rational Manager*, Princeton: Kepner-Tregoe, Inc.

Kepner, C. & Tregoe, B., 1965. *The Rational Manager: A Systematic Approach to Problem Solving and Decision Making* Reprint., Princeton: Kepner-Tregoe, Inc.

Khatri, V. & Brown, C. V., 2010. Designing Data Governance. *Communications of the ACM*, 53(1), pp.148–152.

Kimball, R., 2011. *The Evolving Role of the Enterprise Data Warehouse in the Era of Big Data Analytics*, Boulder Creek, CA. Available at:  
<http://www.montage.co.nz/assets/Brochures/DataWarehouseBigDataAnalyticsKimball.pdf>.

## References

- King, E., 2015. The Real World of the Database Administrator. , (March). Available at: <http://www.dbta.com/DBTA-Downloads/ResearchReports/The-Real-World-of-the-Database-Administrator-5237.aspx>.
- Knight, B. et al., 2007. *Professional SQL Server 2005 Administration*, Indianapolis: Wiley Publishing.
- Kolb, D.A., 1984. *Experimental Learning: Experience as the Source of Learning and Development*, Englewood Cliffs: Prentice Hall.
- Kolb, D.A., Rubin, I.M. & McLntyre, J., 1971. *Organizational Psychology: an Experimental Approach to Organizational Behaviour* Reprint., Englewood Cliffs: Prentice Hall.
- Korth, H.F. & Silberschatz, A., 1997. Database Research Faces the Information Explosion. *Communications of the ACM*, 40(2), pp.139–142.
- Kotonya, G. & Sommerville, I., 1998. *Requirements Engineering Processes and Techniques* Reprint., Chichester: John Wiley & Sons.
- Kralj, M., 2008. The Need for an Architectural Body of Knowledge. In W. Don & P. Priess, eds. *The Architecture Journal*. Microsoft, pp. 17–20. Available at: <http://msdn.microsoft.com/en-us/architecture/cc505966>.
- Krueger, R.A. & Casey, M.A., 2009. Focus Groups A Practical Guide for Applied Research. In *Planning the Focus Group Study*. SAGE Publications, Inc, pp. 17–34. Available at: [www.sagepub.com/upm-data/24055\\_Chapter2.pdf](http://www.sagepub.com/upm-data/24055_Chapter2.pdf).
- Kurtz, C.F. & Snowden, D.J., 2003. The new dynamics of strategy : Sense-making in a complex and complicated world. *IBM Systems Journal*, 42(3).
- Landrum, R., 2009. *SQL Server Tacklebox*, Cambridge: Simple Talk Publishing.

- Lane, A. et al., 2012. *Systems Diagramming*, Milton Keynes, U.K.: The Open University. Available at:  
[http://www.open.edu/openlearn/ocw/pluginfile.php/75346/mod\\_oucontent/oucontent\\_download/epub/8dc29e6e9c724088d2fd81ddf4cad7611704ebef/systems\\_diagramming.epub](http://www.open.edu/openlearn/ocw/pluginfile.php/75346/mod_oucontent/oucontent_download/epub/8dc29e6e9c724088d2fd81ddf4cad7611704ebef/systems_diagramming.epub).
- Lenharth, B.Y.A., Nguyen, D. & Pingali, K., 2016. Parallel Graph Analytics. *Communications of the ACM*, 59(5).
- Lewin, R., 1993. *Complexity: Life at the edge of chaos*, London: Phoenix.
- Lewins, A. & Silver, C., 2014. *Qualitative Data Analysis and CAQDAS*, London: SAGE Publications Inc.
- Lewis, M.W. & Grimes, A.J., 1999. Theory Building Metatriangulation : Paradigms From Multiple. *Academy of management review*, 24(4), pp.672–690.
- Liamputtong, P., 2011. *Focus Group Methodology Principles and Practice First.*, London: SAGE Publications Ltd.
- Liamputtong, P., 2009. *Qualitative research methods 3rd Editio.*, Melbourne: Oxford University Press.
- Liu, L. & Huang, Q., 2009. A Framework for Database Auditing. *2009 Fourth International Conference on Computer Sciences and Convergence Information Technology*, pp.982–986. Available at:  
<http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5369543>  
[Accessed March 26, 2011].
- Loshin, D., 2009. *Master Data Management*, Burlington: Morgan Kaufmann Publishers.

## References

- Lund, T., 2012. Combining Qualitative and Quantitative Approaches: Some Arguments for Mixed Methods Research. *Scandinavian Journal of Educational Research*, 56(2), pp.155–165.
- Lungu, I., Velicanu, M. & Botha, I., 2009. Database Systems – Present and Future. *Informatica Economica*, 13(1), pp.84–100.
- Lyon, J.K., 1971. The Role of the Data Base Administrator. *Data Base*, 3(4), pp.11–12.
- Macfarlane, I. & Rudd, C., 2001. ) *IT Service Management version 2.1.a*, Reading: itSMF Ltd.
- Magnabosco, J., 2009. *Protecting SQL Server Data*, Cambridge: Simple Talk Publishing.
- Markus, M.L., 2011. Historical Reflections on the Practice of Information Management and Implications for the field of MIS. In R. D. Galliers & W. L. Currie, eds. *The Oxford Handbook of Management Information Systems*. Oxford: Oxford University Press, pp. 3–15. Available at: [fds.oup.com/www.oup.com/pdf/13/9780199580583\\_chapter1.pdf](http://fds.oup.com/www.oup.com/pdf/13/9780199580583_chapter1.pdf).
- Maslett, M., 2012. Database Landscape Map. *451 Group*. Available at: [http://blogs.the451group.com/information\\_management/2013/02/04/updated-database-lanscape-map-february-2013](http://blogs.the451group.com/information_management/2013/02/04/updated-database-lanscape-map-february-2013).
- Maughan, A., 2010. Six reasons why the NHS National Program for IT failed. *Computer Weekly*. Available at: <http://www.computerweekly.com/opinion/Six-reasons-why-the-NHS-National-Programme-for-IT-failed> [Accessed August 11, 2012].



- May, D.B. & Etkina, E., 2002. College physics students' epistemological self-reflection and its relationship to conceptual learning. *American Journal of Physics*, 70(12), p.1249.
- McBath, F., 2002. *SQL Server Backup and Recovery Tools and Techniques*, Upper Saddle River: Prentice Hall.
- McCririck, I.B. & Goldstein, R.C., 1980. What do data administrators really do? *Datamation*, 26(8), pp.131–134.
- McGehee, B., 2009. *Brad's Sure Guide to SQL Server Maintenance Plans*, Cambridge: Simple Talk Publishing.
- McGregor, M., 2007. When Best Practice is Just Not Good Enough Why and How You Need to be Better than the Best. *BPTrends*, (July), pp.1–2.
- Mckendrick, J., 2011a. Data Cross-Currents : 2011 Survey on Cross-Platform Database Administration. *Unisphere Research, a division of Information Today, Inc*, (June), pp.1–25. Available at: <http://www.dbta.com/DBTA-Downloads/ResearchReports>.
- Mckendrick, J., 2016. Database Lifecycle Management Emerges To Address Ever-More Complex Data Sites : 2016 Survey on DLM Strategies. *Unisphere Research, a division of Information Today, Inc*, (June). Available at: [bit.ly/29GJTIK](http://bit.ly/29GJTIK).
- Mckendrick, J., 2014a. DBA – Security Superhero: 2014 IOUG Enterprise Data Security Survey. *Unisphere Research, a division of Information Today, Inc*, (October), pp.1–29. Available at: <http://www.dbta.com/DBTA-Downloads/ResearchReports>.

## References

- Mckendrick, J., 2014b. Efficiency Isn't Enough: Data Centers Lead the Drive to Innovation. 2014 IOUG IT Resource Strategies Survey. *Unisphere Research, a division of Information Today, Inc*, (February), pp.1–23. Available at: <http://www.dbta.com/DBTA-Downloads/ResearchReports>.
- Mckendrick, J., 2013. From Database Clouds to Big Data: 2013 IOUG Survey on Database Manageability. *Unisphere Research, a division of Information Today, Inc*, (October), pp.1–28. Available at: [http://www.oracle.com/webapps/dialogue/ns/dlgwelcome.jsp?p\\_ext=Y&p\\_dlg\\_id=13789408&src=7897724&Act=33](http://www.oracle.com/webapps/dialogue/ns/dlgwelcome.jsp?p_ext=Y&p_dlg_id=13789408&src=7897724&Act=33).
- Mckendrick, J., 2011b. Managing the Rapid Rise in Data Growth: 2011 IOUG Survey on Database Manageability. *Unisphere Research, a division of Information Today, Inc*, (March), pp.1–26. Available at: <http://www.oracle.com/us/products/database/ioug-managing-db-growth-355001.pdf>.
- Mckendrick, J., 2015. The Rapidly Accelerating Cloud-Enabled Enterprise: 2015 IOUG Survey on Database Manageability. , (May). Available at: <http://www.oracle.com/us/products/database/2015-ioug-survey-db-manageability-2542988.pdf>.
- Mckendrick, J., 2014c. The Vanishing Database Administrator: Survey of Data Professionals' Career Aspirations. *Unisphere Research, a division of Information Today, Inc*, (September), pp.1–30. Available at: <http://www.dbta.com/DBTA-Downloads/ResearchReports>.
- McNaught, C. & Lam, P., 2010. Using wordle as a supplementary research tool. *Qualitative Report*, 15(3), pp.630–643. Available at:

<http://www.scopus.com/inward/record.url?eid=2-s2.0-77953058705&partnerID=tZOtx3y1>.

Mell, P. & Grance, T., 2011a. The NIST definition of cloud computing. *NIST Special Publication*, 145, p.7. Available at: <http://www.mendeley.com/research/the-nist-definition-about-cloud-computing/>.

Mell, P. & Grance, T., 2011b. *The NIST Definition of Cloud Computing Recommendations of the National Institute of Standards and Technology*, Available at: <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>.

Monk, A. & Howard, S., 1998. The Rich Picture: a tool for reasoning about work context. *Interactions*, 5(2), pp.21–30. Available at: <http://portal.acm.org/citation.cfm?doid=274430.274434>.

Moore, D.S., 2010. Long-term data archiving. *Analytical and bioanalytical chemistry*, 396(1), pp.189–92. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19838825> [Accessed March 27, 2011].

Moreno, J.L., 1953. *Who Shall Survive? Foundations of Sociometry, Group Psychotherapy and Sociodrama*, Beacon: Beacon House.

Morgan, D.L., 1998. Practical Strategies for Combining Qualitative and Quantitative Methods: Applications to Health Research. *Qualitative Health Research*, 8(3), pp.362–376.

Morgan, D.L. & Krueger, R.A., 1997. When to Use Focus Groups and Why. , pp.3–19. Available at: <http://csde.washington.edu/~scurran/files/readings/590QM/Week 7/Frey & Morgan Articles.pdf>.

## References

Morgan, G., 1986. *Images of Organization*, Thousand Oaks: SAGE Publications Inc.

Morris, C., 2008. *Quantitative Approaches in Business Studies Seventh.*, Harlow: Pearson Education Limited.

Mosley, M. et al., 2009. *The DAMA Guide to The Data Management Body of Knowledge (DAMA-DMBOK Guide)*, Bradley Beach: Technics Publications.

Mourad, M.B. Al & Hussain, M., 2014. The Impact of Cloud Computing on ITIL Service Strategy Processes. *International Journal of Computer and Communication Engineering*, 3(5), pp.367–371. Available at: <http://www.ijcce.org/index.php?m=content&c=index&a=show&catid=43&id=416>

Muller, H., 2009. *DCC Briefing Paper: Database archiving*, Available at: <http://www.era.lib.ed.ac.uk/handle/1842/3346>.

Mullins, C.S., 2012. *Database Administration: The Complete Guide to DBA Practices and Procedures Second.*, Upper Saddle River, New Jersey: Addison-Wesley.

Mundy, J., Thornthwaite, W. & Kimball, R., 2011. *The Microsoft Data Warehouse Toolkit*, Indianapolis: John Wiley & Sons.

Namey, E. et al., 2007. Data reduction techniques for large qualitative data sets. In *Handbook for team-based qualitative research*. pp. 137–163. Available at: [http://web.stanford.edu/~thairu/07\\_184.Guest.1sts.pdf](http://web.stanford.edu/~thairu/07_184.Guest.1sts.pdf).

National Audit Office, 1998. *A Practical Guide to Sampling*, Available at: [www.nao.org.uk/publications/Samplingguide.pdf](http://www.nao.org.uk/publications/Samplingguide.pdf).

Nattermann, P.M., 2000. Best practice does not equal best strategy. *The McKinsey*

*Quarterly*, 2.

Netz, R. & Noel, W., 2007. *The Archimedes Codex*, London: The Orion Publishing Group Ltd.

O'Brien, J., 1998. *Introduction to Information Systems: An Internetworked Enterprise Perspective Second Alt.*, Boston: Irwin/McGraw-Hill.

O'Donovan, B., 2014. Editorial for Special Issue of SPAR: The Vanguard Method in a Systems Thinking Context. *Systemic Practice and Action Research*, 27(1), pp.1–20.

Olofson, C., 2015. *The Analytic-Transactional Data Platform*, Available at:  
<http://sapassets.edgesuite.net/sapcom/docs/2015/04/886d6fe5-1f7c-0010-82c7-eda71af511fa.pdf>.

Oracle, 2011. *Database as a Service : Reference Architecture – An Overview*,

Oringderff, J., 2004. "My Way": Piloting an Online Focus Group. *International Journal of Qualitative Methods*, 3(3), pp.1–10. Available at:  
[http://www.ualberta.ca/~iiqm/backissues/3\\_3/pdf/oringderff.pdf](http://www.ualberta.ca/~iiqm/backissues/3_3/pdf/oringderff.pdf).

Otey, M., 2010. The rise of cloud computing: is it a resurgence of mainframe/thin computing, or is it the future of our business?". *Windows IT Pro*, May, p.69.

Available at:

<http://find.galegroup.com.libezproxy.open.ac.uk/gtx/infomark.do?&contentSet=IAC- Documents&type=retrieve&tabID=T003&prodId=CDB&docId=A227282593&source=gale&srcprod=CDB&userGroupName=tou&version=1.0>.

Oxford University Press, 2016. Oxford Dictionaries. Available at:

## References

<https://en.oxforddictionaries.com/>.

Pantula, S.G., 2011. Statistics: A Key to Innovation in a Data-Centric World! *Journal of the American Statistical Association*, 106(March), pp.1–5.

Patton, M., 1990a. Purposeful Sampling. In *Qualitative Evaluation and Research Methods*. Beverly Hills, CA: Sage, pp. 169–186.

Patton, M., 1990b. *Qualitative Evaluation and Research Methods*, Beverly Hills, CA: Sage.

Patton, M.Q., 1999. Enhancing the quality and credibility of qualitative analysis. *Health services research*, 34(Patton 1990), pp.1189–1208.

Paulk, M.C. et al., 1993. Capability Maturity Model for Software, Version 1.1. *Software, IEEE*, 98(February), pp.1–26. Available at: <http://www.sei.cmu.edu/library/abstracts/reports/93tr024.cfm>.

Pavlou, K.E. & Snodgrass, R.T., 2008. Forensic analysis of database tampering. *ACM Transactions on Database Systems*, 33(4), pp.1–47. Available at: <http://portal.acm.org/citation.cfm?doid=1412331.1412342> [Accessed March 27, 2011].

Peters, T.J. & Waterman, R.H., 1982. *In Search of Excellence: Lessons from America's Best- Run Companies*, New York: Harper & Row.

Peterson, M., 2004. *Strategic Profile Information Lifecycle Management: A vision for the future*,

Pettinger, R., 2000. *Mastering Organizational Behaviour*, Basingstoke: Palgrave.

Polimeni, J. & Polimeni, R., 2006. Jevons' Paradox and the myth of technological

- liberation. *Ecological Complexity*, 3(4), pp.344–353. Available at:  
<http://linkinghub.elsevier.com/retrieve/pii/S1476945X07000098> [Accessed July 17, 2011].
- Pollard, C. & Cater-Steel, A., 2009. Justifications, Strategies, and Critical Success Factors in Successful ITIL Implementations in U.S. and Australian Companies: An Exploratory Study. *Information Systems Management*, 26, pp.164–175.
- Potgieter, B.C., Botha, J. & Lew, C., 2005. Evidence that use of the ITIL Framework is effective. *18th Annual Conference of the National Advisory Committee on Computing Qualifications, Tauranga*, pp.160–167.
- Potter, S., 2006. *Doing Postgraduate Research* Second. S. Potter, ed., Milton Keynes, U.K.: Open University in association with SAGE Publications.
- Prahalad, C.K., 2010. Best Practices Get You Only So Far. *Harvard Business Review*, (April), p.32.
- Prell, C., 2012. Becoming familiar with social networks. In *Social Network Analysis: history, theory & methodology*. London: SAGE Publications Ltd, pp. 7–18.
- Pugh, D.S., 1990. *Organizational Theory selected readings* Third Edit. P. D.S., ed., London: Penguin Books.
- Quatrani, T., 2000. *Visual Modelling with Rational Rose 2000 and UML*, Upper Saddle River: Addison-Wesley.
- Reiner, D. et al., 2004. Information Lifecycle Management : The EMC Perspective. *Group*.
- Reynolds, D.M. et al., 2014a. Diagramming for development 1 - Bounding realities. Available at: <http://www.open.edu/openlearn/science-maths->

## References

technology/computing-and-ict/systems-computer/diagramming-development-1-bounding-realities/content-section-3.

Reynolds, D.M. et al., 2014b. Diagramming for development 2 - Exploring interrelationships. Available at: <http://www.open.edu/openlearn/science-maths-technology/computing-and-ict/systems-computer/diagramming-development-2-exploring-interrelationships/content-section-0>.

Richard L Nolan, 1973. Computer Data Base: the future is now. *Harvard Business Review*, pp.66–82. Available at: <http://pro.unibz.it/staff/ascime/documents/Computer Data Bases.pdf>.

Richardson, R., 2015. Disambiguating Databases. *Communications of the ACM*, 58(1), pp.54–61.

Rindler, A. & Hillard, R., 2013. *Information Development Using MIKE2.0*, Henderson: Motion Publishing.

Roche, J., 2013. Adopting DevOps Practices in Quality Assurance: Merging the art and science of software development. *ACM Queue*, 11(9), pp.1–8.

Rousseau, D. & Wilby, J., 2014. Moving from Disciplinarity to Transdisciplinarity in the Service of Thrivable Systems. *Systems Research and Behavioral Science*, 31(5), pp.666–677.

Ryan, G.W. & Bernard, H.R., 2000. Data Management and Analysis Methods. In N. K. Lincoln & D. & Y. S., eds. *Handbook of Qualitative Research*. Thousand Oaks, CA: Sage, pp. 769–802. Available at: <http://nersp.nerdc.ufl.edu/~ufruss/documents/ryanandbernard.pdf>.

Saldana, J., 2013. *The Coding Manual for Qualitative Researchers Second.*,



London: SAGE Publications Ltd.

Sanwal, A., 2008. The Myth of Best Practices. *Journal of Corporate Accounting & Finance*, 19(5), pp.51–60.

Savage, N., 2014. The Power of Memory. *Communications of the ACM*, 57(9), pp.15–17. Available at:  
<http://epress.lib.uts.edu.au/journals/index.php/csrf/article/view/2118%5Cnhttp://epress.lib.uts.edu.au/journals/index.php/csrf/article/download/2118/2288%5Cnhttp://epress.lib.uts.edu.au/journals/index.php/csrf/article/view/2118/2288>.

Schein, E., 2010. *Organizational Culture and Leadership* 4th Editio., San Francisco: Wiley.

Schein, E., 1980. *Organizational Psychology* Reprint 19., Englewood Cliffs: Prentice-Hall.

Schoen, H. et al., 2013. The power of prediction with social media. *Internet Research*, 23(5), pp.528–543. Available at:  
<http://www.emeraldinsight.com/10.1108/IntR-06-2013-0115> [Accessed January 21, 2014].

Schuh, P., 2001. *Agile DBA*, Available at: [www.sdexpo.com](http://www.sdexpo.com).

Schumpeter, J.A., 2010. *Capitalism, socialism, democracy* Reprint 19., Abingdon: Routledge Classics.

Seddon, J., 2003. *Freedom from command and control*, Buckingham: Vanguard Press.

Seddon, J., 2008. *Systems thinking and the public sector*, Axminster: Triarchy.

## References

- Seddon, J. & Caulkin, S., 2007. Systems thinking, lean production and action learning. *Action Learning: Research and Practice*, 4(1), pp.9–24.
- Segal, B. et al., 2000. Grid computing: The European data grid project. *IEEE Nuclear Science Symposium and Medical Imaging Conference*, 1, p.2.  
Available at:  
[http://www.researchgate.net/publication/3914275\\_Grid\\_computing\\_the\\_European\\_Data\\_Grid\\_Project/file/72e7e52864de518172.pdf](http://www.researchgate.net/publication/3914275_Grid_computing_the_European_Data_Grid_Project/file/72e7e52864de518172.pdf).
- Senge, P.M., 1990. *The Fifth Discipline: The Art and Practice of The Learning Organization* Reprint., Chatham: Mackays of Chatham.
- Shankar, R. & Menon, R., 2010. MDM Maturity : Pragmatism , Business Challenges , and the Future of MDM. *Business Intelligence Journal*, 15(3), pp.19–26.
- Shoham, Y., 2015. Why knowledge representation matters. *Communications of the ACM*, 59(1), pp.47–49. Available at:  
<http://dl.acm.org/citation.cfm?doid=2859829.2803170>.
- Shu, N.C., Wang, H.K.T. & Lum, V.Y., 1983. Forms approach to requirements specification for database design. *ACM*, pp.161–172.
- Silberschatz, A. et al., 1996. Strategic Directions in Database Systems - Breaking Out of the Box. *ACM Computing Surveys*, 28(4), pp.764–778.
- Silberschatz, A., Stonebraker, M. & Ullman, J., 1995. Database Research : Achievements and Opportunities Into the 21st Century. *NSF Workshop on the Future of Databases Systems Research*, pp.1–17.
- Silberschatz, A., Ullman, J. & Stonebraker, M., 1991. Database Systems: Achievements and Opportunities. *Communications of the ACM*, 34(10),

pp.110–120.

- Sivarajah, U. et al., 2017. Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, pp.263–286. Available at: <http://dx.doi.org/10.1016/j.jbusres.2016.08.001>.
- Smith, S., 1997. *Create that change readymade tools for change management* S. Smith, ed., London: Kogan Page Ltd.
- Snowden, D.J. & Boone, M.E., 2007. A Leader's Framework for Decision Making. *Harvard Business Review*, 85(11), pp.68–76.
- Sockett, G.H. & Iyer, B.R., 2009. Online reorganization of databases. *ACM Computing Surveys*, 41(3), pp.1–136. Available at: <http://portal.acm.org/citation.cfm?doid=1541880.1541881> [Accessed March 27, 2011].
- Sowa, J.F. & Zachman, J.A., 1992. Extending and formalizing the framework for information systems architecture. *IBM Systems Journal*, 31(3), pp.590–616.
- Stein, M.-K. et al., 2015. Coping with information technology: Mixed emotions, vacillation, and non conforming use patterns. *MIS Quarterly*, 39(2), pp.367–392. Available at: <http://misq.org/coping-with-information-technology-mixed-emotions-vacillation-and-nonconforming-use-patterns.html>.
- Stonebraker, M., 2012. New opportunities for New SQL. *Communications of the ACM*, 55(11), pp.10–11. Available at: <http://dl.acm.org/citation.cfm?doid=2366316.2366319> [Accessed December 2, 2012].
- Stonebraker, M., 2016. The land sharks are on the squawk box. *Communications of*

## References

- the ACM*, 59(2), pp.74–83. Available at:  
<http://dl.acm.org/citation.cfm?doid=2886013.2869958>.
- Stonebraker, M. & Robertson, J., 2013. Big Data is “Buzzword du Jour;” CS Academics “Have the Best Job.” *Communications of the ACM*, 56(9), pp.10–11. Available at: <http://dl.acm.org/citation.cfm?doid=2500468.2500471> [Accessed October 13, 2013].
- Storey, V.C. & Goldstein, R.C., 1993. Knowledge-Based Approaches to Database Design. *MIS Quarterly*, (March), pp.25–47.
- Strauss, A.L., 1987. *Qualitative analysis for social scientists*, Cambridge: Cambridge University Press.
- Szulanski, G., 1996. Exploring internal stickiness: Impediments to the transfer of best practice within the firm. *Strategic Management Journal*, 17(S2), pp.27–43. Available at: <http://doi.wiley.com/10.1002/smj.4250171105>.
- Takeda, K. et al., 2010. Data Management for All: the Institutional Data Management Blueprint Project. In *6th International Digital Curation Conference*. Chicago. Available at: <http://eprints.soton.ac.uk/169533/> [Accessed May 22, 2011].
- Tallon, P.P. & Scannell, R., 2007. Information Life Cycle Management. *Communications of the ACM*, 50(11), pp.65–70.
- Tashakkori, A. & Teddlie, C., 2003. *Handbook of Mixed Methods in Social & Behavioral Research*, Thousand Oaks, CA: SAGE Publications Inc.
- Taylor, F.W., 1947. *Principles of Scientific Management* first published 1911, ed., New York: Harper.

- Teddlie, C. & Tashakkori, A., 2009. *Foundations of Mixed Methods Research*, Thousand Oaks, CA: SAGE Publications Inc.
- The 451 group: Cloudscape, 2011. *Cloud Codex*, Available at:  
[http://www.the451group.com/cloudscape/cloudscape\\_free\\_report\\_detail.php?id=1485](http://www.the451group.com/cloudscape/cloudscape_free_report_detail.php?id=1485).
- The Guardian, 2015. Nearly 157000 had data breached in talk talk cyber attack. *The Guardian*. Available at:  
<https://www.theguardian.com/business/2015/nov/06/nearly-157000-had-data-breached-in-talktalk-cyber-attack>.
- The Open Group, 1996. TOGAF - the Enterprise Architecture Framework. Available at: <http://www.opengroup.org/subjectareas/enterprise/togaf> [Accessed March 1, 2015].
- The Open University, 2012. T552 Diagramming. *Systems Thinking and Practice: T552 Diagramming*. Available at: <http://systems.open.ac.uk/materials/T552/>.
- Thurmond, V.A., 2001. The Point of Triangulation. *Journal of nursing scholarship : an official publication of Sigma Theta Tau International Honor Society of Nursing / Sigma Theta Tau*, 33(3), pp.253–8. Available at:  
<http://www.ncbi.nlm.nih.gov/pubmed/11552552>.
- Trauth, E.M., 1989. The evolution of information resource management. *Information & Management*, 16(5), pp.257–268. Available at:  
<http://www.sciencedirect.com/science/article/pii/0378720689900037>.
- Tucker, A.L., Nembhard, I.M. & Edmondson, A.C., 2007. Implementing New Practices: An Empirical Study of Organizational Learning in Hospital Intensive

## References

- Care Units. *Management Science*, 53(6), pp.894–907. Available at: <http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1060.0692> [Accessed September 26, 2014].
- Tuckett, A.G., 2015. Applying thematic analysis theory to practice - A researcher's experience. *Contemporary Nurse*, 6178(December), pp.75–87.
- Turner, V. et al., 2014. The Digital Universe of Opportunities: Rich Data and Increasing Value of the Internet of Things. *IDC White Paper*, (April), pp.1–5. Available at: <http://idcdocserv.com/1678>.
- Utterback, J.M., 1996. *Mastering the Dynamics of Innovation*, Boston: Harvard Business School Press.
- Varga, S., Cherry, D. & D'Antoni, J., 2016. *Introducing Microsoft SQL Server*, Redmond: Microsoft Press.
- Venkatesh, V., Brown, S.A. & Bala, H., 2013. Bridging the Qualitative – Quantitative Divide : Guidelines for Conducting Mixed Methods Research in Information Systems. *MIS Quarterly*, 37(1), pp.21–54.
- VersionOne, 2013. *7th Annual State of Agile Development Survey*, Available at: <http://www.versionone.com/pdf/7th-Annual-State-of-Agile-Development-Survey.pdf>.
- VersionOne, 2015. *9th Annual State Of Agile Survey*, Available at: <http://www.versionone.com/pdf/state-of-agile-development-survey-ninth.pdf>.
- Viekzke, R., 2009. The Internet2 Community and the Large Hadron Collider. *World Wide Web Internet And Web Information Systems*.
- Villar, M. & Kushner, T., 2010. A framework to map and grow data strategy.

*Information Management*, (Nov/Dec), pp.24–27.

Wagner, E.L., Scott, S. V. & Galliers, R.D., 2006. The creation of “best practice” software: Myth, reality and ethics. *Information and Organization*, 16(3), pp.251–275. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S1471772706000121> [Accessed July 15, 2014].

Wagner, S. & Dittmar, L., 2006. The Unexpected Benefits of Sarbanes- Oxley The Unexpected Benefits of Sarbanes- Oxley. *Harvard Business Review*, April, pp.1–10.

Waldrop, M.M., 1992. *Complexity: The emerging science at the edge of order and chaos*, New York: Simon & Schuster.

Watzlawick, P., Weakland, J. & Fisch, R., 1974. *Change Principles of problem formation and problem resolution*, New York: W.W. Norton & Company.

Wellstein, B. & Kieser, a., 2011. Trading “best practices”--a good practice? *Industrial and Corporate Change*, 20(3), pp.683–719. Available at: <http://icc.oxfordjournals.org/cgi/doi/10.1093/icc/dtr011> [Accessed July 30, 2014].

Westrum, R., 2004. A typology of organisational cultures. *Qual Saf Health Care*, 13(Suppl II), p.ii22-ii27.

Wettinger, J., Breitenbacher, U. & Leymann, F., 2015. Dyn Tail - Dynamically Tailored Deployment Engines for Cloud Applications. *2015 IEEE 8th International Conference on Cloud Computing*, pp.421–428. Available at: <http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7214073>.

## References

Whalen, E. et al., 2001. *Microsoft SQL Server 2000 Performance Tuning Technical Reference*, Redmond: Microsoft Press.

Wikipedia, 2016. Database Engine. Available at:  
[https://en.wikipedia.org/wiki/Database\\_engine](https://en.wikipedia.org/wiki/Database_engine).

Woody, B., 2002. *Essential SQL Server 2000: An Administration Handbook*, Boston: Pearson Education.

Woody, B., 2003. The SQL Server Runbook. Available at:  
<http://www.informit.com/guides/content.aspx?g=sqlserver&seqNum=278>.

Yin, R.K., 2009. *Case Study Research Design and Method*, Thousand Oaks: SAGE Publications Inc.

Yuhanna, N., Gilpin, M. & D'Silva, D., 2009. *The Forrester Wave™ : Enterprise Database Management Systems , The Forrester Wave™ : Enterprise Database*, Available at:  
<https://www.forrester.com/The+Forrester+Wave+Enterprise+Database+Management+Systems+Q2+2009/fulltext/-/E-RES46643>.

Zachman, J.A., 1987. A Framework for Information Systems Architecture. *IBM Systems Journal*, 26(3), pp.276–292.

Zairi, M. & Whymark, J., 2000. The transfer of best practices : how to build a culture of benchmarking and continuous. *Benchmarking: An International Journal*, 7(1), pp.62–79.