



DATA POLICIES FOR BIG HEALTH DATA AND PERSONAL HEALTH DATA

By

PEPUKAYI, DAVID, JUNIOR CHITONDO

Thesis submitted in fulfilment of the requirements for the degree

MASTERS OF TECHNOLOGY

In Faculty of Informatics and Design

at the Cape Peninsula University of Technology

Supervisor: Prof. Retha de la Harpe

Cape Town

CPUT copyright information

The dissertation/thesis may not be published either in part (in scholarly, scientific or technical journals), or as a whole (as a monograph), unless permission has been obtained from the University

DECLARATION

I, Pepukayi David Junior Chitondo, declare that the contents of this dissertation/thesis represent my own unaided work, and that the dissertation/thesis has not previously been submitted for academic examination towards any qualification. Furthermore, it represents my own opinions and not necessarily those of the Cape Peninsula University of Technology.



19/10/2016

Signed

Date

ABSTRACT

Health information policies are constantly becoming a key feature in directing information usage in healthcare. After the passing of the Health Information Technology for Economic and Clinical Health (HITECH) Act in 2009 and the Affordable Care Act (ACA) passed in 2010, in the United States, there has been an increase in health systems innovations. Coupling this health systems hype is the current buzz concept in Information Technology, 'Big data'. The prospects of big data are full of potential, even more so in the healthcare field where the accuracy of data is life critical. How big health data can be used to achieve improved health is now the goal of the current health informatics practitioner. Even more exciting is the amount of health data being generated by patients via personal handheld devices and other forms of technology that exclude the healthcare practitioner. This patient-generated data is also known as Personal Health Records, PHR. To achieve meaningful use of PHRs and healthcare data in general through big data, a couple of hurdles have to be overcome. First and foremost is the issue of privacy and confidentiality of the patients whose data is in concern. Secondly is the perceived trustworthiness of PHRs by healthcare practitioners. Other issues to take into context are data rights and ownership, data suppression, IP protection, data anonymisation and re-identification, information flow and regulations as well as consent biases. This study sought to understand the role of data policies in the process of data utilisation in the healthcare sector with added interest on PHRs utilisation as part of big health data. The main research question was: How do health data policies influence health data and big data utilisation? Data was collection of two cases, a public and a private healthcare service provider. Semi-structured interviews were conducted with relevant personnel from the two providers. Documents analysis of the two service providers current data policies was conducted as well as co-design sessions with patients serviced by the both fraternities of healthcare. The method of data analysing the data corpus of the study was thematic analysis. This study presents a significant step forward in understanding what policymakers need to do to delicately manoeuvre around the serious issues impeding full data access and use. Some policy direction recommendations are proposed at the end of the study.

Keywords: Big data; Data policies; Healthcare data; Personal Health Records; Health informatics; Data analytics.

ACKNOWLEDGEMENTS

I wish to thank:

- My supervisor Prof Retha de la Harpe for the guidance throughout the duration of this study and for a genuine interest in me beyond my academic life. The Lord's blessings in your direction.
- My fellow work colleagues Mrs Sophie Bhebe, Ms Moreen Achieng, Mrs Jane Freitas, Mr Kruben Naidoo, Mr Martin Mandioma and Mr Wendal Koopman for their support. Your presence was invaluable.
- My research fellows Mr Oluwamayowa Ogundaini, Mr Zandisile Mkubukeli, Mr Sam Asoba for the encouragement.
- My friends Mr Wishes Sibanda, Mr Innocent Maronedzwe, Mr Stewart Sangu, Mr Godknows Chidzidzi, Ms Yeukai Mhembere, Ms Nyasha Wenhla, Ms Noleen Mkombeza, Ms Ropafadzo Luthuli, Mr David Oliver, Mr Michael Melato, the Mbekela family, the AMEN family for the prayers, for the financial support in tough times, for the smiles and laughter. You guys made it possible.
- My family Mr Kudzai Mumanyi, Ms Tinotenda Mumanyi, Ms Tanyaradzwa Jeche, Ms Tariro Mumanyi, Dr and Mrs Mumanyi, Mrs Mataire, Mr and Mrs Jeche, Mainini Thuba, my grandmothers, Gogo Jeche, Gogo MaDube and Mbuya Chitondo for your thoughts and love.
- My sisters Rudo Katsande, Tafadzwa Chitondo, Ratidzo Katsande and Tatenda Chitondo for keeping me going. This is for you.
- My mothers for the care that made me who I am today.
- The Almighty God the Creator of the heavens and the earth for taking care of me and giving me the strength to see this through. Without His guiding angels, I wouldn't have made till now.

To all of you, I say thank you from the bottom of my heart.

Many thanks to the staff of Netcare and the Department of Health (WC) for availing themselves as my research participants and to the CPUT URF fund for the financial assistance that they also rendered me.

DEDICATION

For Divara

In memory of my father who passed away during the course of this study. May his soul rest in peace as he awaits the coming of our Lord and Saviour Jesus

TABLE OF CONTENTS

DECLARATION	ii
ABSTRACT.....	iii
ACKNOWLEDGEMENTS.....	iv
DEDICATION.....	v
TABLE OF CONTENTS	vi
LIST OF FIGURES	xii
LIST OF TABLES	xii
GLOSSARY	xiii
CHAPTER ONE.....	1
INTRODUCTION	1
1.1 Background to Research Problem.....	1
1.2 Statement of Research Problem	3
1.3 Research Aim.....	4
1.4 Research Questions	4
1.5 Research Objectives.....	5
1.6 Methodology.....	5
1.6.1 Research Philosophy.....	5
1.6.2 Strategies	6
1.6.3 Time Horizon.....	6
1.6.4 Unit of Analysis	6
1.6.5 Sampling.....	6
1.6.6 Data Collection Techniques.....	6
1.6.7 Research Approach.....	6
1.6.8 Data Analysis	7
1.7 Ethical Considerations	7
1.8 Delineation of the Research	7
1.9 Contribution of the Research.....	7
1.10 Research Map.....	8

CHAPTER TWO	11
LITERATURE REVIEW	11
2.1 Patient Data	11
2.1.1 Health Records	11
2.1.2 Paper Based Records	12
2.1.3 Electronic Health Records (EHR).....	12
2.1.4 Patient Centred Design of EHR	14
2.1.5 Meaningful Use of Electronic Records.....	15
2.1.6 Personal Health Records.....	16
2.1.7 Summary of Patient Data	19
2.2 Big Data.....	19
2.2.1 Big Data in Business.....	21
2.2.2 Big Data in Healthcare.....	23
2.2.3 Summary of Big Data.....	24
2.3 Data Policies	24
2.3.1 Nature of Current Data Policies.....	25
2.3.2 Data Policies on Electronic Health Records	25
2.3.3 Policy Models in Electronic Health Records	27
2.3.4 Data Policies on Personal Health Data.....	28
2.3.5 Data Policies on Data Rights and Ownership.....	29
2.3.6 Data Policies and Data Analytics	30
2.3.7 Data Policies on Big Data.....	30
2.3.8 A Big Health Data Policy Theoretical Framework.....	31
2.3.9 Summary of Data Policies	35
2.4 Data Utilisation	36
2.4.1 Consensual Access VS Non-consensual Access to Patients Health Data	36
2.4.2 Suppression of Data	37
2.4.3 The Cost of IP Protection in Health	37
2.4.4 Regulations on Information Flow	38
2.4.5 Anonymisation and Re-Identification.....	38
2.4.6 The Open Data Case.....	38
2.4.7 Summary of Data Utilisation	39
2.5 The South African Scenario.....	39

2.6	Conceptual Model	41
2.7	Conclusion – The Policy Lacuna	43
CHAPTER THREE		44
RESEARCH METHODOLOGY AND DESIGN		44
3.1	Statement of Research Problem	44
3.2	Research Aim.....	44
3.3	Research Questions and Objectives	44
3.3.1	Research Questions	44
3.3.2	Research Matrix and Objectives	45
3.4	Research Methodology	46
3.4.1	Research Philosophy.....	46
3.4.2	Research Approach.....	49
3.4.3	Research Strategy.....	50
3.4.4	Units of Analysis and Sampling.....	50
3.4.5	Time Horizon.....	51
3.4.6	Data Collection Techniques.....	51
3.4.7	Thematic Analysis.....	52
3.4.8	Data Presentation	54
3.4.9	Research Quality Management.....	54
3.5	Conclusion	56
CHAPTER FOUR		57
FIELDWORK REPORT		57
4.1	Data Collection.....	57
4.2	Undertaking Documents and Literature Review.....	57
4.3	Undertaking the Semi-structured Interviews	58
4.3.1	Interview Protocol and Process.....	58
4.3.2	Transcription of the Interviews	59
4.4	Undertaking Co-Design Sessions.....	59
4.5	Summary of Data Collection	62
4.6	Undertaking Data Analysis Using Thematic Analysis	62
4.6.1	Generating Codes.....	63
4.6.2	Generating Themes	64

4.6.3	Summary of Data Analysis	65
4.7	Reflections and Experiences	66
4.8	Conclusion	67
CHAPTER FIVE		68
DISCUSSION OF FINDINGS		68
5.1	Access to Information Formulation	68
5.1.1	Acquiring Consent	69
5.1.2	Politics and Information Access	70
5.1.3	6 W's of Information Access.....	70
5.1.4	Stigmatisation and Prejudice as Results of Information Access	71
5.1.5	Robust Audit Mechanisms.....	71
5.1.6	Presentation of Policies	71
5.1.7	Summary of Access to Information Formulation	72
5.2	National Wide Healthcare Repository	72
5.2.1	National Patient Indexing System	73
5.2.2	National Reporting Structures	73
5.2.3	National Algorithms for Data Analytics.....	74
5.2.4	Central Repository of all National Healthcare Data	74
5.2.5	The Future with Personal Health Records	75
5.2.6	Ownership of Healthcare Data	75
5.2.7	Summary of National Wide Healthcare Repository	76
5.3	Healthcare Open Data Policy Management	76
5.3.1	Adoption of an Open Data Policy	76
5.3.2	Mapping BHD Activities to Patient Outcomes.....	77
5.3.3	Huge Turnaround of ICT Personnel in Health.....	77
5.3.4	Partnerships Between the Private and Public Healthcare Sectors.....	78
5.3.5	Prioritising Information Sharing	79
5.3.6	Handling Add-on Infrastructure and Systems	79
5.3.7	Summary of Open Data Policy Management.....	80
5.4	Restricted Field Open Data Policy.....	80
5.5	Information Accountability.....	80
5.5.1	Auditing Mechanisms.....	80
5.5.2	Legal Consequences of Information Abuse	81

5.5.3	Summary of Information Accountability.....	81
5.6	Holistic Patient Visualisation.....	81
5.6.1	Continuous Care.....	81
5.6.2	Patient Life Cycle.....	82
5.6.3	Summary of Holistic Patient Visibility.....	82
5.7	Conclusion.....	82
CHAPTER SIX.....		84
CONCLUSION.....		84
6.1	What the Literature Had Been Saying.....	84
6.2	What the Research Problem Was.....	84
6.3	Methods Employed in the Research.....	85
6.4	A Framework to Develop a Big Health Data Environment.....	86
6.5	Summary of Findings and Results.....	86
6.6	Research Questions Revisited.....	88
6.6.1	RSQ1.1: How are personal health data and big health data policies affecting data use?	88
6.6.2	RSQ1.2: How are healthcare data policies factoring into consideration foresighted changes in data use strategies?	88
6.6.3	RSQ2.1: How do current personal health data and big health data policies promote information access and sharing?	89
6.6.4	RSQ2.3: What framework can be used to improve healthcare data policies in the context of big health data?	89
6.6.5	Main Research Question 1: How are healthcare policies influencing data use trends like big data and personal health records?.....	89
6.6.6	Main Research Question 2: What is the nature of current data policies in the context of healthcare data use?	90
6.7	Research Limitations.....	90
6.8	Evaluation of the Research.....	91
6.8.1	The Fundamental Principle of the Hermeneutic Circle.....	91
6.8.2	The Principle of Contextualisation.....	91
6.8.3	The Principle of Interaction Between the Researchers and the subjects.....	91
6.8.4	The Principle of Abstraction and Generalisation.....	92
6.8.5	The Principle of Dialogical Reasoning.....	92

6.8.6	The Principle of Multiple Interpretations	92
6.8.7	The Principle of Suspicion	92
6.9	Further Research.....	93
REFERENCES		94
APPENDICES.....		101
A1	- Interview Themes	101
A2	- Interview Questions.....	101
A3	- Interviewees List.....	105
A4	- Interview Transcripts are Available and Stored as Part of Research Data on the Institutional Research Repository.....	106
A5	- Co-Design Sessions Plan	106

LIST OF FIGURES

Figure 1-1: Research Map	10
Figure 2-1: Data Ecosystem.....	21
Figure 2-2: Big Health Data Policy Theoretical Framework.....	32
Figure 2-3: Big Health Data Theoretical Framework: Skillset requirements	33
Figure 2-4: Big Health Data Theoretical Framework: Regulation requirements	33
Figure 2-5: Big Health Data Theoretical Framework: Patient Involvement requirements.....	34
Figure 2-6: Big Health Data Theoretical Framework: Data Use requirements.....	35
Figure 2-7: Big Health Data Theoretical Framework: All requirements	35
Figure 3-1: Four paradigms of analysis of social theory	48
Figure 4-1: Co-design session with public healthcare users in progress	60
Figure 4-2: Co-design session with private healthcare users in progress.....	61
Figure 4-3: Co-Design session's participants	62
Figure 4-4: Data extracts on a coded script.....	63
Figure 4-5: Data extract and subsequent code.....	64
Figure 4-6: Initial Thematic Map.....	64
Figure 4-7: Final thematic Map.....	65

LIST OF TABLES

Table 1-1: Research Questions.....	4
Table 1-2: Research Matrix	5
Table 2-1: The State of Health in the World.....	40
Table 4-2: Map of developed themes to the research questions they are linked to	66

GLOSSARY

BHD: Big Health Data

Big Data: Data and its processing methods that are so complex compared to their traditional counterparts

Data: a set of qualitative or quantitative variables that may or may not be related or organised

DoH: Department of Health

E-Health: Electronic Health

EHR: Electronic Health Records

EMR: Electronic Medical Records

HITECH: Health Information Technology for Economic and Clinical Health Act

ICT: Information Communication Technology

IT: Information Technology

Information: data that has been processed to gain a structure and meaning. NB: In the context of this study the terms data and information have been used interchangeably

MDGs: Millennium Development Goals

Patient: An individual who once or presently is receiving care

PHR: Personal Health Records

PAIA Act: Promotion of Access to Information Act

Policy: Principles and rules to guide decisions and achieve rational outcomes

POPI Act: Protection of Personal Information Act

SANHA: South Africa National Health Act

WCDoH: Western Cape Department of Health

WHO: World Health Organisation

CHAPTER ONE

INTRODUCTION

The information age has seen a rapid increase of electronic information systems development. These information systems, are generating gigantic sums of data which have led to the phenomena in the Information Communications and Technology (ICT) world called big data. A lot has been said about the prospects of big data and its potential in the field of data analytics. Some of these potentials are being realised in the business world and other fields and great strides are being made in data utilisation for better decision making. The medicine and healthcare field has been slow in embracing the use of electronic information systems in their day to today operations. This is because of the nature of their business policies among other things but perhaps more notably because of the way healthcare professionals relate to and utilise information. The last decade, however, has seen an increase in the use of electronic health information systems and electronic health records (EHR). This overwhelming turn to EHR has seen the amount of data generated by health information systems ballooning, therefore, leading to what is called “Big Health Data” in the health informatics world. Health institutions are finding themselves with huge datasets of information and they cannot help but wonder what other knowledge can be gleaned from that data and also if the general prospects of big data can also be realised in their context. One of the main difference between healthcare information and other types of information is the information policies that are used to manage healthcare information. This is mostly because of its sensitivity and personal nature. Privacy concerns around healthcare information have translated to a direction of information policies that are protective of the information making it difficult to access and use the information. What role health information policies play in enhancing or retarding the prospects of “Big Health Data” with a focus on personal health data will be the central focus of this study. Questions around the relationship between the former and the latter will be closely examined and hopefully, some useful outcomes for health informatics and medical professionals, healthcare institutions, government and the health fraternity at large will be reached.

1.1 Background to Research Problem

Health information policies are constantly becoming a key feature in directing information usage in healthcare (Hripcsak et al., 2014; Keen et al., 2013; Berndt et al., 2001). After the passing of the Health Information Technology for Economic and Clinical Health (HITECH) Act in 2009 and the Affordable Care Act (ACA) passed in 2010, in the United States, there has been an increase in health systems innovations and healthcare data use initiatives (Adler-milstein et al., 2014;

Grandia, 2014; Hoffman & Podgurski, 2013b). Coupling this health systems hype is the current buzz concept in Information Technology, 'Big Data' (Ohno-Machado, 2014; Keen et al., 2013; Manyika et al., 2011; Damle, 2014). The prospects of big data are full of potential, even more so in the healthcare field where the accuracy of data is life critical (Crawford & Schultz, 2013; World Economic Forum, 2013; Yao et al., 2010). How big health data can be used to achieve improved health is now the goal of the current health informatics practitioner. Even more exciting is the amount of health data being generated by patients via personal handheld devices and other forms of technology that exclude the healthcare practitioner (Keen et al., 2013; World Economic Forum, 2013; Genitsaridi et al., 2013; World Economic Forum, 2014; Kearney, 2014). Patient-generated data is also known as Personal Health Records (PHR).

Literature has shown that healthcare is not maximising on the opportunities that are presented by big health data and personal health records (Manyika et al., 2011). Healthcare information is being under-utilised and the attitude that needs to be adopted for health data is one that encourages access and sharing so as to achieve realisable meaningful use of healthcare information (Blumenthal & Tavenner, 2010; Studeny & Coustasse, 2014; DesRoches et al., 2013). This will mean healthcare will have to adopt a more open data approach to its view of data utilisation (Zuiderwijk & Janssen, 2014b; Yiu, 2012b). This is not the case currently because of quite valid reasons like privacy and confidentiality (Terry, 2013; Menius Jr & Rousculp, 2014). Policies will have to find ways on how healthcare can be more receptive of the open data.

To achieve meaningful use (Blumenthal & Tavenner, 2010; Studeny & Coustasse, 2014; DesRoches et al., 2013; Crawford & Schultz, 2013) of PHRs as part of healthcare data through big data, a couple of hurdles have to be overcome (DesRoches et al., 2013; Keen et al., 2013; Keen et al., 2012; Hripcsak et al., 2014; Ohno-Machado, 2014; Hoffman & Podgurski, 2013b; Hoffman & Podgurski, 2013a). First and foremost is the issue of privacy and confidentiality of the patients whose data is in concern (Terry, 2013; Menius Jr & Rousculp, 2014). Secondly is the perceived trustworthiness of PHRs by healthcare practitioners (Kearney, 2014; University of California, 2014). Other issues to take into context are data rights and ownership, data suppression, IP protection, data anonymisation and re-identification, information flow and regulations as well as consent biases.

The South African government is faced with various health challenges among those the epidemics of HIV/AIDS, tuberculosis and malaria (Gray & Vawda, 2015; KPMG Africa Limited,

2012; Department of Health South Africa, 2012). It is also faced with under-budgeting for health as a result of the limited resources it has available and the result of this has been huge inequalities in the ability to access healthcare services (Phelan et al., 2010; Statistics South Africa, 2011; Harris et al., 2011). The South African government will need to harness the opportunities that data utilisation technologies such as big data and personal health records provide to improve the health of its population (Kruger, 2016; Department of Health South Africa, 2012; Western Cape Government Health, 2014). The government has recognised this and has bankrolled various projects to this effect but more data use projects are needed and are being planned (Health Information Systems Program-SA, 2015). For these projects to be successful they will require an eHealth policy platform to support them that takes into consideration the nation's challenges, abilities and capabilities. However, the country also has a poor record of health policy implementation (Gray & Vawda, 2015). How to overcome this is still a challenge.

This study will seek to understand the role of data policies in the process of data utilisation in the healthcare sector with a particular interest in PHRs utilisation as part of big health data. The term data utilisation in this scope is being used to denote the access and use of information to give more meaning to the data. The main research question will be: How do health data policies influence personal health data and big data utilisation? The focus on PHRs is because trends forecast PHRs to become the main contributor of healthcare data in the next few years (Anon, 2013; Kearney, 2014), therefore the relevance of this study. This present study represents a significant step forward in understanding what policymakers need to do to delicately manoeuvre around the serious issues impeding full data access and use.

1.2 Statement of Research Problem

Despite some success of big data methods in other fields such as business informatics (Manyika et al., 2011; Terry, 2013), the same results seem not to be as easy to achieve in health informatics (Page, 2014; World Economic Forum, 2013). One main difference between business informatics and health informatics is the orientation of their data policies. Can a link be made between data policies and the success of big data in healthcare? There seems to be a mismatch between the goals of current healthcare data policies and those of big data in healthcare. The role of PHRs in healthcare also seems to be a peripheral issue. This is negatively affecting research efforts in healthcare by hampering the adoption of big data technologies for better data utilisation, access and sharing (Keen et al., 2012; Kearney, 2014; Ohno-Machado, 2014; Vayena et al., 2015). If the status quo prevails, some valuable insights into the health of the world population in general which have the potential of saving lives will continue to be overlooked (Page, 2014). How these two relate needs to be further researched in order to explore possible

ways of better congruency and overcoming the issues around open access and sharing of healthcare data.

1.3 Research Aim

The aim of my research is to explore the relationship between data policies and the prospects presented by big health data with a focus on personal health data. If this could be properly pointed out then the design of policies that support big data utilisation will be greatly facilitated. Hopefully, some recommendations for policy making will be made at the end of this study.

1.4 Research Questions

Table 1-1: Research Questions

Research problem: There seems to be a mismatch between the goals of current healthcare data policies and those of big data in healthcare. The role of PHRs in healthcare also seems to be a peripheral issue.	
Main Research question 1: How are healthcare policies influencing data use trends like big data and personal health records in South Africa?	Main Research question 2: What is the nature of current data policies in the context of healthcare data use in South Africa?
RSQ1.1: How are personal health data and big health data policies affecting data use?	RSQ2.1: How do current personal health data and big health data policies promote information access and sharing?
RSQ1.2: How are healthcare data policies factoring into consideration foresighted changes in data use strategies?	RSQ2.2: What are the differences that could be existing between current healthcare data policies and those of big health data?
RSQ1.3: How are health data policies providing support and a foundation for big health data and personal health data policies?	RSQ2.3: What framework can be used to improve healthcare data policies in the context of big health data?

1.5 Research Objectives

Table 1-2: Research Matrix

Research Questions/ Objectives, data collection and analysis methods.	Objectives	Document Analysis	Semi- structured Interviews	Affinity Diagrams	Thematic Analysis
How are healthcare policies influencing data use trends like big data and personal health records in South Africa?	To determine how aligned current personal health data policies and big health data policies are.	X	X		X
How are personal health data and big health data policies affecting data use?	To gain insights into how data policies can increase data utilisation	X	X		X
How are healthcare data policies factoring into consideration foresighted changes in data use strategies?	To determine the agility and responsiveness of big data systems to data utilisation changes.	X	X		X
How are health data policies providing support and a foundation for big health data and personal health data policies?	To determine how aligned current personal health data policies and big health data policies are.	X	X	X	X
What is the nature of current data policies in the context of healthcare data use in South Africa?	To determine what are the aims of personal data policies and those of big data policies.	X	X		X
How do current personal health data and big health data policies promote information access and sharing?	To determine how information access can be increased by policy measure.	X	X	X	X
What are the differences that could be existing between current healthcare data policies and those of big health data?	To examine what is required to make personal health data policies congruent with big health data policies.	X	X		X
What framework can be used to improve healthcare data policies in the context of big health data?	To determine the basis of formulation of future data policies.	X	X	X	X

1.6 Methodology

1.6.1 Research Philosophy

This research is aligned with the interpretivism philosophy of research which believes in the subjectivity of the meanings of research results. (Saunders & Tosey, 2012). Because big data policing is still a fairly new subject, not much research and implementation has been done on the

subject, therefore, there is little empirical evidence to go by. A positivist paradigm will be difficult to employ in this case.

1.6.2 Strategies

The multiple case study strategy was employed for this study. South Africa's healthcare setup is a two tier system with private healthcare and public healthcare operating almost in parallel to each other. It is because of this setup of healthcare provision that the strategy of this research has been designed as a multiple case study approach as one case would be from the public sector and the other from the private sector.

1.6.3 Time Horizon

This study is a cross-sectional study i.e. its results will relate to a specific time period and cannot be applied over extended periods of time.

1.6.4 Unit of Analysis

The units of analysis comprised of a public healthcare service provider as well as a private healthcare service provider. Personnel and patients of the selected service providers formed the research population of the study ensuring representation of the two tiers of healthcare.

1.6.5 Sampling

Non-probability sampling is going to be the sampling technique used to draw the research sample. Personal health data and big data are still largely new areas, in definition and formation (Vayena et al., 2015). This makes the research population largely unknown and most likely biased towards the middle and upper-class individuals of society (University of California, 2014) as well as affluent health institutions with big data capability. Therefore the methods used to select the sample reflect the non-probability sampling technique.

1.6.6 Data Collection Techniques

Semi-structured interviews, documents reviews, and affinity diagrams were the data collection techniques adopted for the study.

1.6.7 Research Approach

This research employed the inductive approach of theory or hypothesis creation. The inductive approach of research has an inclusive approach to data collection and analysis and this tends to provide research results that are non-biased to the initial perceptions and assumptions of the researcher. All possible leads of truth can be discovered and pursued. This is the reason why

this study follows the inductive approach. This inductive approach is also more suitably aligned to the research philosophies (subjectivism and interpretivism) underpinning this study.

1.6.8 Data Analysis

Thematic analysis of data were employed to analyse and interpret the data collected. Latent themes were identified from the data because a deeper interpretation of the results was required to answer the bulk of the research questions guiding this study.

1.7 Ethical Considerations

Research ethics refer to the epistemic imperative or the moral commitment that researchers are required to make in the search for truth and knowledge (Mouton, 2004). Ethics give researchers guidelines on how they should conduct research. Research design should encompass ethical considerations that protect the well-being of research participants and this should be considered for both humans and non-humans. Ethical guidelines include but are not limited to, informed consent, confidentiality and anonymity as stipulated in the South African National Health Act (SANHA) (South African Government, 2004). Harm to respondents, voluntary participation, respect for study sites, and honest reporting of findings are other ethical considerations to be observed. At every stage of this study, I intend and promise to ensure that these guidelines are followed. I also intend not to proceed with any of the research processes unless written permission has been obtained were necessary from the health research ethics committee of the institution in concern as stipulated in the SANHA (ibid). The Protection of Personal Information Act (POPI) (Parliament of the Republic of South Africa, 2013) also stipulates the importance of ensuring that the ethical code of the institution or individual in concern is respected when engaging with their information. The research plans to respect this premonition.

1.8 Delineation of the Research

This study is not to come up with policies but recommendations that could be incorporated by policy makers. No patient data or information was investigated.

1.9 Contribution of the Research

This research contributes to the Data Management Association (DAMA) Guide to the Data Management Body of Knowledge (DAMA DMBOK Guide). The DAMA DMBOK specifies ten data management functions (Mosley, 2008) of which this research entails aspects that are related to these ten functions. The research also contributes to the Information Systems Development Approach for Societal Development (ISD4D) Architecture design petal (Korpela,

2011). The research also makes a contribution by coming up with data policy guidelines that could be used in healthcare data policy making.

Importantly, one other major contribution that this study makes to the big data world is a big health data theoretical framework that can be a valuable tool in the development of healthcare policy infrastructure and controls. This framework has the potential to cover the gap that currently exists between big health data users and the envisioned advantages and uses of big data healthcare. The ultimate purpose of the study is to increase meaningful usage of healthcare information for the purposes of improving health. If this could be realised, ultimately many lives could be saved.

1.10 Research Map

This thesis is subdivided into six chapters.

Chapter one introduces the background of the research and provides a summarised view of all the research process and concepts. The research questions, aim and objectives guiding this study are outlined. The underlying philosophical stances of the researcher have also been outlined as well as the research design.

Chapter two presents a detailed literature review of the current academic literature on the topics of the research. It is subdivided into four main sections which are in the following sequence: patient data, big data, data policies and data utilisation. The chapter closes by motivating the rationale of the study.

Chapter three presents the philosophical underpinnings of the research, the research methods as well as the research design. The rationale is given why the multiple case study approach was selected in this case as well as reasons for all the other design choices.

Chapter four is the fieldwork report that will give feedback on how data was collected for the research. It outlines how interviews were conducted, with whom and how information was preserved for further analysis. Challenges faced during this process are also outlined. How information gathered was analysed is also discussed in this chapter.

Chapter five presents the findings from the data analysis process carried out and the subsequent discussion of the findings outlined. Results and conclusions about the current situation are drawn from the findings.

Chapter six presents the summary of the whole research. Recommendations based on the results of the study are outlined and possible policymaking guidelines are also discussed. Further possible research projects branching off from this study are also presented.

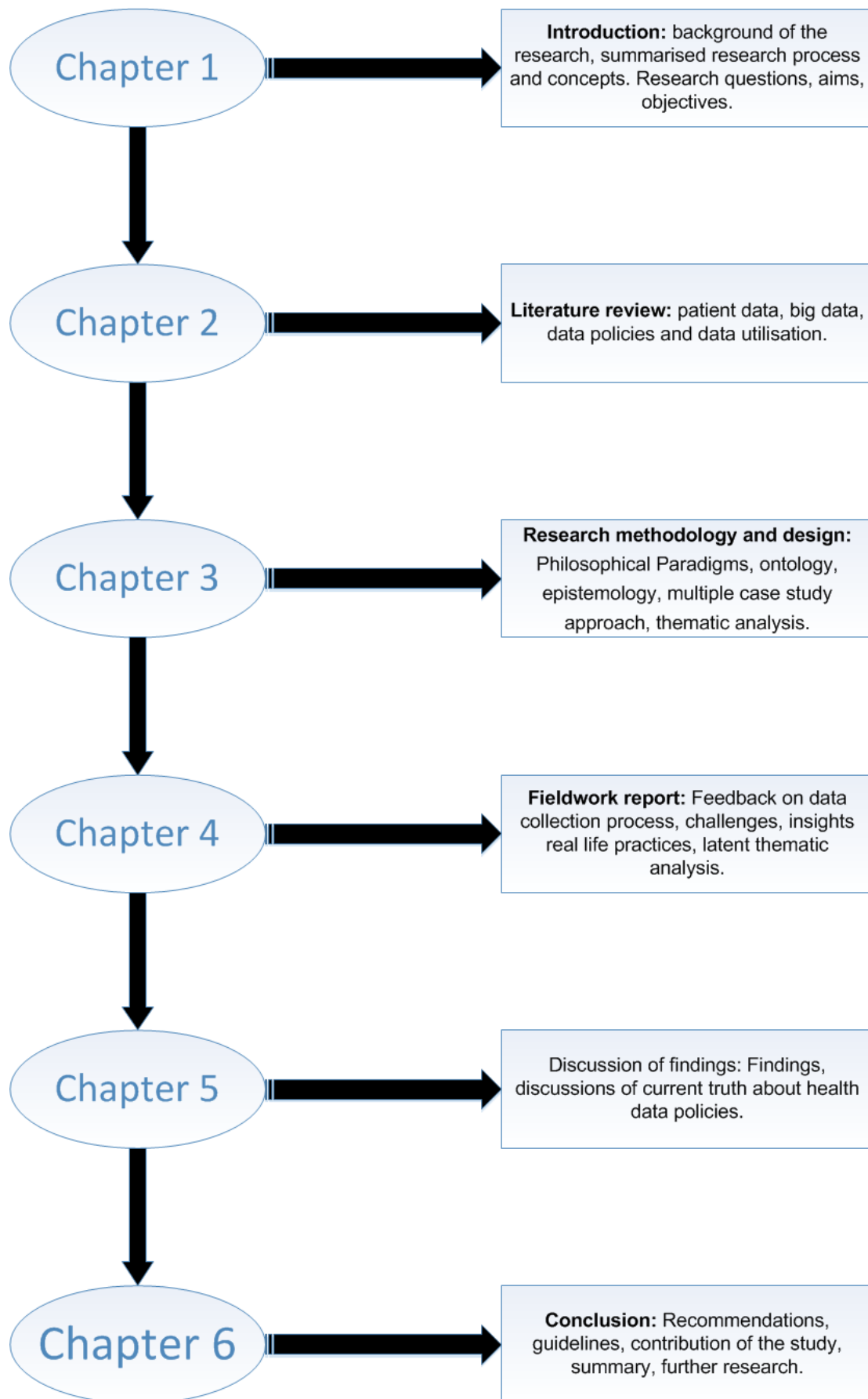


Figure 1-1: Research Map

CHAPTER TWO

LITERATURE REVIEW

Before undertaking this research, a thorough literature review was undertaken to properly understand the theoretical aspects involved and to identify the research gap. This was done to derive the research problem as well as to achieve a comprehensive understanding of the terms and concepts involved in this research. A section of the literature review was also allocated to understand design research as it is going to be an integral part of the research design. Online databases with academic publications were the most used sources of literature. This was because of their ease of availability in many different locations. The most used databases were Google Scholar, Academic Premier, EBSCOhost, ProQuest, PubMed and MEDLINE. The keywords mostly used when searching for literature were big data OR healthcare data OR personal healthcare data OR health informatics OR data analytics OR data policies.

The literature review started by looking at various kinds of patient data and how these are being used in healthcare today. It then went on to focus on big data and its implications on business and healthcare. The literature on data policies was then sourced so as to understand the intentional alignment of future goals in healthcare and big health data. The review then concluded with a look at design research as a method of research in this context. The flow of this chapter will thereby follow the outline mentioned above.

2.1 Patient Data

This section on patient data will deal with health records; electronic health records and their design; meaningful use of health records and is concluded with personal health records.

2.1.1 Health Records

Like any other field of work, there is documentation that comes as a result of the work processes taking place. This documentation is then what we call records and in healthcare, they are called health records. Health records provide a history of health-related happenings concerned with a particular entity which can be a patient, medical units such as a ward or ambulance or even a fully-fledged hospital. Health data can be classified into four main categories i.e. “(1) drug and device, (2) clinical data, (3) claims and related financial data, and (4) patient behaviour and sentiment data” (Terry, 2013: pg 8). This research was mostly concerned with the health records at person level i.e. records about individuals’ health. There are various types of health records depending mostly on the way that that data is

created stored and utilised. Paper-based health records are records that have been captured on hard copy paper or forms that were the original media used to capture patient health-related data. There are still mostly used in developing contexts. There is also electronic health records which are records that have been captured using some form of computerised information system and finally, there are personal health records. The following sections will discuss some of these variations of health records.

2.1.2 Paper Based Records

Paper-based patient records can generally be described as the form of information management that stores and processes patient information using paper. Paper-based patient records are still widely popular in the healthcare sector today (Western Cape Government Health, 2014; Studeny & Coustasse, 2014) despite a growing shift towards electronic health records (Hoffman & Podgurski, 2013b; Hoffman & Podgurski, 2013a). In many cases paper-based patient records are being used in parallel with or supporting electronic health records (Segall et al., 2011). The role that this traditional method of managing patient information will continue to play into the far future cannot be undermined, therefore it is important that they continue to be noted as a useful form of information management (Fielding et al., 2016; Gans et al., 2014).

2.1.3 Electronic Health Records (EHR)

EHRs are defined as “applications for manipulating and processing any information, of an individual, that resides in electronic systems for the purpose of providing health care and health-related services.” (Castillo et al., 2010: pg 2). They are viewed as having the capability to improve how healthcare is administered to the patient by reducing medical errors, supporting decision making and reducing the cost of information sharing among professionals (ibid). EHRs are gaining popularity among healthcare users at a fast pace and this trend is likely to continue (Babbott et al., 2014) but there are many problems lurking on the horizon.

Very large datasets of Electronic Health Records (EHR) are being generated every day by health information systems across the world (Damle, 2014; Hoffman & Podgurski, 2013b). Despite that the healthcare fraternity has been slower in adopting electronic information systems (Chang et al., 2013; Adler-milstein et al., 2014) than some other industries, the last decade has seen a rapid adoption of automated and electronic information systems (Grandia, 2014; Hoffman & Podgurski, 2013b). Massive efforts are being made to create medical databases, at the private sector level, hospital level and even national level

(Hoffman & Podgurski, 2013b). The Hadoop data processing platform has also gained much popularity amongst healthcare data processors as it enables them to manage this influx of data at manageable costs (He et al., 2015).

The goal of EHRs is that each patient has only one record of their health accessible wherever to whoever has authority to access it. Such an arrangement, however, poses a lot of complications which we going to discuss in the next sections.

As EHR systems grow bigger and bigger across organisational, provincial and even national borders, interoperability becomes a serious concern (Appel, 2008). Interoperability refers to the ability of different systems to be able to exchange and share information. Currently, standardisation of healthcare systems is a still goal yet to be achieved (Keen et al., 2012; Studeny & Coustasse, 2014), therefore many healthcare facilities have different information systems based on their different information needs and standards. In some cases where these systems go across national borders, the legal requirements of these systems also evolve which poses the problem of legal interoperability (European Commission, 2014).

Another concern over having such large centralised EHR systems is the increased potential for breaches of privacy and confidentiality. Medical and healthcare providers have experienced 767 security breaches resulting in the compromised confidential health information of over 23 million patients during the period of 2006-2012 in the United States (Privacy Rights Clearinghouse, 2016). In large EHR systems more and more people have access to the EHRs in concern which makes it more complex to track who accesses the information and for what purposes (Shaikh et al., 2011). This also increases the possibility of breaches of privacy and abuse of these records exponentially (Terry, 2013). Accountability, if a breach or something else that is undesirable occurs, is also difficult to obtain (Janssen et al., 2012; World Health Organisation, 2008). In the case of court cases, there have been instances where patient information has been disclosed under court orders, at times even without the knowledge of the patient. Even when the culprits of such breaches are known, prosecuting them is difficult because there are no laws to enforce punishment e.g. in the case where an offender is in another state where what their offence is, is not legally mandated. It is, therefore, evident that there is a need for clear articulation of how EHR systems are going to address the issue of patient information security as current policies are not reflective of this strategy (Pear, 2007; Hripcsak et al., 2014; Bayer et al., 2015).

It has been noted by Castillo et al. (2010) that during the implementation of EHRs system, the workflow of the caregivers can be drastically changed and impacted in a negative way

(Castillo et al., 2010) which in turn leads to a reduction in practice productivity (Gans et al., 2014). In some cases, the changes are minimal though in some cases these changes are considerable. For the mere fact that a new element has been introduced to the system, some change will have to occur. This change to the workflow has been noted to increase the stress levels of physicians (Babbott et al., 2014). This is a very serious side effect of EHRs and their implementation. It cannot be ignored especially when you consider that the experience of the patient influences the well-being of the healthcare professional (Sinsky et al., 2014). A decline in job satisfaction for the healthcare professional may also be a result of such practice and workflow transformations which have a higher demand on the cognitive abilities of the healthcare giver as well as resulting in a time increase (Babbott et al., 2014).

Finally, since EHRs rely heavily on the technology they are based on, any kind of technology failure poses huge risks. If a power outage could occur and these EHRs become inaccessible for various reasons, inability to provide healthcare can occur (Terhune, 2012). This is a problem that developing countries are more prone to than developed countries since they are typically under-resourced. Natural disasters have been known to disrupt technology centres and the risks of technology failure are many (Mata-lima et al., 2013; European Commission, 2016). EHR implementers will have to take measures to prevent such technological mishaps.

2.1.4 Patient Centred Design of EHR

Patient centred design is a design principle that is now being used in the development of EHRs. Its key focus is improving the experience that the patient gets during care by adding value for the patient through technology (Sinsky et al., 2014; Pagliari et al., 2012). This principle is hinged on the recognition that technologies, in this cases EHRs, have the potential of diminishing the experience of the patient. This is often because of the implementation process which is usually beguiled with a lot of hiccups.

For this value addition to be achieved, there is a need for a set of policies that will govern and regulate how EHRs are to be implemented without impeding the positive experience of the patient. Patient engagement should not be lost because of the introduction of technology in care practices as it plays a vital part in enhancing the positive experiences of the patient (Frank et al., 2015), hence the need to consider this aspect during the design and implementation of EHRs. Another aspect that contributes to a pleasant patient experience is a sense that one's personal information is being kept private and confidential. On the contrary, the introduction of EHRs has invoked the opposite. There is, therefore, a need for

policy to address not only these but other aspects and factors that contribute to a pleasant patient experience.

Another important consideration to make is for the patient to have a satisfactory experience of the care they receive and in addition the well-being of the carer or healthcare professional has to also be considered (Sinsky et al., 2014). Therefore measures have to be taken that in the use of EHRs the healthcare professional is not negatively affected in using them. Healthcare professionals have been noted not to be the most technologically savvy people (Babbott et al., 2014). Care has to be taken that they are “not overwhelmed or distracted by EHR-associated tasks” (Sinsky et al., 2014: pg 727). Usually, the electronic workflows associated with EHRs are not aligned to the clinical tasks that the carer would be engaged and could, therefore, be seen as adding to the tasks that they already have. Adequate training on how to use EHRs is then a requirement before implementation of EHRs otherwise extra personnel should be hired to manage this overhead on the healthcare professional.

2.1.5 Meaningful Use of Electronic Records

Adoption is increasingly not the main issue as far as EHRs are concerned as many healthcare facilities are adopting the use EHRs (DesRoches et al., 2013). A more pressing goal now is to achieve meaningful use of EHRs (Blumenthal & Tavenner, 2010). Meaningful use of EHRs refers to their use by healthcare providers to achieve specified meaningful achievements in improving care (ibid). Metrics have to be set to measure meaningful use and these vary depending on the authority assessing this meaningful use. Despite the increased adoption of EHRs, on the other hand, meaningful use is not yet being achieved by many healthcare institutions (DesRoches et al., 2013). This is especially true for smaller institutions. Development of EHR systems will, therefore, have to be coupled with meaningful use objectives to ensure that they are properly utilised. Meaningful use is being promoted strongly in the United States through the Health Information Technology for Economic and Clinical Health (HITECH) Act of 2009, where practising clinicians and institutions are rewarded financially by the government for meeting meaningful use criteria (Iqbal et al., 2013).

Despite the good intentions of this policy on the meaningful use of EHRs, its deployment has however generated a rather negative side effect. Because of the need to achieve the goals of meaningful use as quickly as is possible, some critical design aspects of computerised health records are being side-lined (Studeny & Coustasse, 2014). One of these aspects is

system interoperability. Despite that Studeny and Coustasses research was based on PHRs, their findings can be applied to EHRs because of the many similarities that these two patient data types share.

2.1.6 Personal Health Records

The phrase Personal Health Records (PHRs) definition is something that has not yet been fully agreed upon (Pagliari et al., 2012). The key feature on PHRs on and on top of the EHRs is patient access to the information and or system. Issues of concern are the source and controller of the information, operational platform as well as software functionality (ibid). There are many definitions of PHRs however, the one that blends well with the view of this research is the one which articulates PHRs as electronic applications by which individuals can access, manage and share their health information, and that of others for whom they are authorized, in a private, secure, and confidential environment (Markle Foundation, 2006; Studeny & Coustasse, 2014).

Of the seven billion people living on planet earth today, six billion use mobile phones (World Economic Forum, 2013). These coupled with a plethora of medical sensors, also known as Medical Body Area Network (MBAN) sensors (Wang et al., 2015), self-tracking devices, connected devices and people themselves are generating data relevant to health on a daily basis (World Economic Forum, 2012; Vayena et al., 2015). More and more data about people is being collected passively and actively (University of California, 2014) and all this data is contributing to the amount of personal health records about individuals. “Use of personal data in innovative ways in the health sector can yield significant economic and social value for all stakeholders.” (The Boston Consulting Group, 2013: pg 9). However, most if not all the forms of personal health data are currently not part of the mainstream traditional healthcare data used for research (University of California, 2014). This is because perhaps foremost of the way that it is generated (Kearney, 2014). Its quality is also questionable (University of California, 2014) and its storage, location and ownership as well privacy issues are at the moment too complex to properly define. Yet this complexity should not mean inaction (Kearney, 2014), new approaches are needed to make sure that personal health data utilisation can be increased (Hripcsak et al., 2014).

Terry (2013) says that patients are soon going to become the main source of medical information in big health data pools. Data about patient behaviour and their sentiments are soon becoming important aspects of healthcare data. Initiatives like the Blue Button initiative (Fridsma & Lygeia, 2012), are fast increasing in popularity as a result of the meaningful use

criteria in the HITECH. This is seeing an increase in patient participation and engagement in the creation of healthcare records. The role of PHRs in big data is therefore set to increase and will not be ignorable anymore in the future.

PHRs have improved the way on how healthcare is provided in a number of ways (Pagliari et al., 2012) with benefits for both the providers and patients. For the providers, PHRs have seen improved communication with patients which has in turn improved trust between the two parties. This is a good thing as it is also in line with the patient centred notion of healthcare. Patient enrolment has also been positively impacted by the introduction of PHR. For the patients, there has been a marked patient satisfaction with their health facilities after PHRs were introduced. Patients feel more comfortable and relaxed around their physicians. Most feel that transparency and trust have also improved. It has helped them to take a more positive role in ensuring their own health and well-being. However, some ills are reported in PHR implementation which the next section will look at.

Service integration, workload imbalances and privacy concerns are some of the aspects that need to be resolved when PHRs are introduced (Pagliari et al., 2012). The smooth integration of PHR components into the existing systems is not a trivial matter as it requires some changes to the normal operating routines of the care providers. Workload and workflow issues have been noted to arise after the introduction of PHR systems and there is a need for managing this change in such a way so as to alleviate these imbalances (ibid). Another rising issue is on maintaining the confidence of the patients that the privacy of their information is still kept. Despite the seriousness of these issues, providers seem to be confident they can be resolved.

The methods that are currently being used for the entry, validation, and display of PHRs make it difficult to adopt them for use in clinical practice (Kim & Johnson, 2002). Most still do not contain the basic information sets required for administering healthcare i.e. they are not comprehensive (Mxoli et al., 2014) and their accuracy is questionable (ibid). This means that despite the heightened excitement about how PHR will promote and increase the bounds of healthcare, their recognition as valuable informational resources has not yet been cemented. Regarding PHRs as primary sources for diagnoses and other clinical functions is therefore still a farfetched idea.

There is a need for the creation of interoperable PHRs systems that can talk with each other (Studený & Coustasse, 2014). The foundational principle behind PHRs which is the diversified creation and sharing demands that the systems that handle PHRs be highly

interoperable. From a big data perspective, these systems should achieve not only the basic level of interoperability but the highest level of semantic interoperability. This level of interoperability is essential if the use of PHRs in big data analytics is to be achieved. Promotion of interoperability between PHRs and EHRs and subsequently big health data are areas that still need more work. Studeny and Coustasse (2014) in an evaluation of the limitations affecting increased interoperability of PHRs cite legislation and policy support as lacking to this effect (Studeny & Coustasse, 2014; Mxoli et al., 2014).

By involving patients in the processes of creation and management of their own health data, co-creation of value is taking place (Pasquale, 2013). Patient engagement is improved (Frank et al., 2015) and this has the effect of improving the experience of the patients hence alleviating the stress that comes with diseases or the need for healthcare. Patients are in essence having a first-hand innovation in the products that they themselves will use. They essentially become prosumers i.e. both producers and consumers of their own health data. This increases the sense of ownership the patients have about their own data which in turn can lead to a better interpretation of that information ultimately improving patient-centric healthcare services.

Patient access to their health information has many benefits. It helps patients to better understand their own health and therefore they are empowered to make informed decisions about their own health. There is an increased transparency of the activities of the healthcare provider hence ensuring that the provider and the patient have the same understanding. Patients also have the ability to check the accuracy of their own information and if needed they can query discrepancies. The accuracy of information could, therefore, improve (Fridsma & Lygeia, 2012). PHRs also have the potential to reduce the amount of time lag between the outbreak of a trend and the time authorities react to that trend since patients will be contributing to their health records in real time (World Economic Forum, 2012; Vayena et al., 2015). Here lies the possibility of kerbing localised outbreaks, thereby averting full-blown epidemics and pandemics before they cause widespread carnage.

How PHRs are going to be integrated with the mainstream EHR systems that healthcare providers use is currently blurry (Pagliari et al., 2012). In the context of big data, the ideal scenario will be when EHRs and PHRs will be put in one pool, the data being blended so that it can all be used in data analytics. However, because of the differences in standards across many critical areas like data structure, data creation, ownership and so forth, such a blend of EHRs and PHRs is currently not practical. It can, therefore, be seen that there is a need for developing more guidelines on implementing PHRs (Mxoli et al., 2014) in such a

way as to better position their integration with EHRs. The same is true for EHRs. When guidelines are produced to this effect, PHRs deployment will see growth as its usefulness will also start to be appreciated by the EHRs community.

So far this review on personal health records has in general been in support of PHRs. PHRs, however, have serious setbacks to their adoption as they pose “substantially higher levels of risk regarding security, privacy, and confidentiality” (Terry, 2008: pg 220) than EHRs which ironically provide more benefits like completeness, quality, universal availability and better interoperability. These aspects will be further discussed in a later section titled Data Policies. The thrust of this document, however, is to emphasise how PHRs can be a major player in big health data and how this can be achieved by policy, which leads me to the following section on big data.

2.1.7 Summary of Patient Data

In summary, patient data has now evolved from its previous position of merely being a paper trail of past care episodes with patients. With the inception of electronic records and the realisation of their potential to improve healthcare, there is much talk about the need to treat patient data as a potential energy that could be harnessed for the patients. This school of thought has given birth to the concepts of patient centred care, meaningful use and personal health records, all founded on the potential secondary uses of patient data. Stakeholders in the healthcare sector need to adopt this approach relative to patient data.

2.2 Big Data

Since the inception of the world up until the year 2003, it is estimated that the world produced 5 quintillion bytes of digital data yet today it only takes the whole world two days to generate the same amount of data (Marchibroda, 2013). Even more worrying is that this figure is growing (Assunção et al., 2015) by 150% year on (Yiu, 2012b). These facts all allude to what the phenomenon of big data is today.

““Big data” refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyse.” (Manyika et al., 2011: pg 1). These large datasets, which are often disparate, require new technologies that provide management and processing capabilities (Peters & Buntrock, 2014) that are not readily available today. The above definition includes dependencies in its terminology. For instance how big datasets have to be, to be classified as big data is subjective (Manyika et al., 2011).

Peters and Buntrock (2014) further clarify what big data is and identify four characteristics that data has to have for it to be classified as Big Data which are commonly known as the

four Vs. These are volume, velocity, variety and veracity (ibid). Peters and Buntrocks definition sees an addition of the fourth V (veracity) to the original three Vs. (Laney, 2001) as defined by Laney in an article titled 3D Data Management: Controlling Data Volume, Velocity, and Variety, which is widely viewed as the genesis of the process of defining big data.

It is said that of all the data in the world today, 90% of it was generated in the last two years (van Riimenam, 2013; Laney, 2001). This means that huge amounts of data are being stored away and the sheer size of this data is what is referred to by the first V of big data, volume. "Velocity refers to the rate of change of data, often connected with the analysis of streaming information" (Peters & Buntrock, 2014: pg 207). Variety refers to the different and diverse forms of data both structured and unstructured (Bowden, 2014; Laney, 2001). Veracity refers to the uncertainty of the data and finally volume refers to the scale and size of data (Bowden, 2014).

However, as the big data concept is evolving, some authors have argued that there needs to be more V's to sufficiently define big data (van Riimenam, 2013). An additional three V's, variability, visualisation and value are now being added to the definition of big data by some experts, where variability refers to the contextualisation of data which means that the same data can have different meanings in different contexts.

Visualisation refers to how the data and its knowledge result sets are presented clearly. This is a worthy additional V, as it is important to be able to access the information visually so as to reach some findings. It was noted that companies that use visualisation tools or visual data discovery mechanisms like Tableau Software have higher big data meaningful use throughput with less IT personnel involvement (Jewett, 2014). Visual analysis also helps in easily changing the data being looked at as well as the way it is being looked at (ibid).

Lastly, value, which to me is the most important of all the V's, refers to the gain that can be associated with that data if fully utilised. This last V is in essence what big data is all about, value, which can be in the form of money, lives, well-being, better health or simply put peace.

The diagrammatic visual below (Figure 2-1) illustrates in a very simplified way the ideal concept of big data. Individuals, the public sector and government enterprises, as well as the private sector, are all generating information which is being collated in various massive repositories strategically named data commons (World Economic Forum, 2013). Using powerful data analytics and mining tools, these data commons will then be gleaned of their

forecast potential wisdom which will then result in informed and wise decision making. Such is the future of the information world! Great promises are made by big data for a better life in general. This relationship between the information generators, the information containers, the information treatment and the results is termed the Data Ecosystem (ibid).

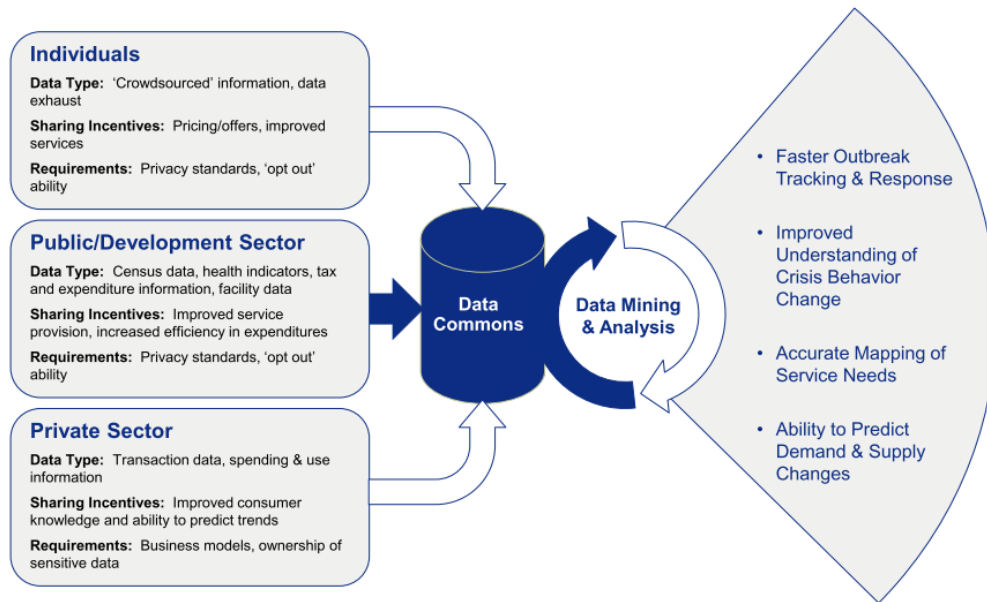


Figure 2-1: Data Ecosystem
(Adapted from World Economic Forum 2012)

To achieve such an ecosystem is a daunting task. Such an ecosystem will require an environment where information privacy and security concerns would have been fully addressed and the people are also willing to share their data. For such data commons to be useful, the information within them should also be identifiable i.e. its origins and belonging should be available. One other requirement of such an ecosystem is a rich kind of human capital that can be able to fully harness the benefits of this kind of data. There is no evidence in the literature that such an environment does exist and can only be achieved when government's take the first initiative to create such open data environments (World Economic Forum, 2012).

2.2.1 Big Data in Business

In today's business world gaining a competitive advantage over your competitors is key for business success unless you're a monopoly organisation. Constantly businesses are trying to find new innovative ways of surviving and big data has been identified as a potential competitive advantage gap providing value for the organisation (Assunção et al., 2015).

There has been a realisation in the business world, of the role that information will now be playing in business success (Schroeder & Schroeder, 2016) and most organisations are trying to position themselves to benefit from this data-driven landscape. Customer profiling among other big data levers has been taken up very heavily by most retailers and other business sectors and this has the potential of generating valuable insights (Damle, 2014).

“In the coming years, the continued adoption and development of big data levers have the potential to further increase sector-wide productivity by at least 0.5 percent a year through 2020. Among individual firms, these levers could increase operating margins by more than 60 percent for those pioneers that maximise their use of big data.” (Manyika et al., 2011: pg 64).

The figures presented above show the massive potential big data has in increasing the profitability of businesses. The retail industry has been identified as the most obvious beneficiary of big data (EMarketer, 2012). However, it is not all business sectors that are set to benefit directly from big data analytics.

Market segmentation has been identified as a field that is now being greatly influenced by big data analytics (Terry, 2013). The behaviour of people through internet surveillance is providing marketing researchers with very rich data about the interests, likes, dislikes and fancies. The digital footprint is, therefore, making precise target marketing a reality. The commercial gain involved here is huge (Terry, 2013; Manyika et al., 2011; World Economic Forum, 2012).

Customer behaviour is one other area that business is now using big data to understand. Almost all retail shops use loyalty cards and the data that these generate is very rich (Jewett, 2014). By analysing information from loyalty cards, organisations like Unilever are able to produce products for the global market and are able to anticipate the demand for popular products and news buying trends. This information can be the difference between gaining an advantage over your competitors and losing out on of revenue.

It can be seen that the business sector has adopted the data sciences already and is already benefiting at some level from the data economy. There is yet to be seen massive paradigm changes affected by big data but on a small scale this has already started to happen (Schroeder & Schroeder, 2016). However, various challenges have also been identified that

will need to be addressed to reach the envisioned data use climax (ibid). The discussion of these challenges is beyond the scope of this study.

2.2.2 Big Data in Healthcare

The shift to EHRs has seen a lot of information that is health related being collected in large volumes (Uller et al., 2007) setting the foundation for Big Health Data. However, can Big Data be a concept applicable to healthcare data? Addressing this question Ohno-Machado says, “Big Data is a big deal in biomedical research and healthcare.” (Ohno-Machado, 2014: pg 193). Peters and Buntrock (2014) go on to say that “Big Data, across millions of patients, offers the opportunity for large-scale population analytics and research.”(Peters & Buntrock, 2014: pg 208). There is a lot of excitement about the prospects of Big Data in healthcare systems across the globe (Keen et al., 2013; Damle, 2014). These large EHR datasets can be comprehensively analysed by researchers in observational studies including millions of patients (Hoffman & Podgurski, 2013a) and the potential of such studies is huge (Keen et al., 2012).

The benefits of Big Health Data could amount in monetary terms to hundreds of billions of dollars annually (Manyika et al., 2011). Some of the other benefits of this kind of data exploitation could be reduced life losses (Page, 2014), increased monitoring of infectious diseases, disease outbreaks and chronic illnesses (Hoffman & Podgurski, 2013a; Keen et al., 2013) and it could also enhance decision making to be based on Evidence Based Practices (World Health Organisation, 2008; Anon, n.d.).

Terry (2013) once argued that all information is healthcare related information (Terry, 2013), and this is one concept that big health data analytics can benefit immensely from. For example the surname of a person can be associated in a medical database with their families history for a particular medical abnormality, a supermarket loyalty card can record the purchase of a pregnancy testing kit, a bus card system can record a drop-off at a cancer specialisation clinic, a social media site e.g. Facebook, can record a “like” of a picture advertising a TB drug, and a web monitoring system can record excess times spent on a heart surgery page. The list is endless and proves how all information can be linked to the health of an individual. What differs is just the strength of relevance otherwise, all information is health related.

Vayena et.al (2015) identified various ethical risks that are associated with the adoption of big data in healthcare. They argue that contextual sensitivity, the application of big data

methodologies and the legitimacy of big data technologies in health care, all pose ethical incongruences that need to be examined and properly anticipated for. Further complicating this ethical risk is its ability to develop imperceptibly with the proponents or culprits not being aware of these ethical trespasses. Big health data will, therefore, need to develop this ethical eye beforehand and enunciate how these ethical challenges will be overcome (ibid).

The complexities that are involved in big health data analytics can then be seen to be enormous (Kankanhalli et al., 2016). Real big health data analytics will have to somehow incorporate all kinds of information. Here lies the challenge of creating an environment that will reflect all input yet at the same time observe the strict regulations healthcare data has to adhere to. How this will be achieved is still a big mystery and Terry (2013) suggests that the regulation of big data will increasingly become more challenging.

2.2.3 Summary of Big Data

The concepts of big data have been in existence for some time now but only until recently has healthcare adopted its use. By definition, big data is a combination of many factors namely, the seven V's of big data and resource capabilities that surround the data. It can be concluded that health data is now at a level where it can be treated as big data. Unlike in business, various hurdles need to be overcome so as to realise tangible applications of big data in healthcare. Only after these have been resolved can expectations of big health data be realised.

2.3 Data Policies

Data policies refer to a set of documentation that outlines the guidelines for ensuring the proper management of an organisation's digital information (Torjman, 2005; Cole, 2016). Such guidelines often involve how data is captured, stored, processed and or eventually discarded. It is data policies that generally shape the information culture of an organisation, therefore, it is imperative that they reflect and support the vision of the organisation. In a healthcare context where that vision can generally be described as bettering healthcare, data policies should be designed to support this core function of healthcare (Pasquale, 2013; Ohno-Machado, 2014). The following sections will discuss various aspects of data policies in the healthcare sector.

2.3.1 Nature of Current Data Policies

Key to the success of exploiting healthcare data are the information policies that dictate how to access and use this information (Keen et al., 2013; Berndt et al., 2001; Vital, n.d.). Foremost in healthcare data policy concerns is the issue of confidentiality (Thilakanathan et al., 2014). Because of the highly intimate nature of healthcare information, current policies around healthcare information are very protective and defensive of that information (Keen et al., 2012; Pasquale, 2013; Terry, 2013). This controlling nature of the current data policies has the potential of hampering the opportunities presented by Big Health Data (Hoffman & Podgurski, 2013a; Pasquale, 2013). According to the CEO of Google (Page, 2014), 100 000 lives are being lost each and every year because of such kind of restrictions and policies around healthcare data, however, this is a moot point. Further aggravating this problem is the fact there is so much personal information about people on the internet such that efforts to strip identifying elements of the data are not effective (Keen et al., 2013).

2.3.2 Data Policies on Electronic Health Records

The Health Information Technology for Economic and Clinical Health (HITECH) Act which was passed in the US in 2009, is one of the most significant policies driving the changes and innovations that are taking place in the health information systems front (Adler-milstein et al., 2014). It is in this act that incentives to providers were made available if they managed to achieve meaningful use of EHRs as well as visible improvements on quality of care, safety efficiency, cost and care coordination (Pasquale, 2013; Kankanhalli et al., 2016). It also advocated for a standardised way of certifying EHR systems. This was a welcome incorporation as there was a void in the regulations in addressing certification, which was leading to a plethora of EHR systems which were potentially unsafe finding their way into the market (Hoffman & Podgurski, 2011). After it was passed there was a sharp increase in the number of providers implementing EHR technologies to qualify for these incentives (DesRoches et al., 2013). The results were obvious; increasing volumes of digital healthcare data being collected and stored.

The Affordable Care Act (ACA) passed in 2010 is another driving policing in healthcare information systems innovation. Through it, there is now seen a common need of having health information systems that are standardised, interoperable and cost driven yet maintaining the quality of care (Adler-milstein et al., 2014). It is seen here that the days of having a healthcare service provider being managed solely by a physician are numbered. Successful health enterprises will need management that is robust, agile, with the ability to make not only decisions in the surgeon room but also from a business and IT perspective.

Another set of legislation that has influenced the shape and direction of health data is the Health Insurance Portability and Accountability Act of 1996 (Terry, 2013). This act is divided into two sections with the first section seeking to protect people who would have switched or lost their jobs by ensuring that they continue receiving medical insurance. The second section deals mostly with administrative efficiencies like standardisation of healthcare information's systems and ensuring confidentiality is highly upheld in the handling of health-related information. Its primary goal is summarised as "to make it easier for people to keep health insurance, protect the confidentiality and security of healthcare information and help the healthcare industry control administrative costs" (Tennessee Department of Health, 2016:pg: 1). Nothing in the laws goal speaks to improved data access and utilisation. On the contrary, these aspects (access and utilisation) are made cumbersome. Terry argues for this point and says that there is no current policy nor regulation that emphasises a balance between patient decisions of how their data can be used in research initiatives (Terry, 2013).

In South Africa, the Protection of Personal Information Act (POPI) (Parliament of the Republic of South Africa, 2013) and the Promotion of Access to Information Act (PAIA) (South African Government, 2000) are two major national policy documents that are shaping the way healthcare information is being accessed and used. Many institutions are finding their operational requirements at the back-foot and having to adjust their operations to not find themselves at the wrong side of these bills (de Bruyn, 2014). When both bills are juxtaposed one seems to be seeking a platform for better access (PAIA) whilst the other (POPI) seems to be seeking to control and balance the environment of this access and use. How other bills are relating to these two and the impact they are having on data utilisation are subjects that still need to be studied. The National Health Act (South African Government, 2004) is one other bill that seeks to determine the direction of healthcare and subsequently that of EHRs.

Like any other project, an evaluation of the successes and failures of EHR systems is crucial. To assess the financial success of EHRs, the return on investment (ROI) on EHRs has to be evaluated. There is a lack of policy structure to ensure standardised evaluation of ROI across EHR systems (Adler-milstein et al., 2014). Common assessment criteria, therefore, need to be identified as well standardised metrics for the evaluation. This is an issue that needs addressing through policies.

2.3.3 Policy Models in Electronic Health Records

In trying to achieve privacy and confidentiality protection of data, Terry (2013) suggests that there are six models (Terry, 2013: pg 14) that policymakers use to design policies. The following section will have a look at these models in order to ascertain their ideological bearings.

- i) Limiting data collection – this model seeks to prohibit the collection of data of a certain type.
- ii) Regulating the disclosure or distribution of collected information – this model seeks to monitor and restrict sharing of information where necessary.
- iii) Breach notification – here whenever a regulation has been violated all stakeholders or parties involved should be notified accordingly.
- iv) Limiting processing of collected data – here policies seek to restrict the use of healthcare data for secondary uses.
- v) Devaluation of data – this model emphasises the de-identification of data being collected i.e. all elements that can enable the identification of the patient or entity involved should be avoided.
- vi) Data Security – these set of policies seek to ensure that the collected data is only accessed by permitted users.

It is either one or more of these models that are applied in policy design to ensure that privacy and confidentiality metrics are met. Most of them can be said to make data collection and sharing cumbersome and neither of them promotes information access. All have a permeating character of being protective of the data in concern. In the age of big data, such a strong protective nature will have to fade away or be loosened. Terry (2013) however has a different opinion as data protection is concerned. He concludes that the tougher measures on data collection and privacy controls are to be the way forward. If his stance will be adopted by policymakers big data prospects will continue to be greatly hampered (Menius & Rousculp, 2014; Hripcsak et al., 2014; Pasquale, 2013).

Commenting on collection based models of regulation, those that prohibit collection, Terry (2013) argues that they will not be sufficient to address the problem in concern as people will always try to find a loophole in the system to achieve their required outcomes. De-identification on its own is becoming difficult as re-identification is becoming easier and easier with time. Also, there can only be a limit to the amount of de-identification that can be done before data becomes useless as Ohm suggests that when data becomes perfectly anonymous its usefulness ceases to exist (Ohm, 2010).

Processing based models of regulation, those that regulate access and use, are also not well suited to address the privacy issue because they are prone to law technicalities and function creep (ibid).

2.3.4 Data Policies on Personal Health Data

“There is a need for a fundamental re-working of the legal model applicable to all electronic health records” (Terry, 2008:pg 1) especially in the case of PHRs (Pagliari et al., 2012). Personal health data as a subset of EHRs is generating a lot of talk and dialogue around its direction and use (Kearney, 2014). Adoption of this technology has been slow though it is expected to pick up soon because of the meaningful use incentives stipulated in the HITECH act (Wynia et al., 2011). One of the criteria of meaningful use is giving patients more access to their health information. This has increased the hype around PHR and has the potential to increase the adoption of PHR.

From policy makers viewpoint this is in line with their vision of seeing technology increase the engagement of patients in their own health (Wynia et al., 2011). However, there is a need for policy makers to understand physicians views and perspective on adopting PHRs before policies can be developed. It has to be taken into account that different physicians will approach PHR use in different ways, depending on various factors ranging from social, biological and practice-related factors. There is a need for physicians to be educated and convinced of the potential benefits of PHR as most do not as yet see it. Most physicians currently are sceptical of using PHR for administering healthcare with most fearing data inaccuracies in these records. These viewpoints of the provider should be treated as critical input to any policy formulation that would be made around PHRs. They cannot be ignored (ibid).

Personal health data is fast becoming a major contributor to big health data. Its volumes are on the increase and it is increasingly becoming difficult to manage trust issues around its use (Kearney, 2014; Hripcsak et al., 2014). Individuals need to be abreast on how all this data about themselves is being used (Kearney, 2014). It is essential for the public to buy into the idea of their data being used for secondary uses and policies need to recognise this and support it (Hripcsak et al., 2014; Keen et al., 2013). However, a balance has to be struck between maintaining patient privacy and fruition of personal health data exploits.

Key to bringing patients to this understanding is the context of the use of the data. Defining policies that will uphold privacy and security of individuals yet promoting data use will have

to disclose the context of use (Kearney, 2014). This has been proven to make individuals more comfortable about their data being exploited when they know what it is being used for (Kearney, 2014). Creation of personal health data is also so diverse that there is a need for policies to manage personal health data complexities like the flow of data and transparency (Kearney, 2014).

2.3.5 Data Policies on Data Rights and Ownership

Another data policy issue around Big Health Data is ownership of the information in these datasets. Whether it is the government or the individual patient or the organisation where the data resides that have the rights and ownership to that information is an issue that is blurred in many health information systems (Menius Jr & Rousculp, 2014; Evans, 2011). If ever this information is to be used consent of the owner will have to be acquired, however, who the owner of the information is, is a moot point.

The argument that the patient should be the owner of the data comes from a good motif. There is a belief that when the patient has ownership to their health data, there will be fewer instances of it being accessed without the consent of the owner thereby reducing privacy and confidentiality breaches (Evans, 2011). Proponents of patient data ownership also want patients to have the power to stop data use if consent will not have been given. However, Evans (ibid), continues to argue that this is not the case as non-consensual access to that data will continue happening because the patients do not have the appropriate means and power to enforce consensual access. This power still resides with the data stewards which are the healthcare organisations and government. Furthermore, current federal laws as in the case of the United States do allow non-consensual access to patient owned information (ibid).

Another point that pro-ownership proponents seem to have ignored is that “having a property right does not ensure property-rule protection” (Evans, 2011: pg 79). What this means, in short, is that rights to a property or information can be taken away at any time by the government or courts under the banner of public protection. What this then means is that the argument for ownership will not resolve privacy and confidentiality issues. Some other solution is needed here which does not restrict information access by owner consent.

It is important also when talking about data ownership and consent for its use to differentiate between the primary uses and the secondary uses of healthcare data (Keen et al., 2012). Primary use refers to the use of that information in treating the patient and all other uses are

secondary use i.e. planning services, managing budgets, accounting for resources used and research (Laurie & Sethi, 2011). These issues about ownership and consent do not affect the primary use of healthcare data. The concerns of this paper fall under secondary uses, therefore, it is important to consider these issues.

2.3.6 Data Policies and Data Analytics

There is also an argument that Damle (2014) supports that despite the analytical opportunities that are presented by Big Health Data, there seems to be a mismatch between the process in which current data analytical packages generate their insights and the science of medicine (Hoffman & Podgurski, 2013b). It is not a clear cut process to infer causal effects in healthcare as current data analytical software's generally do (Hoffman & Podgurski, 2013a). This is because many other factors that are not visible to the naked eye like climate, pollution, social aspects and the geography of an area also have an impact on the health of one (Thacker et al., 2012). Great care has to be taken about how Big Health Data will be analysed to produce pointers that are relevant and usable. An example of how things can go wrong in Big Health Data is the Google Flu Trends (GFT) project which has not produced any trustworthy insights since its inception (Lazer et al., 2014). An alignment of data analytics and healthcare procedures and policies is therefore needed otherwise healthcare analytics efforts and innovations will continue to be irrelevant to medical science (Damle, 2014; Herasevich et al., 2010; Hoffman & Podgurski, 2013a).

2.3.7 Data Policies on Big Data

There are a number of policy considerations that will have to be taken if ever the full value of big data is to be realised both in the private sector and the public sector (Jewett, 2014; Schroeder & Schroeder, 2016). Before diving into this maze of big data, one needs to be sure that they are not stepping on other people's toes i.e. individual people, organisations and governments. Information protection is a serious thing and of late many laws have been published addressing information use. So the big data user will have to be familiar with the laws that govern their data environment. These policies vary by region and practice.

Establishing strong governance has been noted as a major challenge facing big data value creation (Marchibroda, 2013; Schroeder & Schroeder, 2016). There is already a massive shortage of qualified personnel who are suitably qualified to implement big data successfully (TechAmerica, 2014; Manyika et al., 2011) under the current information governance policies. It is also foreseen that the gap between analytical and managerial positions will

continue to grow and this is one area that big data policy will have to address immediately (TechAmerica, 2014).

Another key area that big data policy will have to address is public concern about big data (Hamburger & Tsukayama, 2014) as there are serious concerns about how big data will be used especially personal data. Most of these concerns revolve around privacy and confidentiality issues, and closure on what data collected will be used for (Hamburger & Tsukayama, 2014; TechAmerica, 2014). All the above policy issues are being identified as things yet to be done therefore it would be worthwhile for any party interested in big data to take these issues into consideration.

2.3.8 A Big Health Data Policy Theoretical Framework

Research in big data policies is relatively limited (Manyika et al., 2011; European Commission, 2014). This is because the topic in itself is relatively new (Vayena et al., 2015). There is, therefore, a growing need for research on this subject (European Commission, 2014) to substantiate the many ideas and hypotheses that have become synonymous with big data.

A study on the creation of a theoretical framework that can assist in the design and development of big data healthcare policies could not been done in literature. This is evident because of the lack of a framework that allows measurement of policies to ideal characteristics of a big health data environment. It is in this light that the study will propose a theoretical framework that will be used to validate and substantiate the findings and the results of this study.

The following components or variables have been identified, through literature analysis as key factors that contribute to the presumed effect:

- i) Recognition and development of a Big Health Data skillset (Manyika et al., 2011; European Commission, 2014).
- ii) Regulation of health data that leans towards ease of access and sharing as much as possible, yet maintaining acceptable levels of information security (ibid).
- iii) More power to the users (patient) to create and manage their health data and information in an acceptable manner (Clarke & Meiris, 2006; Pasquale, 2013).
- iv) Strong emphasis on the use of data across all levels of users (Blumenthal & Tavenner, 2010; DesRoches et al., 2013).

As such, the foundational and theoretically ideal big health data environment would be one where all these four spheres are intentionally addressed and reflected. The intentional aspect would be visible in clearly outlined guidelines, procedures and policies that would seek to accentuate these four spheres. The proposed theoretical framework, therefore, can be represented as shown in the figure below.

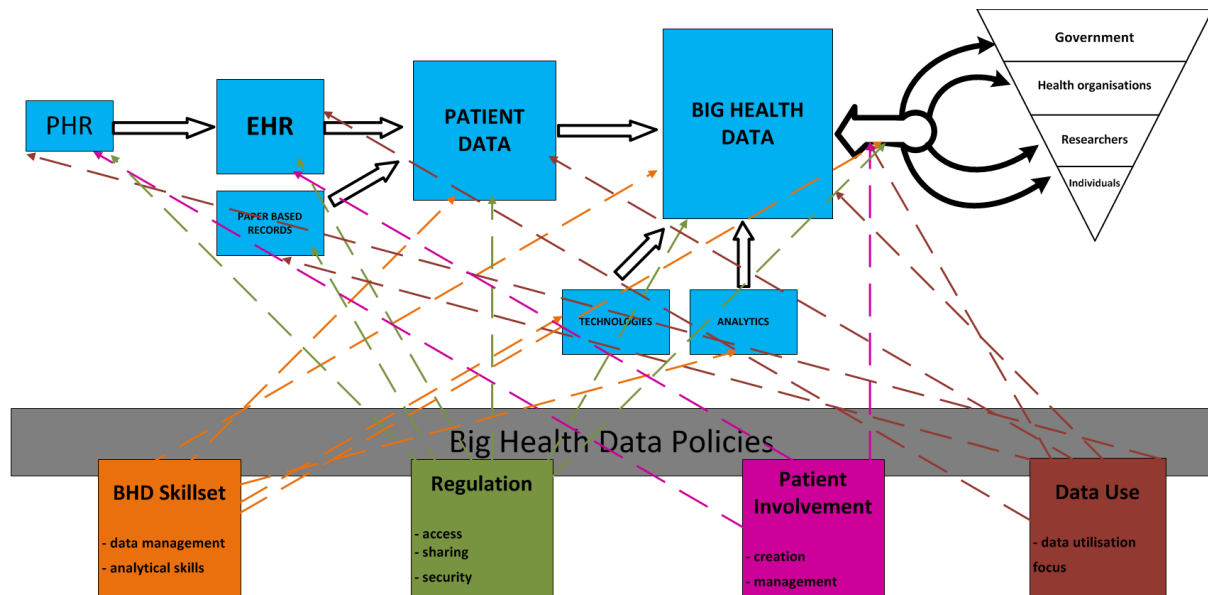


Figure 2-2: Big Health Data Policy Theoretical Framework

The diagram shows the normal big health data environment with its various touch points as being supported by a foundation of the four pillars identified above. Each pillar shows the interaction between it and the various components that feed into the bigger picture of big health data. In order to visualise the proposed framework better, the following section looks at each of the four pillars and their perceived interaction with the big health data components.

Figure 2-3 shows how important a health-focused niche of data analytics and data management human resources is. Health data is more complicated than normal business data. It has its own interpretation, inference and propensity rules that are not the same as of other data (European Commission, 2014). Furthermore, there are the common issues of privacy and confidentiality (Bayer et al., 2015), that are at another level other than that of other data. This, therefore, calls for a data management skillset that is well vexed with the ditches, humps and potholes of health data management.

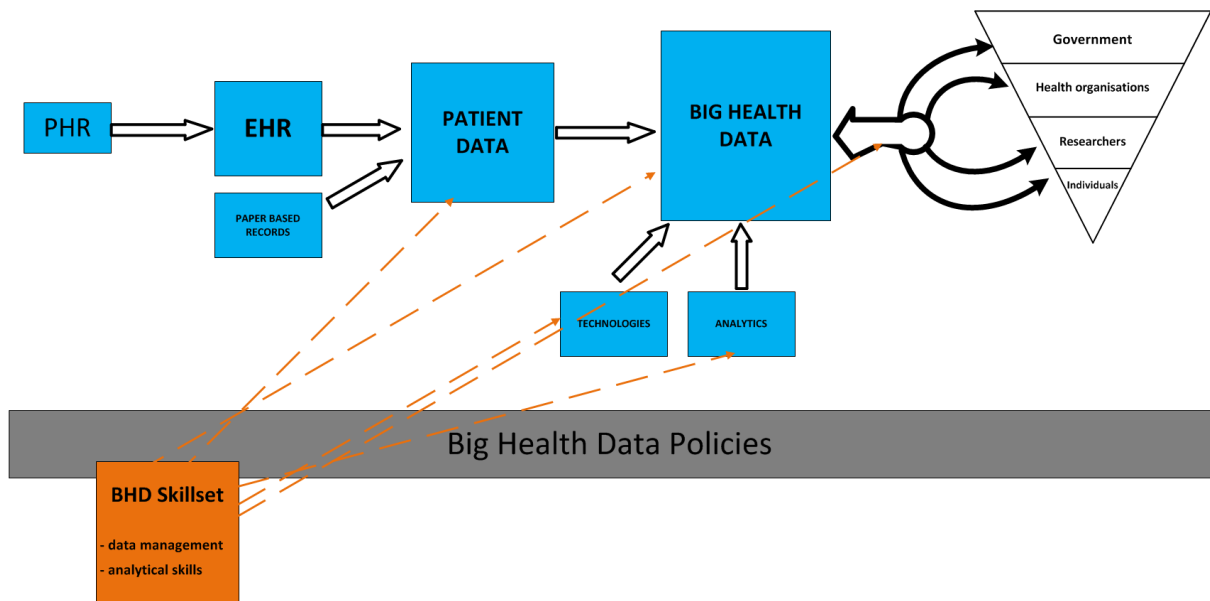


Figure 2-3: Big Health Data Theoretical Framework: Skillset requirements

It will be imperative then for policy platforms to address this need of ensuring that the right kind of sophistication for health data analysis is acquired at all the levels as shown in the diagram.

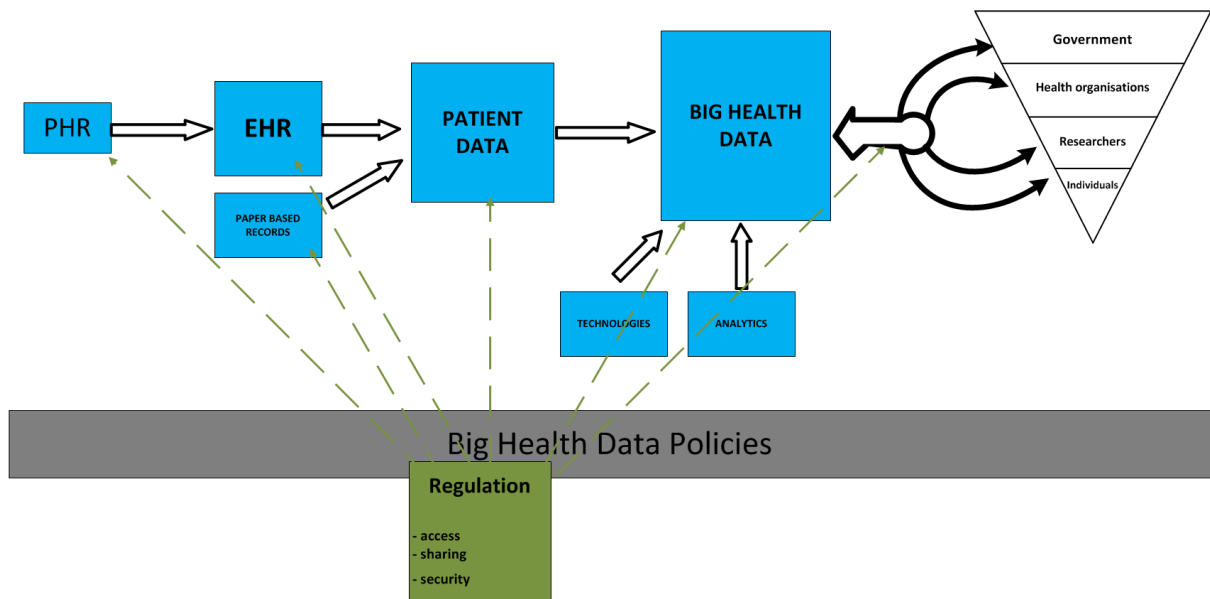


Figure 2-4: Big Health Data Theoretical Framework: Regulation requirements

Figure 2-4 shows the importance of regulatory structures at all the points shown in the diagram. These regulatory structures should speak to all the six W's (Who, When, Where, What, Why, How) of data access, sharing and its security. These regulations should be in a format that is accessible and understandable to all levels. An important component of these

regulations will be enforcing accountability of data access, data sharing, and data security to ensure that data is being used for the right purposes (European Commission, 2014). Accountability here will include laying out the measures, such as legal action, that would be meted to those who violate their undertakings.

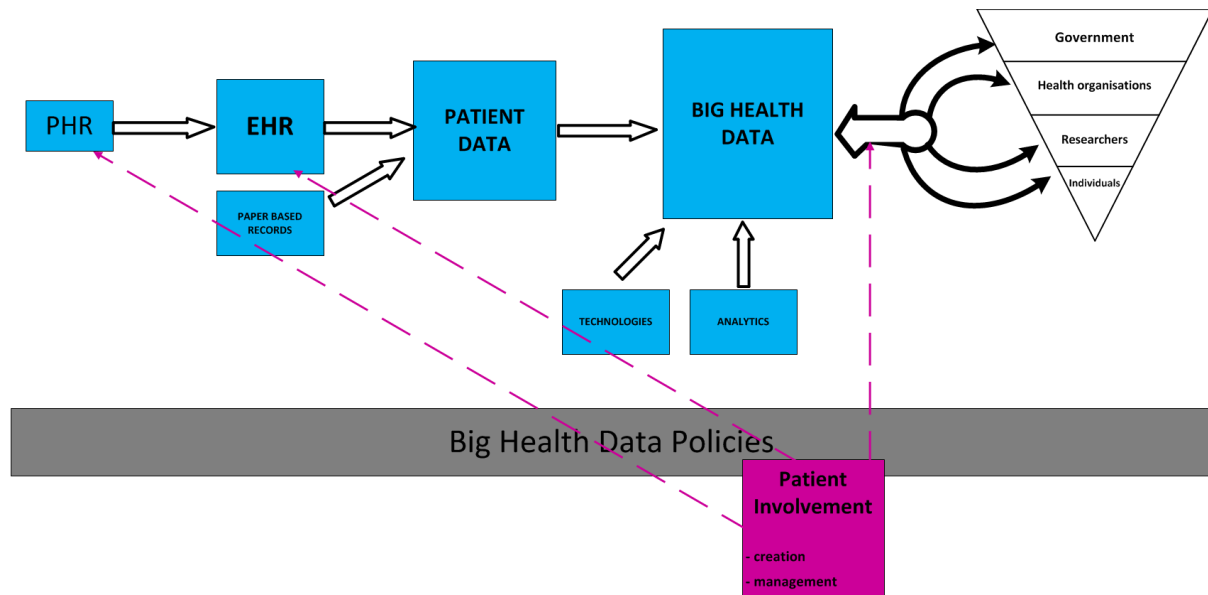


Figure 2-5: Big Health Data Theoretical Framework: Patient Involvement requirements

Figure 2-5 shows the touch-points where patient participation should be perceptible. If not then big health policies should promote this participation. Two key areas where this participation is paramount are patient data creation and patient data management. Personal Health Records and their subsequent proper implementation could be appropriate plugins to ensure that this aspect is sufficiently recognised (Genitsaridi et al., 2013).

The final pillar of the proposed big health data theoretical framework is about promoting data utilisation (World Economic Forum, 2013) and not only data utilisation but meaningful data utilisation (Iqbal et al., 2013). It is evident that the perception of the role that data can play in the various successes has to change. Health organisations still need to learn the value of their information and to use it to promote effectiveness, economics as well as better patient outcomes. This is important in this era of the information age. Where this is not common knowledge, policy should make it common knowledge.

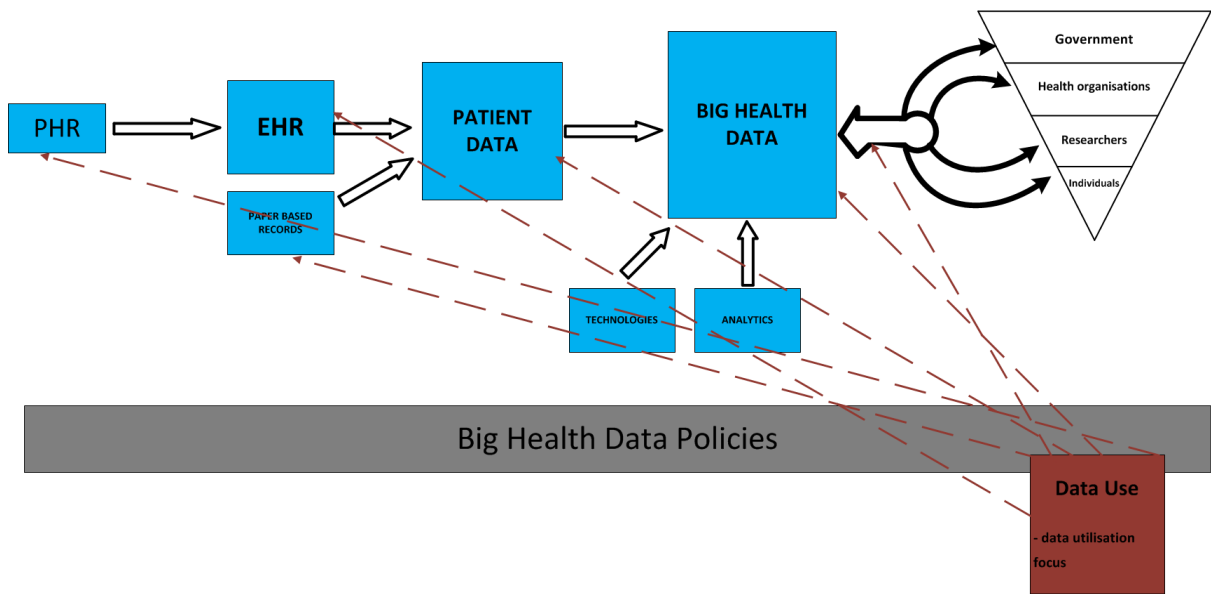


Figure 2-6: Big Health Data Theoretical Framework: Data Use requirements

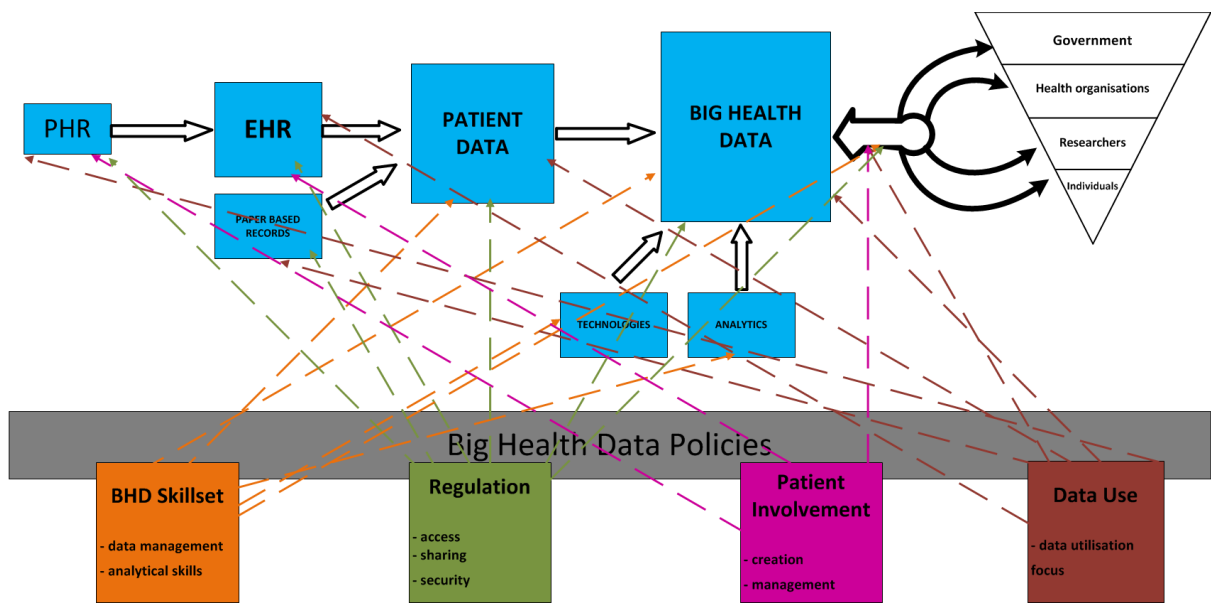


Figure 2-7: Big Health Data Theoretical Framework: All requirements

A combination of all these four pillars then provides us with the above framework for health data use.

2.3.9 Summary of Data Policies

The role that policy making plays in advancing the health of the population should never be underestimated. There is few tangible legislation that speaks to the EHRs, PHRs, data

ownership and how all these fit together in promoting healthcare utilisation. Where the policies are not expressly silent on these things, they (policies) seem to be in contrast to efforts promoting utilisation. A need to readdress policy direction on health data has therefore been noted and a framework for that can assist in developing big health-data policies has been proposed.

2.4 Data Utilisation

It is becoming increasingly clear that a shift in the way data is handled and perceived is needed (Kearney, 2014) if real value is to be realised about the data. It has been realised that there needs to be a shift of focus of policies and governance to be able to promote more usage of data (ibid). Despite that large amounts of healthcare data is being collected and stored, data utilisation is still low and there is a lot of room for improvement (Hripcsak et al., 2014). Manyika et al. (2011) also point out that the healthcare fraternity has generally been slower in implementing better data use for improved operational performance (Manyika et al., 2011). One of the main challenges that have been identified as culprits here is policy issues that need to be addressed to promote value creation from the data (Hripcsak et al., 2014). “Use of personal data in innovative ways in the health sector can yield significant economic and social value for all stakeholders.”(The Boston Consulting Group, 2013:pg 9). Kearney (2014) points out that for this value to be realised, data use strategies to be adopted will have to adhere to the following three key aspects that promote ease of use data: i) Transparency, ii) Accountability and iii) Empowerment of individuals (Kearney, 2014). Another key aspect will be the identification of the context of use of the data (World Economic Forum, 2014) where it is argued that a proper clarification of what data will be used for building trust between the data user and the data owner.

2.4.1 Consensual Access VS Non-consensual Access to Patients Health Data

In a scenario where patients have full rights to their information, with the power to prosecute those who breach their rights, two extremes of consensual access and non-consensual access would yield two distinct result sets (Evans, 2011).

According to a study by Tu and colleagues (Tu et al., 2004), the group of people that will agree to informed consent about their data use have different medical needs and conditions from those who will not agree to informed consent. What this means is that either of the two groups, those that are pro-informed consent and those that are anti-informed consent or those indifferent about it, can never be true representatives of the general public (Evans, 2011). This then introduces a consent bias to the results of any study undertaken on data from consensual information or that from non-consensual information.

It can be argued therefore that consent bias can be potentially a huge problem that can affect data utilisation adversely if not properly managed. Consensual approaches cannot produce unbiased data. On the other hand, non-consensual approaches seem to be more favourable for the purposes of research, analytics and for public health interests.

2.4.2 Suppression of Data

There is a considerable amount of health related information that has never been published or made available for public scrutiny (Pasquale, 2013). Pharmaceutical companies are popular because of this, hiding data about the performance and the negative effects of their products (Pearlstein, 2011). It is also very popular that companies will tend to misrepresent research results on their products to suit them and engage in selective reporting, avoiding publishing negative effects of their products. In fact, the amount and value of health information out and available to the public is greatly understated. This is just one area where this problem is prevalent though it can be safely concluded that most health research results, especially those that affect big cooperation's financially, are kept under the carpet.

Competition is another reason why companies will tend to want to keep their health information exclusive to them. To maintain a competitive edge over other providers, institutions will try to find ways on how they can, through their information, keep ahead of others. Information about pricing standards, quality of doctors, mortality rates and services provided is often tailored to suit the providers (Pearlstein, 2011).

There is no culture of sharing vital and important findings in healthcare (Pasquale, 2013). Data siloing is a common trend as providers try to out-compete each other. This is impeding the kind of large-scale data analysis that other industries are engaged in. The result of this exclusive behaviour means that some organisations might pursue endeavours like research into diseases and best operating practices that another organisation would have already found a solution for. This is costing lives.

2.4.3 The Cost of IP Protection in Health

Closely related to this data exclusive behaviour is the issue about intellectual property laws and their application in the healthcare fraternity. This is a worrying reality, that a scientist might discover something new, extend the bounds of health knowledge and not make that discovery available to others for some personal interests. This is a moral issue that can be addressed from a policy perspective as our generations keep getting more and more

plenteous and morally degraded. Policies have to find ways of minimising the effects of IP laws in healthcare with the vision of even nullifying all of them in health.

2.4.4 Regulations on Information Flow

Regulations on information flow have the potential to hamper innovations (Pasquale, 2013). The flow of health information is seriously monitored and guarded and access is even more barricaded. An example of this phenomenon is disclosure laws which make it difficult for practitioners to use patient data in any secondary use initiatives without the patients consent.

2.4.5 Anonymisation and Re-Identification

The information privacy debate in support of data utilisation and sharing has often overstated the power of data anonymisation (Ohm, 2010). Data anonymisation is when data that is related to users is stripped of its identifying characteristics so that sharing and secondary utilisation can be done without the original owners of the data being revealed. However, the growing power of re-identification tools is making proper anonymisation become an illusion. When data gets to a point where it is completely anonymous, then it would have also reached a state where is not usable for secondary purposes (Keen et al., 2012; Ohm, 2010). Studies have shown that basic information combinations like date of birth, postal code and sex can be used to identify users uniquely by experienced analytics. It can then be concluded that data usefulness and data anonymisation are indirectly proportional. Such complications to the information privacy debate, further complicate advances in data utilisation.

2.4.6 The Open Data Case

One other pro data utilisation case is that of the adoption of an open data policy (Yiu, 2012a; Zuiderwijk & Janssen, 2014a). The argument made here is that opening up data in a reusable format strengthens engagement and participation hence yielding new innovative solutions to social and business constraints (Huijboom & Broek, 2011). The possibility of such opening of data can only be implemented if the authorities in concern take an open data policy approach to the development of its data access policies.

According to Yiu (2012) the benefits of adoption of open data policies are three-fold, i) the economy is given the ability to treat data at par with assets like transport infrastructure and money systems, ii) social advances can be realised that “increase participation, interaction, self-empowerment and social inclusion” (Zuiderwijk & Janssen, 2014: pg19) and iii) political improvements and awareness that increases perceptions of transparency and accountability which are good for government. The general idea that literature suggests is that adoption of

open data policies will promote data utilisation and result in the better promotion of sharing for the purposes of finding solutions. This is a position that the researcher also feels inclined to adopt.

2.4.7 Summary of Data Utilisation

In the context of data utilisation, data suppression, IP protection and regulations on information flow are therefore currently disadvantageous rather than having advantages. Analytics is being greatly hampered and even where possible the analytics results will be distorted as they would not have taken into perspective the whole picture of the situation. The research imperative for data access needs serious attention from a law and policy perspective and there is now a serious call for a movement to full access to health information of the kind only policy changes can achieve. Government involvement in addressing this issue is, therefore, a key feature because of its role in the control of flow and analysis of information. Alone, the government cannot achieve the changes of the kind needed here, it is going to be a process that will need the cooperation of both the public and the private sector. The general idea that literature therefore suggests is that implementation of open data policies has the potential of improving data utilisation which in turn results in other wider benefits for the community at large.

2.5 The South African Scenario

Having discussed patient data, big data, data policies and data utilisation, it is important for this literature platform to explore the prevalence of these aspects in the South African context so as to give perspective to the direction and location of this study.

Africa, in general, has the poorest healthcare services of the rest of the continents in the world (KPMG Africa Limited, 2012). Table 2.1 shows how Africa compares to the rest of the World using some key health indicators. Tuberculosis, Malaria and the Human Immunodeficiency Virus / Acquired Immune Deficiency Syndrome (HIV/AIDS) are among the prevailing struggles that the continent battles with, with huge losses of life each year. Most governments in Africa battle to find enough funding for their healthcare activities and initiatives (ibid) and rampant corruption worsens the amount of state expenditure on health. Only four African nations were found to be compliant with the health-related Millennium Development Goals (MDGs) pledge set by the African Union (AU) in 2010 (ibid) and South Africa is not part of them.

Table 2-1: The State of Health in the World

Selected health indicators				
INDICATOR	WORLD	AFRICA	EASTERN MEDITER- RANEAN	EUROPE
Life expectancy at birth, 2009	68	54	66	75
Physicians per 1,000 people, 2010	14.0	2.3	11.0	33.3
Under-5 mortality per 1,000 live births, 2011	51	107	58	13
Maternal mortality ratio, per 100,000 live births, 2008	260	620	240	21

Source: WHO, UNICEF. 'Eastern Mediterranean' includes North Africa except Algeria

(Adapted from KPMG Africa Limited, 2012)

Zooming into South Africa, we find that the state of health of the nation is also not at a satisfactory level. According to the South African Health Review 2014/15 report, there is mismanagement of expenditure at various levels, many various vacant important posts, areas where healthcare facilities are still inaccessible and a lack of policy implementation and awareness (Gray & Vawda, 2015). Financial barriers were one of the top reasons for people failing to access healthcare facilities (World Health Organisation, 2012) worldwide and this is also true for South Africa (Gray & Vawda, 2015; Harris et al., 2011). There is also an imbalance in the approach that public healthcare is using to tackle the health burden of the nation as compared to the approach of private healthcare. This refers to the problem of inequalities across the different socio-economic strata's of the South African population. The government still needs to address the fact that its population need to be able to access healthcare services regardless of their socio-economic status (Phelan et al., 2010; Statistics South Africa, 2011; Harris et al., 2011). Despite a general dearth of satisfactory healthcare, there are success stories and these keep the nation's population optimistic that things will improve.

It will, therefore, be imperative that South Africa and other African countries adopt high technologies such as big data processing and analytics to try and improve their health

(Kruger, 2016). Africa has the least studied data of all the continents in the world (ibid) and various opportunities exist if already existent data can be accessed and analysed. The South African government has at least taken note of this opportunity and has already rolled out various projects like the National Health Information Repository and Data Warehouse Project (NHIRD) (Health Information Systems Program-SA, 2015) to try and harness the power of its health data. One other such project is the Data Harmonisation Project (Western Cape Government Health, 2014) of the Western Cape government. More projects in this direction are needed.

A look at the National eHealth Strategy, South Africa 2012/13-2016/17 document (Department of Health South Africa, 2012) shows that the SA government is trying to increase its usage of healthcare data to improve health. The document alludes to the fact that there is a need to use data more in research initiatives and to also create functional interoperable systems across the country's different provinces. It also mentions the need for the ability to view the country's data repositories from one perspective and although there is no specific mention of the words big data and personal health records, the notion of these concepts is brought out as something that has not been implemented yet but that the DoH and government would like to gravitate towards.

2.6 Conceptual Model

Below is a visual presentation of how the concepts that have been discussed in this chapter are connected to show the direction of this research. Patient data is shown as a composition of paper-based data as well as EHRs. PHRs are shown as a subset of EHRs, but here a rift exists because of a number of issues that make it difficult to use PHRs as part of the bigger body of big health data. It is this rift as well as how to better harness big health data, which the study wants to understand how current policy is endeavouring to address. The issues represented here are the challenges which have been discussed in the previous sections of this chapter (see Section 2.4). The study follows the idea that if these issues could be resolved by the redemptive measures that policies, if designed consciously, could be formulated to provide. These redemptive measures denote corrective or preventative elements that policies could provide to deal with the identified and possible future issues that can affect meaningful data utilisation of healthcare data.

The model also shows the main stakeholders involved and how they interact in delivering healthcare as a service to the community.

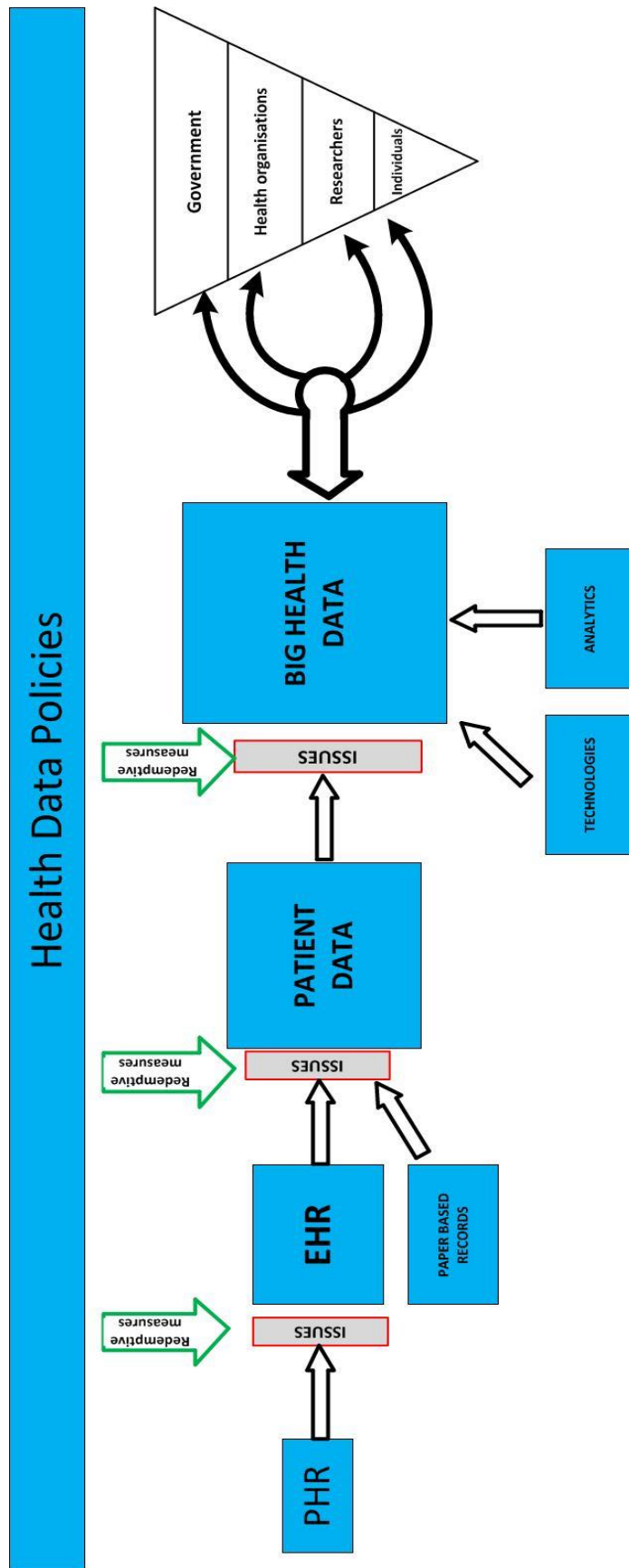


Figure 2-8: Research Conceptual Model

2.7 Conclusion – The Policy Lacuna

It is argued by Hripcsak and colleagues (Hripcsak et al., 2014) that most of these tricky issues around data use can be alleviated by a reworking of the policies that govern the access and use of healthcare data. Keen et al. (2013) also pick up this argument as they identify information governance as one of the three main areas that need attention if ever proper data utilisation of Big Health Data is to be realised. The other two are data management which has many links with information governance and technological infrastructure. There seems to be a lack of separate policing between the primary and secondary uses of healthcare information (Keen et al., 2012). There is also a lack of policy infrastructure especially adapted to support making healthcare information available for secondary uses (Ohno-Machado, 2014; Hripcsak et al., 2014; Pasquale, 2013).

It is in this light and in the direction of advice given by Hripcsak et al. (2014), Keen et al. (2013), Ohno-Machado (2014) and Hoffman and Podgurski (2013a) that this research is pointed in. “Big Data is a big deal in biomedical research and healthcare.” (Ohno-Machado, 2014: pg 193). There is a lacuna between the prospects of Big Health Data and the policies needed to properly support such kind of data use and or exploitation as envisioned by Big Health Data enthusiasts (Menius & Rousculp, 2014).

Pasquale (2013) clearly identifies the need for a policy stance on health data that does not yet exist to ensure the full realisation of the potentials of health information technology. Top of the agenda for this policy stance is making available all key data that is being collected in health systems and yet at the same time ensuring that metrics being used to watch privacy, confidentiality and intellectual property are still met. This will be the key challenge for the health policy maker of today and tomorrow.

CHAPTER THREE

RESEARCH METHODOLOGY AND DESIGN

3.1 Statement of Research Problem

Despite some success of big data methods in other fields such as business informatics, the same results seem not to be as easy to achieve in health informatics. One main difference between business informatics and health informatics is the orientation of their data policies. Can a link be made between data policies and the success of big data in healthcare? There seems to be a mismatch between the goals of current healthcare data policies and those of big data in healthcare. This is negatively affecting research efforts in healthcare by hampering the adoption of big data technologies for better data utilisation. If the status quo prevails, some valuable insights into the health of the world population in general which have the potential of saving lives will continue to be overlooked. How these two relate needs to be further researched in order to explore possible ways of better congruency.

3.2 Research Aim

The aim of my research is to understand the relationship between data policies and the prospects presented by big health data with a focus on personal health data. If this could be properly pointed out then the design of policies that support big data utilisation will be greatly facilitated. Hopefully, some recommendations for policy making will be made at the end of this study.

3.3 Research Questions and Objectives

3.3.1 Research Questions

Table 1-1: Research Questions

Research problem: There seems to be a mismatch between the goals of current healthcare data policies and those of big data in healthcare. The role of PHRs in healthcare also seems to be a peripheral issue.	
Main Research question 1: How are healthcare policies influencing data use trends like big data and personal health records in South Africa?	Main Research question 2: What is the nature of current data policies in the context of healthcare data use in South Africa?
RSQ1.1: How are personal health data and big health data policies affecting data use?	RSQ2.1: How do current personal health data and big health data policies promote information access and sharing?
RSQ1.2: How are healthcare data policies factoring into consideration foresighted changes in data use strategies?	RSQ2.2: What are the differences that could be existing between current healthcare data policies and those of big health data?
RSQ1.3: How are health data policies providing support and a foundation for big health data and personal health data policies?	RSQ2.3: What framework can be used to improve healthcare data policies in the context of big health data?

3.3.2 Research Matrix and Objectives

Table 1-2: Research Matrix

Research Questions/ Objectives, data collection and analysis methods.	Objectives	Document Analysis	Semi- structured Interviews	Affinity Diagrams	Thematic Analysis
How are healthcare policies influencing data use trends like big data and personal health records in South Africa?	To determine how aligned current personal health data policies and big health data policies are.	X	X		X
How are personal health data and big health data policies affecting data use?	To gain insights into how data policies can increase data utilisation	X	X		X
How are healthcare data policies factoring into consideration foresighted changes in data use strategies?	To determine the agility and responsiveness of big data systems to data utilisation changes.	X	X		X
How are health data policies providing support and a foundation for big health data and personal health data policies?	To determine how aligned current personal health data policies and big health data policies are.	X	X	X	X
What is the nature of current data policies in the context of healthcare data use in South Africa?	To determine what are the aims of personal data policies and those of big data policies.	X	X		X
How do current personal health data and big health data policies promote information access and sharing?	To determine how information access can be increased by policy measure.	X	X	X	X
What are the differences that could be existing between current healthcare data policies and those of big health data?	To examine what is required to make personal health data policies congruent with big health data policies.	X	X		X
What framework can be used to improve healthcare data policies in the context of big health data?	To determine the basis of formulation of future data policies.	X	X	X	X

The objective of this study is to explore the data policies of healthcare data specifically those of personal health data and those in big health data. Guidelines on how these policies can be developed to increase data utilisation were reached at the end of the study.

3.4 Research Methodology

3.4.1 Research Philosophy

The term research philosophy refers to how one understands what constitutes knowledge and how knowledge is created (Saunders et al., 2009). During the process of a research, the actions that are taken by a researcher are highly influenced by their general view of what constitutes life and their values and character. These values maybe called the assumptions of the researcher and these assumptions also had a significant bearing on how the results of the study were formulated.

3.4.1.1 Ontology

The ontological position that this study subscribes to is the subjectivist position. Ontology refers to the way how one perceives reality and whether the existence of entities, in reality, is dependent upon other relationships (Burrell & Morgan, 1979).

There are two main ontological positions that a research can be aligned to. The first one is called objectivism, and objectivists believe that social entities like humans can well exist as objects independent of their context or surroundings. What this means is that if for example, we could take a law as a social entity, that law would be absolute globally.

On the other hand, the second position would interpret the same law, as a social entity which is relative to the context that it finds itself in. The subjectivist holds the opinion that for them to understand a social phenomenon, the underlying social details of the situation also need to be studied and put into perspective for the results of the study to be meaningful (Saunders et al., 2009).

It was best to employ a subjectivist position for this study because since the study is concerned with understanding a relationship between the different concepts as they are socially constructed. Relations do not exist independent of social constructs and how humans attach meaning to them. Furthermore, policies in themselves are contextual i.e. their relevance and interpretations are dependent on external perceptions and situations.

3.4.1.2 Epistemology

The epistemological position of this research is a critical interpretivism (Pozzebon, 2003). The term epistemology refers to how knowledge is evaluated as relevant to a field of study (Saunders et al., 2009). In other words, epistemology is the classification of knowledge as either constituted of empirical data and non-empirical data.

Positivism is the philosophical stance that defines reality as observable or empirical entities. Results of positivist studies are normally generalised across all social entities that are similar to the participants of the study. The positivist approach also believes that the researcher is neither affected nor affecting the subject and results of the researcher. Research which reflects positivism is usually quantitative in nature and the methods are structured in such a sense that they can be easily replicated and repeated.

On the other hand, there is an epistemological stance that is called interpretivism (Klein & Myers, 1999; Walsham, 2006) which this study subscribes to. With interpretivism, social realities can be non-empirical entities such as feelings that cannot be observed. Interpretivism is the direct opposite of positivism and believes that rich insights about our surroundings are lost if research is conducted with the positivist philosophy. Interpretivism understands that humans are subjective creatures which relate to information differently and does not treat them as mere objects. Therefore the behaviour of humans has implications to what reality is i.e. they (humans) are significant contributors to what we experience though at times implicitly. To the interpretivist meaning is relative, unique and even contextual. “Law-like generalisation”(Saunders et al., 2009: pg 116), are not recognised as valuable knowledge platforms.

Another important consideration of the interpretivist is their own role in the process of development of new knowledge. Unlike the positivist, the interpretivist believes that they have a significant role in determining what the results of the research become. The interpretivist is an involved researcher (Walsham, 2006). As they try to understand the research subjects from their own point of view, their own points of views are unavoidable. How they conduct the research, their feelings and views introduce a bias to the overall findings. This philosophy is called axiology. It is, therefore, the goal of the interpretivist to be both neutrally involved as well as a full action researcher. They somehow need to find this balance.

3.4.1.3 Research Paradigm

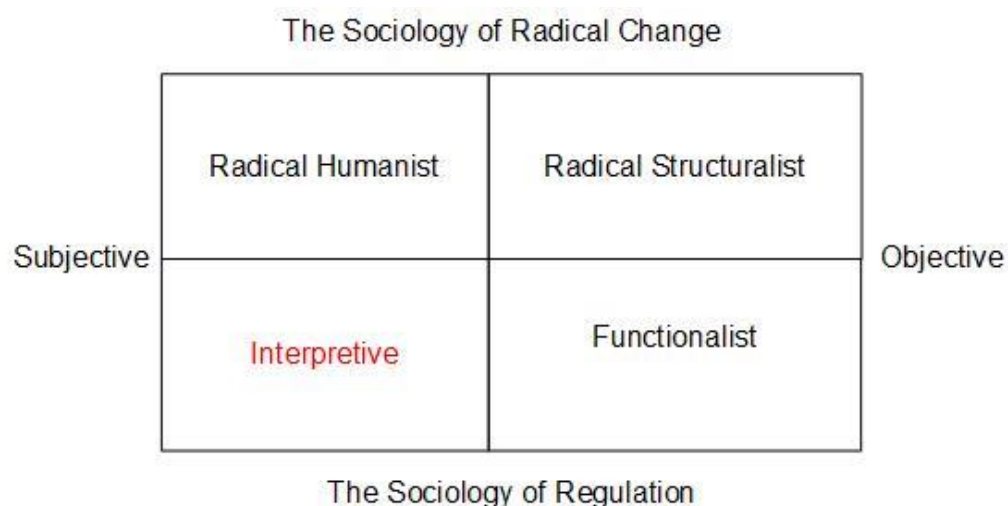
Now that the philosophical ontologies and epistemologies that reflect the philosophical stance of this research have been discussed, we will now summarise this section by explaining the research paradigm of this study which is the interpretive paradigm.

A research paradigm is a topological view of all the philosophies that are held by the researcher (Walsham, 2006). It shows the overall implications of the direction of the study and what the possible intended outcomes will be (Burrell & Morgan, 1979).

Burrell and Morgan (1979) in their work on sociological paradigms and how organisations are analysed, moved from an era where the ontological positions of subjectivism and objectivism were viewed, used and analysed separately from the epistemological positions of radical change and regulation. Depending on social issues that will be current in a particular society, they argued that at any point in time, the philosophical positions of social theorists have always reflected ideologies and assumptions of either one of the ontological positions (subjectivism or objectivism) and also ideologies and assumptions of either one of the epistemological positions (radical change or regulation).

It is then that as they were looking at the two dimensions of ontological and epistemological philosophies that they came up with four paradigms that a social theorist can belong to any one time i.e. the four paradigms are mutually exclusive. The diagram below illustrates these four paradigms.

The four paradigms are like a map showing one's personal perspective and understanding of their social surroundings. They represent, like in the animal kingdom, the intellectual territory of the theorist, as well as the meta-theoretical assumptions that they have about how they conceive their reality.



**Figure 3-1: Four paradigms of analysis of social theory
(Adapted from Burrell and Morgan 1979)**

According to the above four paradigms, this research falls more into the interpretive paradigm. Under the interpretive paradigm, the interpretive theorist believes that they can achieve most of what they need by emphasising on regulations which will be followed by humans who are generally peaceful, ordered and integrated. They seek to understand the social world subjectively, not overlooking connections that to the objectivist are not important. They are concerned about the fundamental meanings that shape society and believe in the ever continual process of development as contexts change. It is the strong belief in regulations that excite the researcher and confirm that this research is interpretive in nature. Through properly aligned data policies, effective utilisation of already existing data through technologies such as big data can be achieved where they are not achievable now. By understanding the constructed current environment better policies can be constructed that are contextualised and reflect the meaning of data that can be varying across many cultures and fields of work.

3.4.2 Research Approach

There are two approaches that a research can embark on. The first is called the deductive approach and the second is called the inductive approach. This research employed the inductive approach of theory or hypothesis creation (Patton, 2003a; Saunders et al., 2009).

The deductive approach is a method that first starts by creating a theory then formulating a way of proving or testing it. It starts by formulating a hypothesis, then collecting mostly quantitative data so as either to support the hypothesis or to disapprove of it. Results of studies that are deductive are usually generalised i.e. they can be applied in other cases of similarity since they employ structured methodologies of data collection and also because deductive studies normally take an objectivist stance though not always. Another characteristic of deduction is that it uses control and experimental research participants to test the hypothesis. It is normally used in the natural sciences.

Unlike the deductive approach, the inductive approach, when presented with phenomena to study, will start by a rigorous data collection process which may not be structured and then after analysis of the findings come up with a theory to explain the phenomena. The deductive approach has limitations on the branches of truth that it may pursue whereas with the inductive approach, more possible leads of truth can be discovered and pursued therefore producing more reality. This is the reason why this study follows the inductive approach and not the deductive approach. The inductive approach is also more suitably aligned to the research philosophies underpinning this study which have already been discussed in previous sections. These philosophies reflect value being placed on information

that is non-empirical therefore the data collection and analysis methods to be used must support qualitative data.

3.4.3 Research Strategy

This research employed the multiple case study approach (Saunders et al., 2009). Case study research allows the investigation of phenomenon within its real-life context, making it possible to retain context as a constant in a study (Yin, 2013). Case study research is a qualitative technique that supports in-depth gathering of a data from a subject that would have been identified as highly relevant to the study which can be helpful in two ways: i) the risk of concentrating on research subjects that are not relevant is minimised, and ii) data collected is richer and deeper in description (ibid). The following section explains more why the case study strategy was the choice for this study.

3.4.3.1 The Two Tier Healthcare System

Healthcare is defined as a human right by the UN (World Health Organisation, 2015), which is why citizens will generally assume that healthcare provision is the responsibility of the government. Most ethical governments do fill in their defined roles as healthcare providers, making available to the public population basic care, at times freely but normally at a cheaper subsidised rate (Ataguba & Akazili, 2010). This can be referred to as the first tier of healthcare. On the other hand, private organisations also provide specialised healthcare which is a better quality and service than that provided by the government because they have more resources and the services are provided to patients who have the means to pay for the services. This is true for almost all the countries in the world with a few exceptions. In the South African context, the private healthcare sector caters for 20% of the affluent population who can afford insurance whereas the rest have to deal with the public sector which is under-resourced and stretched. In the UK 8% of the population is serviced by the private sector (Raymond Lang, 2011) though their public NHS is adequately funded.

It is because of this setup of healthcare provision that the strategy of this research has been designed as a multiple case study approach. One case was a public healthcare associated data management system and the other a private sector associated data management system. Both had to, in some way, be involved in big data.

3.4.4 Units of Analysis and Sampling

The units of analysis for the research are the Western Cape (WC) Department of Health (DoH) Information Directorate Office and Information Centre and Netcare's Information Management centre. The former is a public service entity and the latter is a private

healthcare service entity. The WC DoH is the sole public healthcare provider in the WC that is why it was selected as the public case. Netcare, a private entity has been selected as the other case as it is one of the largest private hospital networks not only in South Africa but in Africa (Willie & Matsebula, 2007; Netcare, 2016b). Netcare has over 55 hospitals, over 10 000 beds and over 350 theatres across its various hospitals across the country (Netcare, 2016a). Mediclinic is the second largest hospital network with around 52 hospitals, 8 000 beds and 270 theatres across the country (Mediclinic, 2016).

Personal health data and big data are still largely new areas, still being defined and under continuous change (Vayena et al., 2015). This makes the research population unknown and also introduces a bias towards the middle and upper-class individuals of society (University of California, 2014) as well as affluent health institutions with big data capability. Therefore, a random sample representative of the population cannot be drawn. The methods used to select the sample reflect the non-probability sampling technique.

3.4.5 Time Horizon

The research follows a cross-sectional time horizon i.e. its results will relate to a specific time period and cannot be applied over extended periods of time.

3.4.6 Data Collection Techniques

3.4.6.1 Semi-Structured Interviews

Semi-structured interviews are interviews where the interviewer asks questions that might vary depending on the interviewee. An initial list of themes, drawn from the main points as noted in the literature, and questions were designed though some were eventually left out or some added. This allowed for themes that the interviewer had not thought of or anticipated to be captured as well. Semi-structured interviews are good for exploratory studies like this one (Saunders et al., 2009).

Semi-structured interviews were conducted with relevant identified people who are involved in data policy and regulations creation and implementation at the two selected case organisations. A minimum of four interviews were conducted for each case.

3.4.6.2 Literature and Documents Analysis

An analytical review of existing documentation on the topic was conducted. The current data policies related to big data and PHRs data from both cases as well as the government was analysed to provide deepened understanding of the topic. Reviewed articles, books, websites and journals were also be consulted.

3.4.6.3 Affinity Diagrams as a Design Research Tool

Many times during group brainstorming sessions, a lot of ideas are generated. It can be a difficult task trying to make sense of all that information and making sure that some brilliant trains of thought are not lost. Affinity diagrams are a great way of organising such a kind of information overload and also ensuring that valuable information is not lost by random, impromptu summarisations. The term was first coined by a Japanese anthropologist called Kawakita Jiro in the 1960s (Mind Tools, 2015). They can be described as “a creative process used for gathering and organising large amounts of data, ideas and insights by evidencing their natural correlations” (Tassi, 2009:pg 1).

The process that affinity diagrams follow is that related ideas are grouped into themes which help to structure the information. Themes can then be further grouped if too many so that a final set of manageable themes is then drawn out. Affinity diagrams also help in noticing connections between themes or ideas that might not have been comprehended before (Mind Tools, 2015).

Affinity diagrams fit very well in the design research realm as they have a key focus on the interaction between the participants. They are a great way of invoking the deeper social views from participants which are normally ignored by traditional research methods (Duncan of Jordanstone College of Art & Design, 2013). They are also useful to establish a common ground in the group.

As such, the research employed affinity diagrams with the patients of the two cases. The main goal in doing the affinity diagrams with the patients is to gather their views on how they feel policies impact their participation in their own healthcare. The other theme was how they felt about their data being used for secondary uses and how they would like to get involved in managing their PHRs.

3.4.7 Thematic Analysis

The research employed the thematic method of analysis of the research findings. Thematic analysis is a framework that is used for analysing data in which recurring or related data are identified as patterns which then lead to the formulation of themes (Braun & Clarke, 2006). As data is being collected and after collection, the researcher or analyst begins looking for issues that are or may be speaking to the research questions governing the study. Related issues are then developed into patterns or themes whose meaning is then deduced. It is used in qualitative research. In fact, all qualitative research data analysis techniques are to a great degree subsets of thematic analysis (ibid).

Thematic analysis is a subjective method in that the themes developed from the same data by one researcher may be different from the themes developed by another. This flexibility aspect of thematic analysis makes it a popular method as it allows researchers to express their views without being confined to a rigid method of analysis when formulating their themes. What constitutes a theme is, therefore, a subjective matter though it has to reflect a patterned meaning about the data.

Thematic analysis can be categorised into two categories, inductive (bottom up) thematic analysis and theoretical (top down) thematic analysis (ibid). With inductive thematic analysis, the themes identified will be more connected to the data that would have been gathered and may tend not to be related to the research questions in subject. This paradigm tends to produce very rich comprehensive data descriptions. On the other hand, the themes developed when using theoretical thematic analysis tend to be more connected to the theoretical underpinnings of the researcher and their analytic interests. This paradigm tends not to produce as rich descriptions of the data in overall, as compared to inductive thematic analysis, but themes mostly of interest to the researcher will be extensively covered.

Thematic analysis can be divided into yet another two sub-spheres which are semantic thematic analysis and latent thematic analysis (ibid). With semantic thematic analysis, the researcher is only concerned with the surface meaning of the data in concern and is not interested in the deeper underlying stories that are not explicitly stated or observed. Patterns and relationships between data are made but no other meanings beyond the obvious are made. On the contrary, thematic analysis at the latent level further goes beyond the bounds of semantic in order to search out the deeper underlying patterns and meanings of the data. The approach to the data is interpretive in nature not wanting to leave any stone unturned. After a proper latent thematic analysis, results are normally articulated as theories that can then be applied in other studies or that need to be proved. This study employed the theoretical, latent thematic analysis paradigms.

The theoretical framework developed Chapter 2 (see Section 2.3.7) was also be employed as a theoretical lens of interpreting the findings of the study. Employing the framework to interpret the findings of the study introduces deduction into the research whereas the main philosophical underpinning of the study is inductive (Schultze, 2000). The concepts of internal homogeneity and external heterogeneity (Patton, 2003b) was also be applied to ensure the credibility of the themes as separate, distinct and non-conflicting.

3.4.8 Data Presentation

The findings of this research as well as the final outcomes and recommendations are published as part of this thesis at the end of this study. Various presentations of papers will also be published in journals. A poster was also designed that provided a summarised view of the conceptual model of the whole study.

The researcher also used visualisation tools like Tableau to present the findings of the data in a format that is appealing and easy to understand.

3.4.9 Research Quality Management

3.4.9.1 Scope of the Study

This research study was conducted on cases that are both in South Africa in the province of the Western Cape. The study focused on issues that the researcher, pointed to by literature, believed could be resolved by policy adjustments. The study is a qualitative study and was completed within the three years.

3.4.9.2 Assumptions

This research was undertaken with the following principal assumptions of the context in which it is conducted.

- i) The results of the study will not be generalised across the developed world as well as the developing world. This assumption has been taken because although South Africa exhibits a heterogeneous developmental behaviour i.e. it has characteristics of the developed country as well as characteristics of a developing country, more cases are required before the findings can be generalised in terms of developed versus developing worlds.
- ii) More information means better representation which intail means better big data results.
- iii) Although it is not possible yet to predict trends and futuristic insights from analysing the datasets, the derived insights can form the basis for further studies.

3.4.9.3 Delineation of Research

This study will not propose policies but recommendations or guidelines that could be incorporated by policymakers. No patient data or information was investigated.

3.4.9.4 Ethical Considerations

Research ethics refer to the epistemic imperative or the moral commitment that researchers are required to make in the search for truth and knowledge (Mouton, 2004). Ethics give researchers guidelines on how they should conduct research. Research design should

encompass ethical considerations that protect the well-being of research participants and this should be considered for both humans and non-humans. Ethical guidelines include but are not limited to, informed consent, confidentiality and anonymity, harm to respondents, voluntary participation, respect for study sites, and honest reporting of findings. At every stage of the study, the above ethical guidelines were observed. No research processes were conducted unless written permission had been obtained where necessary from the necessary from the health research ethics committee of the institution in concern as stipulated in the SANHA (South African Government, 2004). The Protection of Personal Information Act (POPI) also stipulates the importance of ensuring that the ethical code of the institution or individual in concern is respected when engaging with their information (Parliament of the Republic of South Africa, 2013). The research respected this premonition. The highly ethically sensitive part of the research was when data was collected from patients using affinity diagrams. Uttermost care was taken to ensure participants feel comfortable about the process. Permission to collect data from the Western Cape's DoH as well as Netcare was obtained before data collection was begun.

3.4.9.5 Evaluation and Reflections of the Research

After completion of the research a formal evaluation that will be separate from the thesis, will be conducted. The results of the evaluation will be part of a proposed paper to be written on the experiences of postgraduate students at the CPUT. Klein and Myers (1999) set of 7 principles for conducting and evaluating interpretive research studies were also employed as a validation tool for the research.

The reflections will be used as a basis for further improvements in research processes by the researcher for future research projects envisioned that include a PHD degree and IT consultancy research as a branch of an all-round IT company currently being developed.

3.4.9.6 Contribution of the Research

This research contributes to the Data Management Association (DAMA) Guide to the Data Management Body of Knowledge (DAMA DMBOK Guide). The DAMA DMBOK specifies ten data management functions (Mosley, 2008) of which this research entails aspects that are related to these ten functions. The research also contributes to the Information Systems Development Approach for Societal Development (ISD4D) Architecture design petal (Korpela, 2011). The research made a contribution by coming up with guidelines that could inform healthcare data policy creation and design.

Importantly, one other major contribution that this study makes to the big data world is a big health data theoretical framework that can be a valuable tool in the development of healthcare policy infrastructure and controls. This framework has the potential to cover the gap that currently exists between big health data users and the envisioned advantages and uses of big data healthcare. The ultimate purpose of the study is to increase meaningful usage of healthcare information for the purposes of improving health. If this could be realised, ultimately many lives could be saved.

3.5 Conclusion

In this chapter, the process of how this study was undertaken was laid down. It was noted that the problem that it is trying to address is the fact that there seems to be a misalignment of healthcare policies efforts with trends of data use prevailing today such as big data. Therefore, this study explored this issue to understand it better. The research questions guiding the study were drawn and the methods to be employed in the study were also discussed. It was noted that the study pursued a subjectivist, interpretivist, critical interpretivist and inductive dimension of research. Multiple cases formed part of the units of analysis and the study is cross-sectional. Semi-structured interviews were employed as collection techniques together with the documents and affinity diagrams. Thematic analysis was the analysis method and a proposed theoretical framework was developed to act as a lens of viewing the study's findings and results. Various quality management techniques were also discussed that aided the researcher to ensure that this study meets the standard criterion of a standard research article of reputable standing.

CHAPTER FOUR

FIELDWORK REPORT

This chapter's purpose is to provide an outline of how the process of data collection and its subsequent analysis was done. As a requirement for a masters degree, meticulous standards were followed in these processes to ensure that the data acquired is accurate and of a good quality befitting a masters qualification more so because in the field of ICT quality information is as gold. The researchers experiences, as well as lessons learnt, will also be highlighted upon.

4.1 Data Collection

The process of collecting the data happened between the period of November of 2015 and April of 2016. Semi-structured interviews were conducted with the Western Cape's Department of Health (public entity) between the months of November and December 2015 in Cape Town. The other interviews were done with Netcare (private entity) during the month of February 2016 in Johannesburg. In May of 2016, the last data collection activities, i.e. co-design sessions with some public health care users as well some private healthcare users were conducted in Cape Town.

4.2 Undertaking Documents and Literature Review

The process of document analysis and literature review was a permanent feature throughout the duration of the study. Information sourced through continuous analysis of literature and documents acquired throughout the duration of the project kept enlightening the direction of the study.

During the interviews with participants of the research, some would refer to particular documents and where their relevance to the study was ascertained, the researcher would source them for consideration as possible literature for the study. One such document is the Healthcare 2030 vision of the Western Capes government (Western Cape Government Health, 2014) which provided an overview of the direction that the WCDH envisions to take in the context big health data. Other policy documents would also be alluded to and where needed, they would be included in the body of literature and documents that informed the study.

Generally purposive sampling was applied in choosing and the documents and literature that would inform the study. Documents and literature was sourced from web publications, journals, the internet and other relevant websites.

4.3 Undertaking the Semi-structured Interviews

As part of the data collection process, interviews were set up with the WCDH information directorate personnel during the months of October and November of 2015. It was difficult to set up these interviews as participants cited that they were busy finishing up work before the year ended for the December holidays but eventually I managed to schedule them.

A total of 7 personnel that are involved in the information management and or data policy management projects or departments at the information directorate office were identified by the director of the Information Directorate of the WC, who I also managed to interview. Interviews were then carried out as semi-structured interviews with these participants at 4 Dorp St, Cape Town, which is the address of the offices of the Information Directorate of the WCDH. These represented the public case.

For the private case, interviews were set up with Netcare which is the largest private hospital chain in South Africa (Willie & Matsebula, 2007; Netcare, 2016b). A total of 6 personnel that are involved in the information management and or data policy management projects or departments were identified by one of the IT Governance Officers at Netcare. Interviews were then carried out at the Netcare Head Offices (76 Maude St, Sandton) and also at the Netcare Auckland Park Offices (2 Bunting Road Auckland Park).

One interview was conducted with an expert researcher from the Council for Scientific and Industrial Research (CSIR) in the field of information technology.

Refer to the appendix A3 for more details about the entire body of interviewees who participated in this study.

4.3.1 Interview Protocol and Process

The following guidelines and standards were followed to ensure quality and ethical correctness of the interview processes:

- Interviewees were provided with the interview themes and questions beforehand.
- I would also consider how the interviewees look on the day of the interview to avoid any negative perceptions about them. The purpose was to try and blend in as much as is possible into the environment of the interview as well as with the people (Saunders et al., 2009).
- On the day of the interview, the I would ensure that I was on time for the interview as much as I could, though in some instances this would be impossible due to various factors.

- I would introduce myself, the research interests as well as the purpose of the interview.
- The interviewees would be promised anonymity and confidentiality. They would also sign a consent form to note their unforced participation in the research.
- Ask for permission to record the proceedings of the interview either by audio recording or video recording. Respect the views of the interviewee.
- Technical and theoretical jargon in the questions would be explained so as to be clear and unambiguous as possible.
- Summaries of what I would have understood as the position of the interviewee would be given. This would give the interviewee opportunity to correct any misunderstandings. This also proved to be a great tool for further probing.
- Interviews were always finished at or before the expected finishing time.
- I would always thank the interviewee for their time.

Initially, the plan was to purchase a voice recorder for the interviews but this was not possible before the start of the interviews. Recording of interviews was then done by a portable mobile smartphone with voice recording capabilities. The records were of clear quality. After the interviews, the recordings would then be transferred to the researcher's Dropbox folder which is a password secured online storage platform.

Please refer to appendix A1 for the interview themes and to appendix A2 for the interview questions.

4.3.2 Transcription of the Interviews

This was by far the most time-consuming and agonising process of this whole study after the setting up of interviews. According to the Braun and Clarke (2006) whilst explaining how to conduct thematic analysis, they stress the point that it is beneficial for the researcher to transcribe their own voice data into text as this assists with the first phase of doing thematic analysis which is familiarisation yourself with the data. The interviews were on average about one hour and it would take about a day and a half to transcribe one interview. Refer to appendix A4 for all the transcribed texts of the interviews. For access to the actual voice recordings, you would have to seek permission with the researcher in person.

4.4 Undertaking Co-Design Sessions

Developing the Affinity Diagrams Using Co-design sessions

This was the last step in the data collection process. Two sessions were conducted with two groups at different times. The first group was a group of public healthcare users. This group was comprised of some general workers who alluded to their need for better healthcare opportunities.

The session with this group went well. It was exciting for the participants to be involved. They appreciated that their voices were being considered in such an important matter and they hoped it would effect to change.



Figure 4-1: Co-design session with public healthcare users in progress

The second group was a group of private healthcare users. Initially, the plan was to get participants that use the same private case as that used for the interviews, in this case, Netcare but this was not possible because getting a substantial number of participants that use Netcare facilities proved to be a challenge. The group was then comprised of participants who generally use private healthcare. This session was also fruitful. The participants who happened to also be researchers enjoyed the creativity sessions and marvelled at the simplicity of methods yet producing rich outcomes. This was evidence of the value of design research methods. Each co-design session produced an affinity diagram that was then analysed to draw from it the themes represented by the issues raised by the participants.



Figure 4-2: Co-design session with private healthcare users in progress

The structure of the co-design sessions was such that at the end, participants would have identified issues around their health data management and come with possible solutions to these issues using the affinity diagrams. They would also design a typical personal health records system that they felt would address the issues that they had identified ensuring that their participation in their healthcare information management was increased. Please refer to appendix A5 for the Co-Design session plan.

Before any of the sessions would start the interviewer would introduce themselves as well as their research interests. The researcher would then explain the plan and sequence of the sessions and what was expected of them. Each participant would then be asked to sign a consent of participation form and also promised anonymity and confidentiality. The ethical considerations discussed in Section 4.1.2.1 of this document would also be considered where applicable.



Figure 4-3: Co-Design session's participants

4.5 Summary of Data Collection

During this phase of the study, data was gathered from the literature, from interviews with the case representatives and from the co-design sessions undertaken with the participants. Care was taken so as to ensure that the research population reflected the South African context of healthcare and the researcher believes this was achieved. This phase also consumed more time than initially anticipated and can be eventually singled out as the phase that delayed completion of the study. Various personal lessons were also drawn from the data collection process.

4.6 Undertaking Data Analysis Using Thematic Analysis

When using thematic analysis as an analysis tool, the overall body of data collected for a particular research is called the data corpus. For this study, this includes the interview scripts of both the interviews done at the public case of the study as well as those done at the private case of the study. This also includes the affinity diagrams that were created during the co-design sessions as well as policy documents and literature. The data set is all the data in the corpus that will be used in the analysis of any particular analytic interest. A data item refers to an individual piece of data collected. This might be a picture or an interview script. This then brings us to the last component which is called the data extract. This refers to an individual coded piece of extracted data that would have been identified in a data item.

Having prepared all the corpus of the research study, the actual process of analysing the data was then undertaken. First, the researcher familiarised himself with the data items by reading the scripts and other materials. Care was taken at this stage to ensure that the

researcher understood what was being implied by the interviewees and where the need was clarification was sought from the original source.

4.6.1 Generating Codes

The interview transcripts were then coded. Coding is the initial process in thematic analysis where codes are identified within the scripts that are even in the smallest sense relevant to the research questions, ideas and underpinnings of the study. A challenge encountered during this phase was that often the researcher would notice codes in the information scripts that were not per se connected to the research questions though they seemed relevant. These would be captured and in the next phase of searching for themes they would be discarded if not speaking to the research problem. This was done to maintain consistency with the theoretical approach (top-down) of thematic analysis.

The process of coding scripts to identify the data extracts was such that data extracts were highlighted using the MS Word feature of commenting text. The image below shows an example of a coded script. The codes are the comments on the far right column.

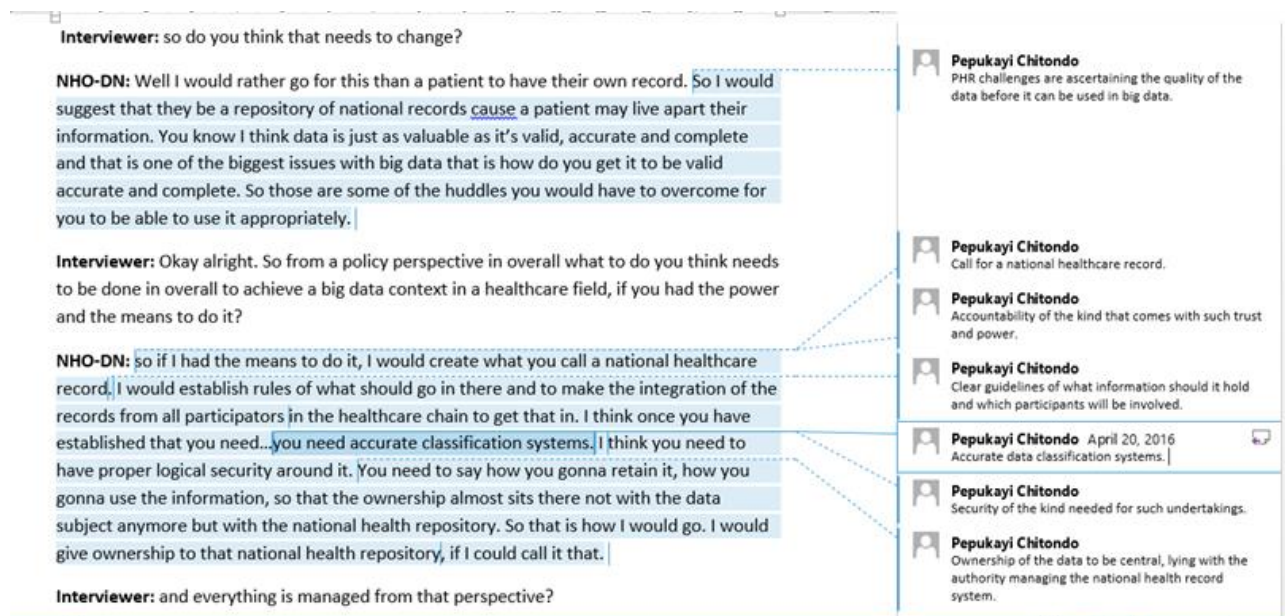


Figure 4-4: Data extracts on a coded script

The image below shows a data extract (left column) and what it has been coded for (right column).

NHO-DN: Well I would rather go for this than a patient to have their own record. So I would suggest that they be a repository of national records cause a patient may live apart their information. You know I think data is just as valuable as it's valid, accurate and complete and that is one of the biggest issues with big data that is how do you get it to be valid accurate and complete. So those are some of the huddles you would have to overcome for you to be able to use it appropriately.

Pepukayi Chitondo
 PHR challenges are ascertaining the quality of the data before it can be used in big data.

Figure 4-5: Data extract and subsequent code

A total of 407 codes were generated for 407 data extracts.

4.6.2 Generating Themes

This phase of the analysis phase constituted of first building a list of all the generated codes which were existing in the different scripts. MS Word 2013 is equipped with a feature to copy all the comments on a document. This was very helpful in this case as it enabled a collation of all the different codes from the different data items into one long list of codes.

This having been achieved, a run through of all the codes and trying to group them into similar categorisations produced an initial thematic map as shown in the diagram below.

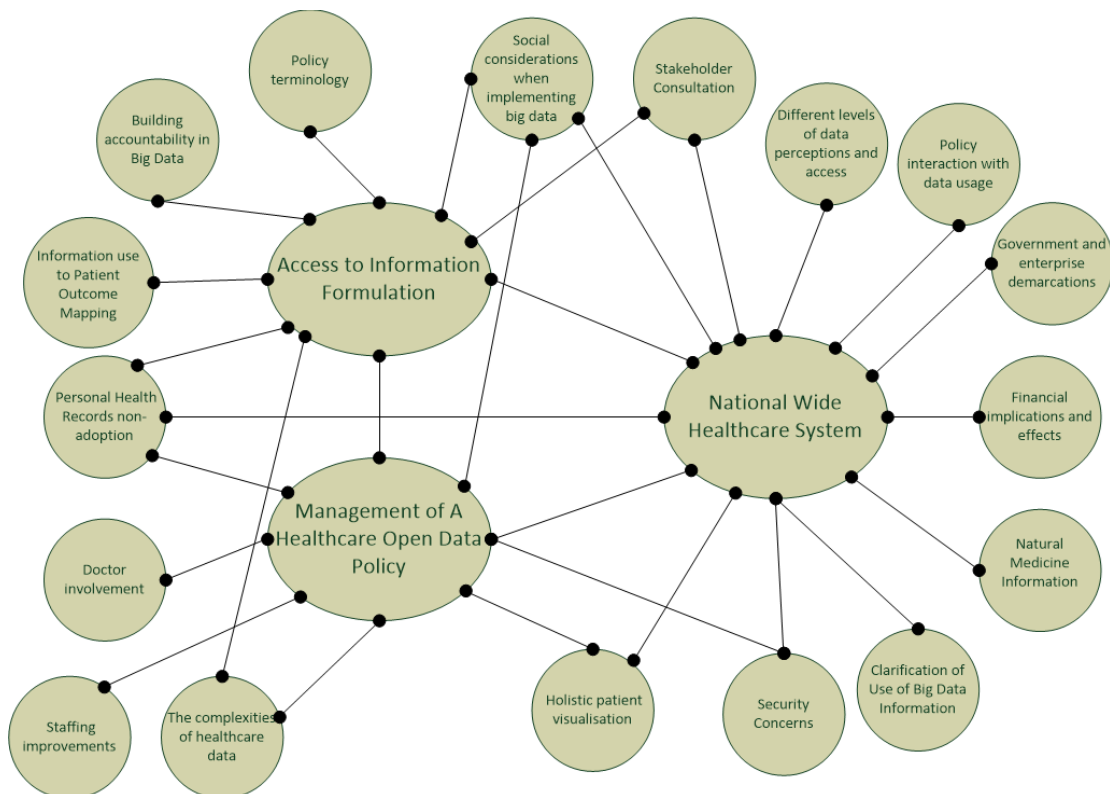


Figure 4-6: Initial Thematic Map

Onto this initial thematic map where the themes generated from the co-design sessions were included as well. The map shows three main themes which are interwoven together with seventeen sub-themes. At this point some of the themes had partial reflections of other themes as would be expected (Braun & Clarke, 2006). This, therefore, necessitated further refinement of themes to ensure unique categorisations of the data within the themes. A final thematic map was then produced with two main themes and four sub-themes. Below is a visual representation of the final themes as reflections of the analysis of the studies data.

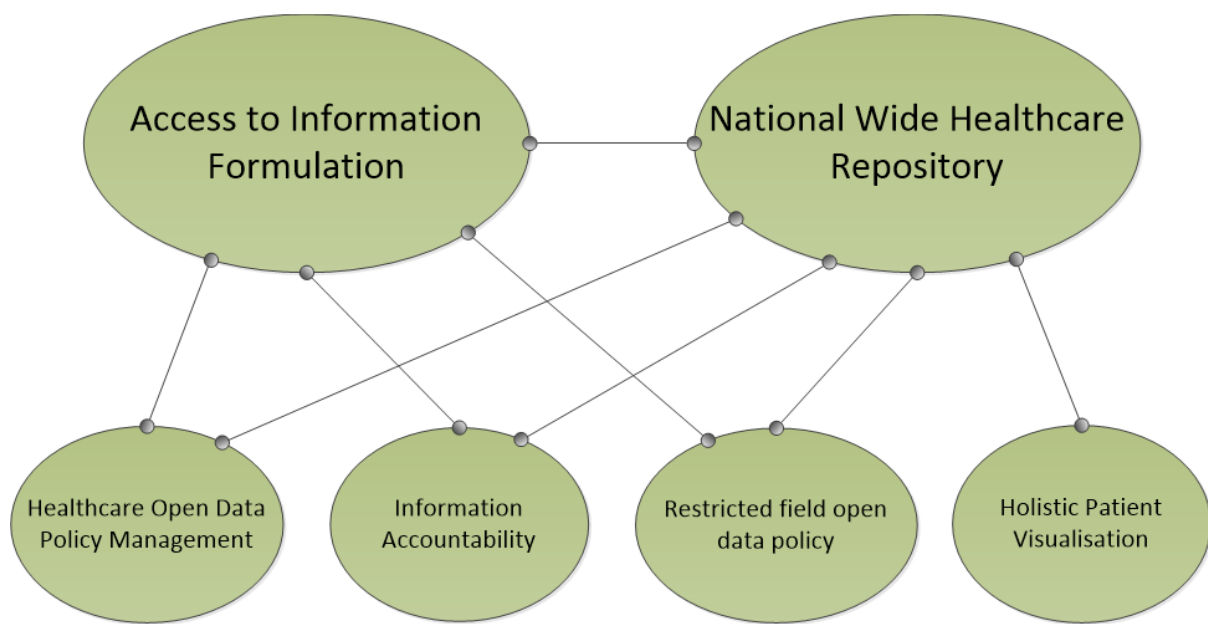


Figure 4-7: Final thematic Map

4.6.3 Summary of Data Analysis

Reflecting on the use of thematic analysis it can be concluded that the method is highly abstract. Care should be taken as the interpretation of the data sets can be very varied depending on the perceptions of the analyser or researcher. It is easy to unintentionally ignore important leads when coding and formulating themes. Due diligence was taken to guard against these possibilities as much as possible in this study. The concepts of internal homogeneity and external heterogeneity (Patton, 2003b) were also applied to ensure the credibility of the themes as separate, distinct and non-conflicting.

Below is a table which shows how these themes map to the research sub-questions.

Table 3-2: Map of developed themes to the research questions they are linked to

Research Questions/Themes	Access to Information Formulation	National Wide Healthcare Repository	Holistic Patient Visualisation	Restricted Field Open Data Policy	Information Accountability	Healthcare Data Open Policy Management
How are healthcare policies influencing data use trends like big data and personal health records?	X	X	X		X	X
How are personal health data and big health data policies affecting data use?	X	X	X		X	X
How are healthcare data policies factoring into consideration foresighted changes in data use strategies?	X	X	X		X	X
How are health data policies providing support and a foundation for big health data and personal health data policies?	X	X			X	
What is the nature of current data policies in the context of healthcare data use?	X		X	X	X	X
How do current personal health data and big health data policies promote information access and sharing?	X				X	X
What are the differences that could be existing between current healthcare data policies and those of big health data?	X		X		X	X
What framework can be used to improve healthcare data policies in the context of big health data?	X		X	X	X	X

4.7 Reflections and Experiences

After having conducted the research as explained in this chapter, the researcher believes he has now gained the skills of a researcher. When this research journey started, there was very little credible research that the researcher could have been noted as being a part of. That was an intimidating reality. However after undertaking the events narrated in this chapter much if not all of this changed.

Some specialised people skills were required to be able to sufficiently organise the interviews and co-design sessions. Those are skills that one is never taught. The researcher also had to exercise much patience and often despair would set in as plans in practice would not be possible as anticipated. This was especially true during the organising and setting up of interviews with external participants from the public and private cases.

The realisation that the chosen method of analysis would require verbatim transcription of the interviews preferably by the researcher was also not easy. After transcribing a few minutes of the first interview, the researcher decided he could not continue. It was a painful process. However, after encouragement from the other researchers and friends, that all tiring process had to be undertaken. The researcher feels there should be a better less painful way of passing this stage. Perhaps more research is needed in this area.

When writing time came, thought was given to which writing case (active, passive) the researcher wanted to pursue. Maintaining the chosen style was not always easy and constantly reverts would have to be done to ensure consistency.

The help of the supervisor during these processes was also invaluable. From the financial support to the moral support, if it was not for the supervisor much would not have fallen into place as it did. Much thanks to her. Thanks also go to the URF fund that the institution awarded me that assisted in covering some of the running costs of undertaking the fieldwork for this study.

In conclusion, of these reflections, the researcher notes that they have been impacted positively by undertaking this path of being a researcher. The researcher mingled with relevant people in the field of this research. The researcher is more confident now and the experience can be regarded as valuable and positive.

4.8 Conclusion

In this chapter, the data collection process report has been given. The subsequent process of analysing the data corpus generated was also described and the resultant themes and their relationships were articulated. Two overarching themes were developed from a process that started with over 400 extracted data codes. The need for a national wide healthcare repository and the need for the formulation of access to information standards were the two main themes developed. Four sub-themes were also developed to further encompass the overall feelings of the study's data corpus. These themes set the tone for the discussions that are undertaken in the next chapter.

CHAPTER FIVE

DISCUSSION OF FINDINGS

This chapter will articulate the findings of the study as gathered from the previous chapter and subsequently a discussion of the findings (see Chapter 4, Figure 4.7: Final thematic Map). The previous chapter ended with a thematic diagram showing the main themes and sub-themes incorporating the researchs findings, this chapter will accentuate the principles expressed in these themes and how they fit into the conceptual model guiding the study (see Chapter 2, figure 3). The findings of the study will then be paralleled with the theoretical framework as discussed in Chapter 3 (see Big Health Data Policy Theoretical Framework) to better understand the current position of healthcare data policies in the context of healthcare data use.

The findings result from the triangulation of data collected from interviews with personnel from the study's public case, and the private case; co-design sessions with public healthcare patients, and private healthcare patients as well as data gathered from the literature. Below are the themes that were generated after the data analysis stage:

Main Themes:

- Access to Information Formulation
- National Wide Healthcare Repository

Sub-themes

- Healthcare Open Data Policy Management
- Information Accountability
- Restricted Field Open Data Policy
- Holistic Patient Visualization

The following section will now discuss each theme in more detail.

5.1 Access to Information Formulation

One clear grey area that the study's findings allude to is how manageable data access will be incorporated into healthcare. The study revealed that:

- There is no set legislation and or policies that define how consent should be acquired from patients.
- Politics has a considerable impact on information access.
- There is no clarification of who will be accessing which data and for what purposes.

- There is no concise definition of the different data access levels.
- There is prejudice and stigmatisation and offenders are not brought to the book.
- Double folded access authorisation and approval to information.
- Little audit mechanisms to enforce standards and policies.
- Policy terminology is abstract and should be made understandable.

The next section will now focus on discussions of the above points.

5.1.1 Acquiring Consent

A look at the conceptual framework developed in Chapter 2 (see Figure 2-8, Conceptual Framework) shows the progression of data sources as an embankment to the main health data repositories. Literature suggests that at each and every stage there have to be forms of data use and for this to be possible, consent should have been provided (Blumenthal & Tavenner, 2010; DesRoches et al., 2013). However, it was found in the study that there are instances where secondary data users of patient information had no knowledge whether the data they were using had been consented for. Furthermore, as data moves from one level to the next there is a need for that consent trail of approval to be visible but in some instances, this trail was not evident.

In another instance, the ethical considerations of how that consent had been sought had moral ethics gaps. An example here would be that giving consent was tied to the provision of the healthcare to offer, therefore, it leaves patients with no option but to consent.

The conclusion, therefore, was that there is not any detailed methodology for access to information that recognises the different levels of users of the data across the conceptual model. There are also problems with folded access requirements i.e. having to require consent for the same information from more than one party. This shows a lack of commitment to the regulations pillar of the Big Health Data (BHD) Theoretical Framework as one of its bases is defined access to information. There is, therefore, a need for formulations of standards that recognise this gap and speak to it. This is the case across public and private healthcare. A well-oiled BHD environment requires this kind of a data access structure otherwise patients information will continue to be accessed without proper consent. Also, some organisations will continue to abuse the current system where blanket and indirectly forced consent is rampant.

5.1.2 Politics and Information Access

The study found out that politics, in some instances, is having a negative impact on useful information access. The definition of a policy in itself means that politics will be involved. As this study has chosen to dwell on healthcare policies it was not amazing to find that it revealed politics as having an impact on information access.

South Africa is sub-divided into nine provinces. There are provinces that have provincial governments that operate individually with different budgets and expenditures as well as health goals. The study revealed that information access and sharing is categorised into these provinces. This, at times, results in failed attempts to share and access information across provincial governments.

At the national level, there are some initiatives or stances that the government has embarked on that have also affected seamless access and sharing of information across organisations. One such case is the role of the Competition Commission of South Africa whose Acts, were cited as hampering access and sharing of some health information. The Promotion of Access to Information Act (PAIA), as well as the Protection of Personal Information Act (POPI), were also cited as preventing useful access and sharing of information that could improve health.

It was concluded therefore that some national policies are currently in a position counter to a BHD environment. This is to be the case for both private and public healthcare though it seems that it is affecting private healthcare more than it is affecting public healthcare. There is a need for the formulation of information access policies that transition the political divides of the country. This is key especially at the interaction that happens between the stakeholders themselves in the pyramid of the conceptual framework and as well as with the data systems in use. Failure of such a national policy ecosystem will continue to see efforts between the two healthcare sectors not being harmonised. More importantly, opportunities to synchronise various national programs to support healthcare will be missed.

5.1.3 6 W's of Information Access

One other notable conclusion that has been drawn in the study is that there is not proper documentation of how information access will be done, when, where and by whom. Also, what will constitute BHD information and why are all things that still need to be identified and formulated into policy templates for all. This speaks to all the pillars of the BHD framework and covers a broad range of factors like relevancy of information, patient involvement and a specialised skill set and managing the complexities of a BHD environment. For instance,

how long should information in the BHD environment be kept? Five, ten forty years? These are issues that currently have not been defined across for all and they will need answers before the realisations of the visions of this study. The private sector seems to be more abreast about this issue than the public sector. Various aspects of the health information of the country are continually being overlooked. It might be that irrelevant information might be still being stored and being accessed by irrelevant people and or programs whereas the right data is not being stored or the right people and programs are not being given access to it. The list of possible consequences if these aspects are not clearly defined is endless.

5.1.4 Stigmatisation and Prejudice as Results of Information Access

There is a huge problem that will continuously haunt information access initiatives and that is victimisation of people because of their sensitive health data. The findings of the study study revealed that there is no explicit talk of how this will be kerbed in current health systems and more so in BHD environment.

So as to promote patient involvement and improved data use via a healthcare open data policy, it will be expedient to also come up with measures to significantly reduce the stigma and prejudice around some diseases and health conditions. This will have to be done prior to the adoption of any kind of an open data policy in a healthcare context. The results of such a policy stance that sets out a method to kerb victimisation of patients, has the potential to improve the reception of a healthcare open data policy. Stigmatisation and prejudice were noted as a problem across both private and public healthcare institutions. In fact, it is a nationwide issue therefore a nationwide solution is needed. It will be difficult in the current environment to advocate for open data policies in healthcare as many people would be affected negatively.

5.1.5 Robust Audit Mechanisms

This issue though appearing under this section of access to information formulation, it is extensively discussed in a later section of this chapter (see Section 5.5).

5.1.6 Presentation of Policies

One other finding drawn from the study's data corpus is that current policies are not easily implementable because they are abstract. This has to do with the terminology and jargon generally used to write out policies among other things. The data showed that there is a considerable need for either an extra layer of documentation in-between the users and the policies to allow lay people to be able to use these policies without fear of doing the wrong

thing. This was particularly expressed by public healthcare respondents though some private healthcare respondents also alluded to this need.

A look at the conceptual model of this study shows that this is a necessary element that still needs to be incorporated into healthcare data policies to facilitate use at elementary levels. If data use and patient involvement as envisioned by the BHD theoretical framework is incorporated, then careful thought will have to be put into the presentation of policies to speak to a category of users that may not be academic, professional or literate.

5.1.7 Summary of Access to Information Formulation

In the literature, there is a constant reference to a need for a well thought through modus operandi for information access that promotes and advances access to information yet still promoting the cornerstones of privacy and confidentiality (World Economic Forum, 2014). More thought will have to be given to the formulation of healthcare data policies to embrace the issues presented in the above discussions. It can be concluded that the current setup of policies that have cut across the issue of access can be greatly improved to improve healthcare provision.

5.2 National Wide Healthcare Repository

The data corpus of the study also revealed the following about policy gaps:

- There is a lack of a national patient indexing system.
- There is no enforcement of compatibility structures across systems.
- There is a lack of national reporting structures.
- There are no national algorithms for data interpretation and structures of how they will be formulated and agreed upon, avoiding biases.
- There is no set structure of overcoming provincial government structures.
- There is a lack of a central repository system for all national health data.
- There are no set guidelines for managing the rate of change and disposal of information.
- There is no Personal Health Records.
- Lack of a patient life cycle from birth till death.
- There are no set structures for determining the relevancy of information to be incorporated into such a national system.
- The issue of ownership of data is not clearly addressed by policies.

The next section will now focus on discussions of the above points.

5.2.1 National Patient Indexing System

The current healthcare environment in SA cannot be viewed from a single focal point nationally. This is perhaps one of the weakest links to the healthcare legislation of this nation, i.e. the fact that patients cannot be indexed across private to public entities and across provincial governments.

The lack of such a national indexing system makes it difficult to currently make studies that can be generalised across the nation rightly so because of the data used in the study. Though policy structures set by the government which have been incepted for years recognise this need, this has still not been achieved to date. Both the conceptual model and the BHD framework beg for such a system in principle.

It can be concluded therefore that current policy structures need to legally force such a system into existence. The benefits of the ability to connect all health systems on the basis of the individual patient can have far-reaching positive results. The development and implementation of a national patient indexing system are being called for both by both the private healthcare as well as the public healthcare systems. Its advantages towards better reaching an optimal BHD environment are vast. It is in effect the only real chance of properly ascertaining what the health burden of the nation is and how it can be tackled. A national patient indexing system is a yesterday need for South Africa.

5.2.2 National Reporting Structures

It was discovered that there is currently no regular reporting needs in policies that cut across public to private healthcare. This reporting should ideally be at a national level. Because public healthcare is generally managed by the various provincial governments, reporting from a public perspective is part of the operational and performance requirements of the public healthcare entities. This is not so though for private healthcare who are only required to report to the government in exceptional instances, such as emergencies, disasters, outbreaks and for specific ailments.

The fact that there is no conscious need from the policies to create a national health reporting structure is counter to a functional BHD environment. This lack, therefore, speaks to the BHD skillset pillar of the BHD framework as well as the regulations pillar. The existence of such a reporting structure means the government can be able to adequately predict the health burden of the country across private and public care, facilitating planning, budgeting and resource allocation at a national level. Care should be taken however when

this element is being considered so as not to kill the autonomy of the private sector as it does have its place in improving healthcare.

5.2.3 National Algorithms for Data Analytics

Among the various grey areas of how a national repository of healthcare data is to be implemented, one notable one is the definition of a set of rules that will be used to infer, correlate and interpret South African healthcare information. This has not yet been thought through to be defined in policy at a national level. It is, therefore, a notable gap for healthcare policy makers to address. This speaks to the BHD skillset pillar of the BHD theoretical framework.

Healthcare information is at a different wavelength than other data's and has complexities not common to other information. Its interpretation is also localised as it is dependent on local elements. This means that the algorithms that will be used for analytics of this data will have to this aspect. This is one area that healthcare policies still need to define, i.e. algorithms that recognise context, avoiding biases and different perspectives.

5.2.4 Central Repository of all National Healthcare Data

There is no central repository system for South Africa's healthcare data. Public healthcare data is currently separate from private healthcare data. Within public healthcare, there is a further division at the provincial level, and in some instances, at the district level. With private healthcare, there is virtually no data interaction between the different private healthcare institutions and very minimal interaction with the national repositories. That also is only when required. This scenario is further aggravated by the fact that private healthcare institutions are in direct competition with each other.

The above scenario is not ideal if the view is to make information a key informant of the national health landscape. Patient life cycles are lost or are not visible and as alluded to earlier, the healthcare burden of the nation cannot be efficiently determined. According to the conceptual model of this study and the BHD theoretical framework, there is a need to view the national healthcare information from a single perspective. More research is needed to find ways so as to achieve this single view impeccably. Directive from a policy point of view addressing this is lacking. The data use pillar of the framework would be easily recognised if such a central repository was a reality. So would the regulations pillar as it would be theoretically easier to regulate such a system. It would be perhaps necessary to mention at this point that the envisioned central repository does not mean that the data would have to reside physically at one location.

5.2.5 The Future with Personal Health Records

On Personal Health Records (PHRs), the study found out that in the strict sense of the definition of PHRs adopted for this study, they are not yet in use both in private healthcare as well as public healthcare in South Africa. In private healthcare, there seem to be few forms of what could be termed PHRs with a few changes. There are virtually none such initiatives in public healthcare. Virtually all the individuals that participated in the co-design sessions (both public and private healthcare users) had no tangible experience with personal health records.

This conclusion means that there is minimal patient involvement in the areas of information creation and management as stipulated by the BHD framework. Looking at the conceptual model, interaction with BHD by the individual so as to better participate in taking care of their health is therefore not visible. There are various reasons why PHRs perhaps haven't gotten a hold in the South African context as yet (Kim & Johnson, 2002; Pagliari et al., 2012; Mxoli et al., 2014). Policies can play an effective role in ironing out these issues. They (policies) are not there yet.

5.2.6 Ownership of Healthcare Data

The study found that the approaches taken by public healthcare and private healthcare, to the issue of ownership of patient information are different. Public healthcare tends to view the information it has as owned by it whereas private healthcare tends to understand personal healthcare information as belonging to the patient. This is despite that both claim to adhere to POPI and PAIA.

It will be expedient for a well-oiled BHD environment to accurately settle this matter of ownership of healthcare information at the different levels it's governed it. The solution to be designed to address this should have in cognisance the fact that consent formulation is dependent on ownership clarification. Many of the issues surrounding health data use could be alleviated better formulation of ownership rules. The definition PHRs which still has been globally agreed upon is also dependent on this, so are many regulatory concerns.

According to the BHD theoretical framework the realisation of a strong regulations pillar and increased patient involvement, are potential results of better legislation tackling the issue of ownership. Policies are not totally silent on this issue but there is a need to approach it, perhaps more head on. Here, researcher tends to favour a position that will give controlled ownership to the central system. More about is discussed later in this chapter (see 5.5).

5.2.7 Summary of National Wide Healthcare Repository

Implementing a national wide healthcare system is definitely a huge component of the solution to achieving big data analytics that will benefit the nation by looking at the whole nation's data. Careful thought, consultation and much discussion will have to go into the foundations of this system with the issues above being highlighted. The researcher would like to agree that such a national system is still a far way long if current legislation remains as is.

5.3 Healthcare Open Data Policy Management

The notable findings that stood out under this theme are:

- There is a resistance to the adoption of an open data policy.
- National policies are access-restrictive.
- Policy implementation is poor.
- There is a need for clarification of what constitutes sensitive data.
- A need for policy alignment across the different levels of government.
- Insurance companies are a nuisance to data safety, therefore, a need to manage their interaction with such a system.
- There is a lack of structures to ensure every action of such a system is patient outcome focused.
- There is currently thin partnerships between the private sector and public sector.
- There is no drive that encourages information sharing to be a priority.
- There is no promotion of an open data.
- There is a lack of structures to identify the national health burden of the country.
- There is a need for clarification of how add-on independent systems will be handled.
- There is a huge turnaround of IT personnel in health.
- There is a need for ensuring that the right ICT human capacity and skill is available for this system.

5.3.1 Adoption of an Open Data Policy

It can be concluded that the adoption of an open data policy in the context to healthcare data is yet still a long way off. This is the case both in private healthcare as well as in public healthcare. The implementation of open data policies is reliant on a number of issues that have to be addressed prior, one of them being a concrete access to information formulation concept (see 5.1).

There is a lack of consciousness about the open data policing itself as well as the issues that surround it. This revealed that there is little or no reflection at a policy level about opening up data for more use. There is no way, then, that this concept of thinking may look adaptable especially considering privacy and confidentiality. Were proposed, very few respondents thought of an open data policy as profitable let alone manageable. This was true for both public healthcare respondents as well as private healthcare respondents.

According to the BHD framework, an open data policy would benefit the data use and the patient involvement pillars. The regulations pillar, however, would need some revolutions to be congruent with open data policies. These foreseeable conflicts would have to be addressed by policy as they are currently not yet doing this, both at private and at national levels.

5.3.2 Mapping BHD Activities to Patient Outcomes

Ensuring that every activity/event/analysis that is performed in the BHD environment will benefit the end patient will be a key component of a well-oiled BHD environment. The necessity of this is highlighted by the fact that there is a high possibility of non-patient-related activities benefiting from BHD activities. An example would be insurance and financial services companies who might use BHD results for their own pursuits. This includes healthcare organisations who might use BHD activities to reduce their operational costs though not reducing the cost to the patient or healthcare in general. There is clearly a need, therefore, for the formulation of rule sets that will map all events in the environment to possible patient outcomes. There is a lack of such rule sets in current legislation and policies across the healthcare fraternity in general.

This phenomenon speaks directly to the data use, regulation and BHD skillset pillars of the BHD theoretical framework. In summary “patient outcome” “meaningful data use” (Hoffman & Podgurski, 2011) structures are not yet laid out from a policy perspective. This is to say that there is no policy framework whose goal is to achieve data utilisation that is focused on improving the results of patient visits to hospitals. Not achieving such a policy structure leaves the door open for big data activities that have nothing to do with the patient which in essence means the main reason for engaging in big data activities is forfeited.

5.3.3 Huge Turnaround of ICT Personnel in Health

It was found that there is a huge turnaround of ICT staff in the healthcare sector in general compared to other sectors. There is a scenario in one of the cases were out of a department staffing of over ten people, only two were left and it was proving difficult to replace the

people who left with new people with the right skills. Various reasons were cited. The strongest was poor remuneration compared to market-related packages. Lack of passion for the healthcare field in ICT professionals was another.

Having alluded to the fact that a BHD skillset would have to be a specialised group, understanding the complexities of healthcare and its interpretation, it will be expedient to turn around this trend. Healthcare policy will need to address this ICT brain drain and embed within healthcare policies solutions to avoid losses of key BHD skills. It will take time to train one passionate individual to become a valuable BHD skilled individual and losing them would not be ideal as it will take the same amount of time to train the another individual who may not be as passionate about health. Policies are currently lacking in this regard.

5.3.4 Partnerships Between the Private and Public Healthcare Sectors

There are currently very limited relations and collaborations between the private healthcare sector and the public healthcare sector. Despite the fact that both theoretically exist for the same purpose, the study found that there is next to none hand to glove partnerships between the two sectors. The problem with the status quo is that there is much harmonisation of efforts that is not being realised that can ensure a more comprehensive coverage of healthcare provision across the population. It was also found out that private healthcare structures feel marginalised by government regulations, policies and efforts which seem to only recognise public healthcare and ignore the presence of private healthcare enterprises.

There are various ways on how inducing partnerships between the private and public healthcare sectors can improve healthcare provision. Often the load on public healthcare structures far more exceeds the capacity of provision that they can accord whereas private structures generally have under-utilised healthcare provision capacity. Partnerships could help in balancing these scales. One other area is that of exchange of information related to management of healthcare facilities where private healthcare generally has a better situation. Sharing of information and resources is another of these benefits.

According to the BHD theoretical framework, stronger partnerships between private and public healthcare would strengthen the BHD Skillsets pillar, the data use as well as the regulations pillars. Though the envisioned NHI has the potential of fostering such relations it was concluded that there is still much to be desired of healthcare policies in relation to harmonisation of private and public healthcare efforts.

5.3.5 Prioritising Information Sharing

Both private and public healthcare were found not to be promoting information sharing with other organisations for the purposes of gaining a better understanding of that information. It was found that most of the times when information is shared, it will be on a request basis and rarely for the purposes of seeking healthcare knowledge from the information. The most common reason for a request to share was found to be for litigation purposes. This means that there is currently little if no policy infrastructure to nurture sharing of information for the reasons of knowledge management. This is a serious problem. What this means is that there is no culture of information sharing which in-turn leads to data siloing and more importantly a failure of realising the big story that the data tells in its entirety.

It is beneficial in a BHD environment to cultivate a culture of conscious sharing and this is something that policies need to find a way to forge. Though not completely silent about this concept, there is a need for further reinforcements of policies to better address this lack of conscious sharing of information. The data use and regulations pillars of the BHD theoretical framework allude to this need. The conceptual model also shows this need by the constant feeding into the BHD pool by the government, health organisations, researchers as well as individuals.

5.3.6 Handling Add-on Infrastructure and Systems

The number of independent ICT resources and infrastructure that are available today have the potential to improve healthcare, especially in the area of cost reductions. However, how they are managed should be adequately laid out. How they will be interacting with the main systems and other add-on resources which potentially are structured differently are all areas that still need to be addressed. The skills of any ICT manager to be able to use these presumably free resources to cut costs and improve operations efficiency are in high demand but problems are prone to such setups because of the underlying differences in the platforms of these add-on resources. Examples of such add-ons are the Bring Your Own Device (BOYD) concept, cloud computing, Virtual Private Networks (VPN) as well as social media platforms.

Careful thought will have to be applied in formulating some guidelines of usage of such infrastructure in a healthcare context otherwise vulnerable, sensitive and confidential health data might be made accessible to unwanted users. This aspect alludes to the regulations pillar of the BHD theoretical framework as a lack of foresight in its management can have serious security repercussions.

5.3.7 Summary of Open Data Policy Management

An open mind discussion of the management of how an open data policy could be implemented in healthcare context leads the researcher to summarise this section by saying that, current healthcare policies are not yet ready to handle the offshoots of implementing a healthcare open data policy. It is of great benefit for the healthcare policy maker to incorporate solutions to the issues that have been highlighted in this section to prepare a solid foundation for open data policing in the healthcare sector.

5.4 Restricted Field Open Data Policy

The huge resistance to the adoption of an open data policy was something that was reflected by all respondents of the interviews. Some did not think it was a good idea at all, yet some saw its value in a big health data environment. It became obvious in these discussions that the current understanding of such data usage would be difficult in the healthcare context. The researcher found out that a modified version of an open data policy, with healthcare considerations, would be an acceptable alternative. One particular respondent coined the phrase, 'Restricted Field Open Data Policy' and this phrase has been adopted by this study to denote a healthcare contextualised open policy.

5.5 Information Accountability

- There is a lack of auditing procedures to ensure proper data use, integrity, and that policies and standards are being followed.
- There is a need for market-related remuneration of Big Data personnel that will ensure security accountability.
- There is a need for clarification of punishments to abusers of data.
- There is a need for mapping of data use initiatives to clinical outcomes to ensure acceptable use.
- There is no encouragement of global research partnerships which lean towards public accountability.

5.5.1 Auditing Mechanisms

Seeing that such an information environment will require constant tracking to see whether events and results are occurring as expected, it was also found that the policies used to manage healthcare would now need to be revised to ensure accountability of all BHD events. Auditing structures and mechanisms currently in use are not sufficient to adequately manage the envisioned BHD environment. This is something that will need to be also resolved by policy. All pillars of the BHD theoretical framework are set to benefit from such improvements to healthcare legislation.

5.5.2 Legal Consequences of Information Abuse

Those who will be found themselves on the wrong side of the law as far as BHD activities are concerned, i.e. mismanagement of information, and thrusts that have nothing to do with the patients but self, will have to be held accountable. There is a great need for the formulation of worthy consequences of the breach of conduct that BHD personnel or any other people who will interact with the BHD systems will do. Policies were found to be needing to pay more attention to this aspect.

5.5.3 Summary of Information Accountability

Embedded in the issue of information accountability is also the responsibility of ensuring that BHD personnel are able to perform their duties without the risk of being enticed one way or the other to breach their codes of conduct. This alone will greatly improve efforts on enforcing information accountability in a big health data context. On top of this, the kind of audit structures worthy of a well-oiled BHD environment as well the legal response to negative results of such audits are issues that policies are currently not fully addressing.

5.6 Holistic Patient Visualisation

- A huge part of the population is uninsured and this impairs continuous care because of the way the public care is structured.
- There is a lack of accurate population figures. This is also aggravated by porous borders.
- There is a lack of coherent continuous recording of all episodes of care for patients.
- Lack of a patient life cycle from birth till death.
- Lack of linkage between public and private healthcare systems and the overall picture is being missed out.

5.6.1 Continuous Care

The idea of being able to view the entire record of all the patient's health encounters across various healthcare systems has always been treasured in healthcare but it has not been realised in the context of this study. It was also found that current policies do not address the problem with the lack of ability to provide seamless continuous care, especially across private and public healthcare services. Often when a patient is transferred to another facility or has to switch doctors the process of transfer of their previous health information is not clear and at times even impossible. Patients themselves have also been noted as having a tendency of changing clinicians over time (Bayer et al., 2015). This then impairs visualisation of the care given to the patient over time.

Continuous care is a key feature of an optimal BHD environment. To properly analyse and summarise data with meaning its bigger picture (continuous care) should be available therefore to be provided for by the different healthcare systems. This is an area that policy solutions will have to be developed.

5.6.2 Patient Life Cycle

Closely related to the issue of continuous care is the issue of being able to view a patient's life cycle from birth till present or death. This is also a key feature of BHD and likewise, its importance is not yet reflected from a policy perspective. This does not only speak to the data use pillar of the BHD theoretical framework but it also speaks to the patient involvement pillar.

5.6.3 Summary of Holistic Patient Visibility

To be able to have a full map of the patient as envisioned by a BHD environment, continuous care should be enabled by healthcare systems. This is a feature that is not common across healthcare systems, therefore, there is a need to establish this commonality. Closely tied with this is the ability to view the life prints of all care episodes that a patient will have in his/her entire life. In both private healthcare and public healthcare, a holistic view of the patients data did not exist. This can have huge benefits for building trends that have proof from the data. Such a holistic view of the patient still has to be founded in healthcare policies. Holistic patient visualisation will only be possible when private healthcare and public healthcare are somehow connected, another area that healthcare policies should still focus on.

5.7 Conclusion

In conclusion, having discussed the findings of this study, one can reflect upon them and say that most of the issues raised in this discussion still need solutions. Various issues were looked at ranging from access formulation to a national healthcare system as well as embedding accountability into all these areas of data use. The government's role in providing a suitable environment for these elements of a big health data setup to flourish was also touched upon. A well-oiled BHD environment was also painted using the BHD theoretical framework as a guiding lens and a conclusion can be drawn that for the BHD environment to flourish, there is a serious need for healthcare policies to take a paradigm shift to their approach to data management. A more open usage environment is required and this is not yet reflected in the current policies. There is also little patient participation encouraged by the current setup of the healthcare environment and its policies and this should be seen to change so as to support a homeostatic BHD state. As a general principle

coming out of the study, more reforms need to be done in the public sector so as to reflect the attributes of a big health data environment, as compared to the private healthcare sector which seems to be ahead in preparing for the foreseen data use paradigms.

CHAPTER SIX

CONCLUSION

In this chapter, a bird's eye perspective of this research study is presented before it is concluded. A brief summary of the literature that formed the basis of this study is presented. The aim of this study is presented and the methods used to investigate the research problem are also considered. The outcomes of this study are summarised. These include a proposed BHD theoretical framework and the thematic map developed to represent the findings of this study are also indicated. Finally, some future research directions are suggested. The researcher hopes that if the suggested research directions are pursued, there can eventually be tangible solutions be provided around the issues highlighted in this study.

6.1 What the Literature Had Been Saying

The basis of pursuing this study was informed by the following information that literature had provided:

- The amount of healthcare information being collated is increasing rapidly by the day (Damle, 2014; Hoffman & Podgurski, 2013b). (see Section 2.1)
- There is a need to refocus efforts from collating the data, to finding ways of how healthcare data can be used to progress healthcare (Blumenthal & Tavenner, 2010; DesRoches et al., 2013). (see Section 2.4)
- The use of healthcare data is being hampered, among other things, by policies which are not aligned with the envisioned usage (Page, 2014; Ohno-Machado, 2014; Pasquale, 2013). (see Section 2.3)
- There is a need to readdress the permeating idea across health policies to achieve optimal use of healthcare data (Hripcsak et al., 2014; Keen et al., 2013). (see Section 2.6)

6.2 What the Research Problem Was

Having looked at the issues presented above, the research problem that needed exploration was then summarised as follows:

There seems to be a mismatch between the goals of current healthcare data policies and those of big data in healthcare. This is negatively affecting research efforts in healthcare by hampering the adoption of big data technologies for better data utilisation.

Having understood the matter at hand of how the current orientation of healthcare policies was hampering data utilisation efforts such as big data the researcher then made it the aim of this study to understand the characteristics of current healthcare policies in relation to promoting data utilisation. The objective was to be able to answer the question of the validity of the research problem intelligently and if so what then could be the probable route to address this anomaly. The researcher decided to focus on big data as the data utilisation feature in concern and then personal health records and patient data, in general, were to be the data features in concern. All these features were to be applied from a healthcare context.

6.3 Methods Employed in the Research

The research was based on two case studies, one from the public healthcare sector and the other from the private healthcare sector. This is a deliberate intention to represent the healthcare platform of South Africa. Semi-structured interviews were conducted with representatives from both cases and in the spirit of embracing design methods of research, co-design sessions were conducted with patients that use both the public and private healthcare fraternities. (See Section 3 for fuller details on the methodology and theoretical underpinnings adopted for the study).

Thematic analysis was used to group the resulting leads of the research data into two main themes and four sub-themes. The diagram below shows these themes in the form of a thematic map. (See Section 4.2)

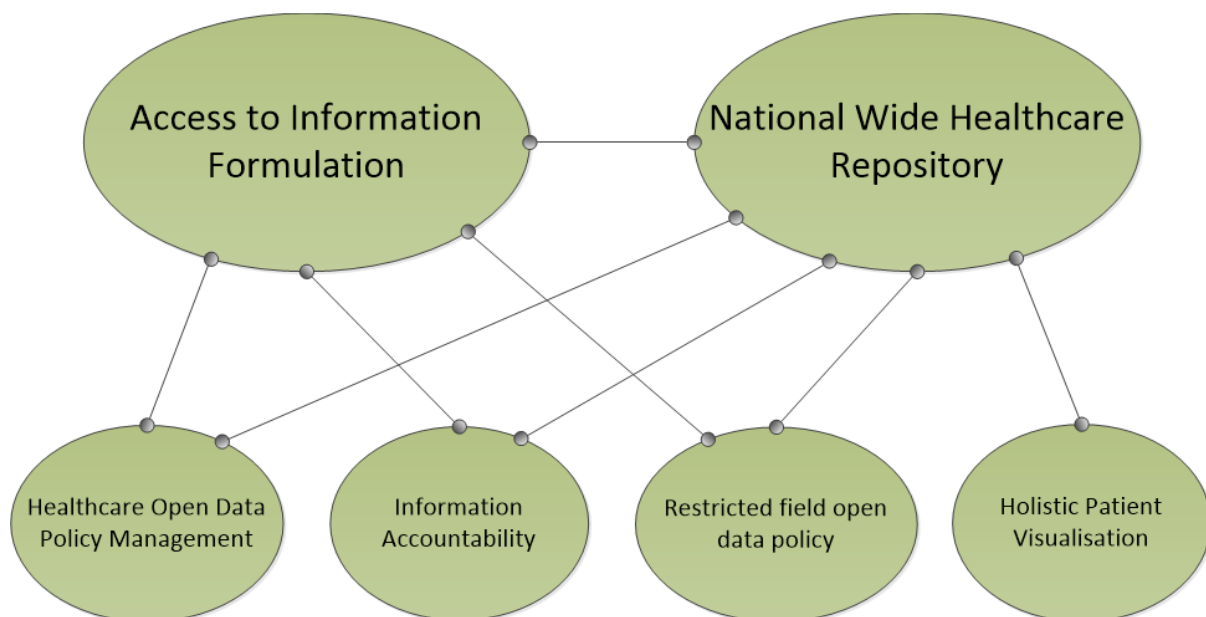


Figure 4.7: Final thematic Map

6.4 A Framework to Develop a Big Health Data Environment

Because the field of research in big data in a healthcare context is still fairly new and underexplored as yet (Vayena et al., 2015), it became necessary to develop a theoretical framework to propose as a blueprint for developing enhanced data utilisation infrastructure. The term coined to represent this infrastructure for this study is the Big Health Data (BHD) Environment. This framework was developed on the basis of applying big data in the healthcare sector, as identified in the literature. For the research’s findings to be put into perspective, this framework was used as a lens to analyse the findings of the study and draw conclusions and results. Below is the proposed BHD theoretical framework.

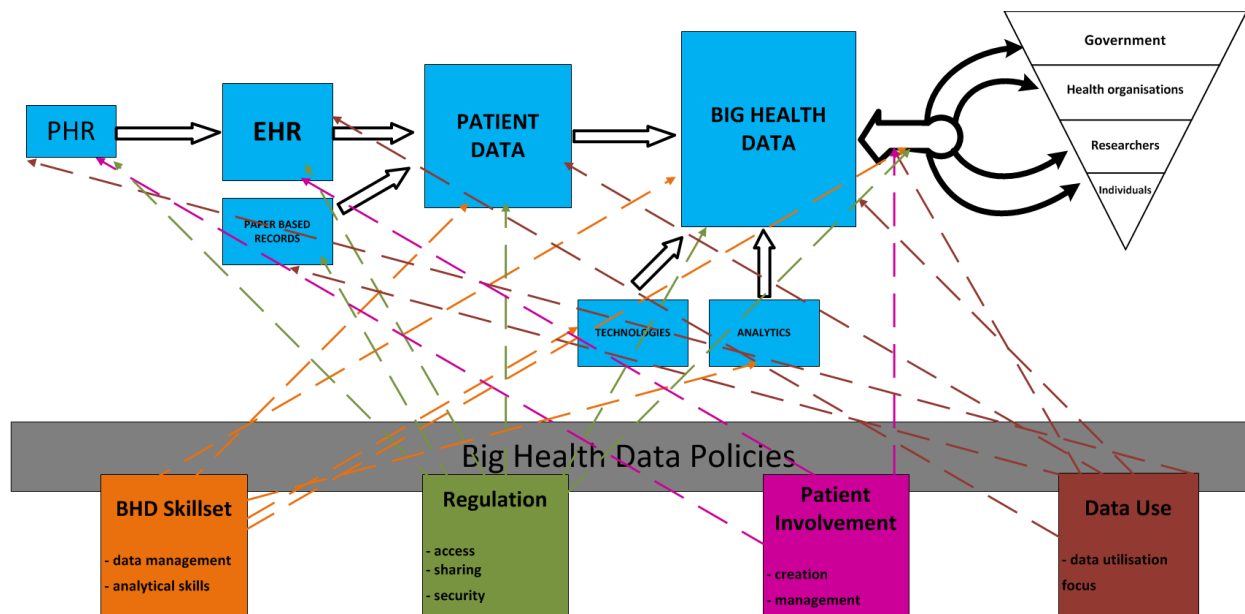


Figure 2-2: Big Health Data Policy Theoretical Framework

Each pillar represents a key feature of big health data policies that would need to be developed so as to support the BHD structure above it appropriately. (See Section 2.3.7)

6.5 Summary of Findings and Results

A summary of the key findings of the study are as follows:

- There are various access issues that healthcare policies are not quite in a position to address as yet. These include how consent is obtained to use health data for secondary uses; the role of politics in the definition of healthcare policies; reducing

stigmatisation and prejudice associated with health data; and the presentation of policies in a format that is interpretable at all levels of literacy. It was seen that current policies do not sufficiently address these issues, therefore, a need was identified to establish a platform that will focus on access to health information formulation to address the anomalies identified above. (See Section 5.1)

- It was found that the need to design and implement a national wide system of storing all the nation's healthcare data is a pre-requisite of a running a credible functional BHD environment. Currently, there is no policy infrastructure available to support and enforce such a system. There is, therefore, a need for national health policies to provide for such a system as this is one of the weak points of the current national healthcare system. Various elements will also have to be considered here and a policy framework developed for them as well. (See Section 5.2)
- It was also found that the implementation of open data policies will be difficult to manage under the current policy constraints. Various improvements and changes would also have to be factored into healthcare policies to enable them to embrace the concepts brought about by opening data for uses as envisioned by proponents of the open data policies. (See Section 5.3)
- Closely related to the above point is the fact that because of the complexities associated with health data, a healthcare contextualised open data policy paradigm would have to be developed. The study adopted a name for such policy, 'Restricted Field Open Data Policy'. The name brings about the key aspects of a restrictive or controlled open data environment. (See Section 5.4)
- The envisioned BHD environment will require robust information accountability measures that are not yet reflected in current healthcare policies. Auditing mechanisms and the legalities of breaches of trust and codes of conduct agreed upon will have to be developed to support the kind of utilisation that is envisioned by big health data. (See Section 5.5)
- It was also found that a key element was missing from current healthcare policies; that which emphasises the visualisation of the patient as a whole. This element incorporates the need to provide continuous care for the patient and the ability to view a patient's whole life cycle from birth till date or death. Current policies are far from needing to represent health information from this dimension. (See Section 5.6)

A concluding statement on these findings is that current health policies are misaligned to the needs of a BHD environment as noted by the above findings and results. They are far from the position that they are to adopt so as to improve data utilisation and promote managed access to health improvement for the benefit of improving the patient and how healthcare is

delivered. There is a mismatch of the goals of current healthcare data policies and that of a BHD environment. How this mismatch will be resolved should be the immediate concern of any futuristic health informatics. The research showed that data policies play a huge role in enabling data utilisation. If they are not properly aligned to the needs of the patient and those of the healthcare professional, they have the ability to counter effective usage of healthcare data. Therefore the underlying nature of current policies needs to shift towards a stance that is more tolerant of easy access and use.

6.6 Research Questions Revisited

This Section revisits the research questions that guided this study and provides an overview of how they were tackled and answered.

6.6.1 RSQ1.1: How are personal health data and big health data policies affecting data use?

The literature reviewed showed various gaps between how policies of PHRs and BHD were tackling the issue of data utilisation. Firstly the literature did not reveal that they were adequate PHRs and BHD policies already in use. Where policies existed they seemed to be at loggerheads with the overall idea of data utilisation. The study's findings and results concurred with this preliminary observation from the literature. Therefore the study concludes that data use is being hampered by poorly designed PHRs and BHD policies as well as an outright lack of these policies.

6.6.2 RSQ1.2: How are healthcare data policies factoring into consideration foresighted changes in data use strategies?

From the literature's perspective, organisations seem to be aware of the fact that their future successes are dependent on them harnessing the power within their data. The how part is where the problem lays. The study's results showed that healthcare organisations are shifting heavily towards EHRs which in turn has led to a huge boom in the amount of digital healthcare data being collected. Healthcare organisations are also beginning to have data analyst teams and activities to ensure that they maximise their use of data. However much can be done if we would compare healthcare data policies with say business data policies. So it can be concluded that healthcare data policies are beginning to place a value on data, its collection, storage and subsequent analysis for better performance. Various technical and ethical challenges still need to be overcome.

6.6.3 RSQ2.1: How do current personal health data and big health data policies promote information access and sharing?

The literature reviewed showed various challenges affecting healthcare information access and sharing. Some policies were found to be restricting access to and sharing of health information. The lack of awareness and gusto to the adoption of open data policing as noted in the study's findings also highlight the challenges that will have to be overcome so as to promote adequate healthcare information access and sharing as required at big data level. It can then be concluded that current personal health data and big health data related policies are not adequately promoting healthcare information access and sharing.

6.6.4 RSQ2.3: What framework can be used to improve healthcare data policies in the context of big health data?

A suitable framework that addresses the needs of big health data design and development was not found in the literature reviewed. This study came up with a big health data theoretical framework (see Section 2.3.7) that highlights the key areas that policy needs to address so as to facilitate a functional big health data environment. The foundation of the framework is based on the issues highlighted in the literature as being imperative for success with big health data.

6.6.5 Main Research Question 1: How are healthcare policies influencing data use trends like big data and personal health records in South Africa?

Investigations of literature and of the study's collected data and findings revealed that personal health records are still a long way to becoming widely relevant data sources for both public and private healthcare. In public healthcare, there were almost zero instances of their occurrence in policies and structures of their operations. Private healthcare also showed very little support of personal health records in its policy structures.

On the other hand, there was found to be various mini big data initiatives with both private and public healthcare institutions. For big data to truly yield big data results, especially in the healthcare context, there needs to be a harmonisation or an amalgamation of not only all big data efforts at public healthcare level but also those at private healthcare level. This is not yet reflected in any policy structures at all levels. It can then be concluded that current data policies are influencing big data to a certain extent which could be improved and that current data policies have not yet influenced personal health records adoption and use.

6.6.6 Main Research Question 2: What is the nature of current data policies in the context of healthcare data use in South Africa?

Optimal data utilisation is business support function that can only be fully realised with the support of all stakeholders involved. This includes the health organisations including the various components within it, the patients as well the civil authorities governing the jurisdiction where the data use activities will be carried out. In the South African context, the South African government's structure and policy environment have elements that make data utilisation efforts complex. Healthcare organisations in themselves at public level as well as at private level were also found to be having data policies that have strong ties to the protection of data rather than to its use, upholding the principles of privacy, confidentiality and consent at the expense of data use. Protection of information in of itself is not a bad thing. The issue was found out to be the method to which this protection is gone about, i.e. secondary uses of information are almost totally stifled out in the process of data protection. It can then be concluded that current health data policies are anti-data use though this is not their motive but their side effect. Adroit policies are needed that will embrace both protection and data use combined. Other ethical considerations will have to be considered as well.

6.7 Research Limitations

The study acknowledges the following limitations that affected the study:

The study was limited to one public case and one private case. A broader case population might have strengthened the research findings.

The study had a time constraint that was determined by the amount of financial resources available. Funding was very limited.

The study was limited to the South African context. A more global representation would have enhanced the study's findings and outputs.

The study was focused on big data elements and more often particular information on this subject matter was not easily available. There is few peer-reviewed and tested theoretical underpinnings that are available in this area and this may have affected the course and ending of the study.

6.8 Evaluation of the Research

There are seven principles that are commonly used to evaluate the validity of interpretive studies (Klein & Myers, 1999). These principles are a guide providing the fundamental aspects that have to be incorporated into an interpretive study as the one that this research followed. The next section will explain these principles and also show how they are reflected consciously and unconsciously in the study.

6.8.1 The Fundamental Principle of the Hermeneutic Circle

This principle is, in essence, the springing platform of all the other seven. “The principle suggests that all human understanding is achieved by iterating between the interdependent meaning of the parts and the whole that they form” (Klein & Myers, 1999: pg 72). The study employed this principle during the literature analysis stage as well as the data collection and analysis stage. Each piece of data considered would be given meaning in relation to the rest of the corpus of the study i.e. the literature as well as data collected. This principle also was observed in the formulation of what eventually became the themes of the study.

6.8.2 The Principle of Contextualisation

This principle “requires critical reflection of the social and historical background of the research setting, so that the intended audience can see how the current situation under investigation emerged” (Klein & Myers, 1999: pg 72). The methods employed for the study observed the needs and demands of the African context and even more, of the South African context of healthcare. There is a whole section in the literature analysis (see Section 2.5) that is dedicated to contextualising the study. The research questions were also geared to observe the context and background of the issue from various angles including ethics, socio-economics, politics and ICT literacy. All the participants interviewed and those who participated in the co-design sessions responded in terms of their understanding of the context of the concepts. The context will, therefore, influence their responses and participation.

6.8.3 The Principle of Interaction Between the Researchers and the subjects

This principle “requires critical reflection on how the research materials (or “data”) were socially constructed through the interaction between the researchers and participants” (Klein & Myers, 1999: pg 72). The research employed semi-structured interviews to collect data and the researcher had personal interaction with the participants during the interviews. The research also employed interactive co-design sessions (see Section 4.1.3) as part of its data

gathering activities. The researcher made an effort to truly reflect on the answers during the coding to derive the right meaning. This clearly shows a reflection of this principle.

6.8.4 The Principle of Abstraction and Generalisation

This principle “requires relating the idiographic details revealed by the data interpretation through the application of principles one and two to theoretical, general concepts that describe the nature of human understanding and social action” (Klein & Myers, 1999: pg 72). The study reflected this principle by adopting a research strategy of multiple case studies that fully represent the healthcare strategy currently being employed by SA. Generalisation of the study’s findings cannot be made with the real limitation that the study is also a cross-sectional study. Abstraction was obtained through the thematic analysis where similar codes were grouped to form the abstractions. This was done iteratively and across the cases.

6.8.5 The Principle of Dialogical Reasoning

This principle “requires sensitivity to possible contradictions between the theoretical preconception guiding the research design and actual findings (“the story which the data tell”) with subsequent cycles of revision” (Klein & Myers, 1999: pg 72). The researcher would like to admit that this was not an easy principle to apply because they had a lot of preconceptions prior to the start of the study. Some were supported by the study though some were in parallel to the research findings. However, discussions on the findings reflected the views of the data and not of the researcher.

6.8.6 The Principle of Multiple Interpretations

This principle “requires sensitivity to possible differences in interpretations among the participants as are typically expressed in multiple narratives or stories of the same sequence of events under study” (Klein & Myers, 1999: pg 72). This principle had to be applied various times as participants would reflect different notions of the same concept. In these scenarios, the researcher would not discard one or either of the interpretations but would consider all in the data corpus of the study.

6.8.7 The Principle of Suspicion

The principle “requires sensitivity to possible” biases” and systematic “distortions” in the narratives collected from the participants” (Klein & Myers, 1999: pg 72). This principle was also employed in the study. Scenarios dealing with narratives of subordinates speaking about their superiors during the interviews would many times require the intervention of this principle.

6.9 Further Research

As the last word, further research areas that could follow up this research that have been identified during this study are: that more research is needed to

- Develop a national patient indexing system that cuts across the public sector and the private sector;
- Further investigate the concept of a restricted field open data policy for healthcare;
- Look at the incorporation of personal health data into the mainstream healthcare information systems;
- Further investigate how national policies can be used as a tool to facilitate improved data utilisation.
- Further investigate the validity of the BHD theoretical framework developed in this study.

REFERENCES

- Adler-milstein, J., Daniel, G., Grossmann, C., Mulvany, C., Nelson, R., Pan, E., Rohrbach, V. & Perlin, J. 2014. *Return on Information : A Standard Model for Assessing Institutional Return on Electronic Health Records*.
- Anon. Evidence-based Practice Key Concepts In Nursing - Credo Reference.
- Anon. 2013. Robert H . McKinney School of Law Legal Studies Research Paper No . 2013 - 04 Nicolas Terry BIG DATA *.
- Appel, J.M. 2008. Why shared medical database is wrong prescription - Orlando Sentinel. *Orlando Sentinel*. http://articles.orlandosentinel.com/2008-12-30/news/OPappel30_1_medical-records-medical-system-electronic-medical 6 November 2014.
- Assunção, M.D., Calheiros, R.N., Bianchi, S., Netto, M.A.S. & Buyya, R. 2015. Big data computing and clouds: Trends and future directions. *Journal of Parallel and Distributed Computing*, 79(80): 3–15. <http://www.sciencedirect.com/science/article/pii/S0743731514001452>.
- Babbott, S., Manwell, L.B., Brown, R., Montague, E., Williams, E., Schwartz, M., Hess, E. & Linzer, M. 2014. Electronic medical records and physician stress in primary care: results from the MEMO Study. *Journal of the American Medical Informatics Association : JAMIA*, 21: e100-6. <http://www.ncbi.nlm.nih.gov/pubmed/24005796>.
- Bayer, R., Santelli, J. & Klitzman, R. 2015. New Challenges for Electronic Health Records Confidentiality and Access to Sensitive Health Information About Parents and Adolescents. *Jama*, 313(1): 29–30. <http://jama.jamanetwork.com/article.aspx?articleid=2088861>.
- Berndt, D.J., Fisher, J.W., Hevner, A.R. & Studnicki, J. 2001. Healthcare Data Warehousing and Quality Assurance. *Computer*.
- Blumenthal, D. & Tavenner, R.N. 2010. The 'Meaningful Use' Regulation for Electronic Health Records. *The New England Journal of Medicine*, 363(6): 501–504.
- Bowden, J. 2014. The 4 V's in Big Data for Digital Marketing. <http://www.business2community.com/digital-marketing/4-vs-big-data-digital-marketing-0914845> 15 October 2014.
- Braun, V. & Clarke, V. 2006. Using thematic analysis in psychology. , 3: 77–101. http://eprints.uwe.ac.uk/11735/1/thematic_analysis_revised_-_final.doc.
- de Bruyn, M. 2014. The Protection Of Personal Information (POPI) Act - Impact On South Africa. *International Business & Economics Research Journal*, 13(6): 2014. <http://search.proquest.com/openview/4ee18ed87793b12df7346b50a25c0a1d/1?pq-origsite=gscholar>.
- Burrell, G. & Morgan, G. 1979. *Sociological Paradigms and Organisational Analysis*. London: Heinemann.
- Castillo, V.H., Martínez-García, A.I. & Pulido, J.R.G. 2010. A knowledge-based taxonomy of critical factors for adopting electronic health record systems by physicians: a systematic literature review. *BMC medical informatics and decision making*, 10(1): 60. <http://www.biomedcentral.com/1472-6947/10/60>.
- Chang, F.S., Nielsen, J. & Macias, C. 2013. data key to quest for quality. *Healthcare Financial Management*, (November).
- Clarke, J.L. & Meiris, D.C. 2006. Electronic Personal Health Records Come of Age. *American Journal of Medical Quality*, 21: 5S–15. <http://ajm.sagepub.com>.
- Cole, B. 2016. What is data governance policy? *TechTarget*. <http://searchcompliance.techtarget.com/definition/Data-governance-policy> 20 January 2016.
- Crawford, K. & Schultz, J. 2013. Big Data and Due Process: Toward A Framework to Redress Predictive Privacy Harms. *Boston College Law Review*, Vol. 55, No. 1, 2014, 55(1): 1–31.
- Damle, A. 2014. An Apple a Day: The Opportunities and challenges of Big Data in

- Healthcare. *Silicon India*, (March).
- Department of Health South Africa. 2012. *National eHealth Strategy, South Africa 2012/13-2016/17*. <https://www.health-e.org.za/wp-content/uploads/2014/08/South-Africa-eHealth-Strategy-2012-2017.pdf>.
- DesRoches, C.M., Charles, D., Furukawa, M.F., Joshi, M.S., Kralovec, P., Mostashari, F., Worzala, C. & Jha, A.K. 2013. Adoption of electronic health records grows rapidly, but fewer than half of US hospitals had at least a basic system in 2012. *Health affairs (Project Hope)*, 32(8): 1478–85. <http://content.healthaffairs.org/content/early/2013/06/27/hlthaff.2013.0308.abstract> 29 October 2014.
- Duncan of Jordanstone College of Art & Design. 2013. Design Research Masters Module: Affinity Mapping. <https://designresearchdundee.wordpress.com/tag/affinity-diagram/> 27 February 2015.
- E-O Ataguba, J. & Akazili, J. 2010. Health care financing in South Africa: moving towards universal coverage. *Continuing Medical Education*, 28(2): 74. <http://www.cmej.org.za/index.php/cmej/article/view/1782>.
- EMarketer. 2012. Trends for 2013: 'Big Data' Gets Bigger - eMarketer. <http://www.emarketer.com/Article/Trends-2013-Big-Data-Gets-Bigger/1009509> 28 August 2014.
- European Commission. 2016. Technological accidents triggered by natural disasters. <https://ec.europa.eu/jrc/en/research-topic/technological-accidents-triggered-natural-disasters> 19 September 2016.
- European Commission. 2014. *The Use of Big Data in Public Health Policy and Research Background information document*. http://ec.europa.eu/health/ehealth/docs/ev_20141118_co07b_en.pdf.
- Evans, B.J. 2011. Much Ado About Data Ownership. *Harvard Journal of Law & Technology*, 25(1): 70–130.
- Fielding, M., Odero, B. & Ochieng, C. 2016. *From paper to data : taking medical health records into the future*. <https://www.sei-international.org/mediamanager/documents/Publications/SEI-2016-ConfPaper-AfyaHealthRecords.pdf>.
- Frank, L., Forsythe, L., Ellis, L., Schrandt, S., Sheridan, S., Gerson, J., Konopka, K. & Daugherty, S. 2015. Conceptual and practical foundations of patient engagement in research at the patient-centered outcomes research institute. *Quality of Life Research*, 24(5): 1033–1041.
- Fridsma, D. & Lygeia, R. 2012. Blue Button Enhances Access to Health Data in Personal Health Record. <http://www.healthit.gov/buzz-blog/consumer/blue-button-initiative-enhancing-consumer-access-health-information/> 22 March 2015.
- Gans, D., Kralewski, J., Hammons, T. & Dowd, B. 2014. Medical groups' adoption of electronic health records and information systems. *Health Affairs*, 24(5): 1323–1333.
- Genitsaridi, I., Kondylakis, H., Koumakis, L., Marias, K. & Tsiknakis, M. 2013. Towards intelligent personal health record systems: Review, Criteria and extensions. *Procedia Computer Science*, 21: 327–334. <http://dx.doi.org/10.1016/j.procs.2013.09.043>.
- Grandia, L. 2014. *Three things healthcare CIOs should know about datawarehouses*.
- Gray, A. & Vawda, Y. 2015. *South African Health Review 2014/15*. <http://www.hst.org.za/publications/south-african-health-review-2014/15>.
- Hamburger, T. & Tsukayama, H. 2014. White House releases big data and privacy report. *The Washington Post*. <http://www.washingtonpost.com/blogs/the-switch/wp/2014/05/01/white-house-releases-big-data-and-privacy-report/> 2 September 2014.
- Harris, B., Goudge, J., Ataguba, J.E., McIntyre, D., Nxumalo, N., Jikwana, S. & Chersich, M. 2011. Inequities in access to health care in South Africa. *Journal of Public Health Policy*, 32(S1): S102–S123. <http://dx.doi.org/10.1057/jphp.2011.35>.
- He, H., Du, Z., Zhang, W. & Chen, A. 2015. Optimization strategy of Hadoop small file storage for big data in healthcare. *The Journal of Supercomputing*.

- <http://link.springer.com/10.1007/s11227-015-1462-4>.
- Health Information Systems Program-SA. 2015. South Africa Projects. <http://www.hisp.org/projects/south-africa-projects/> 25 July 2016.
- Herasevich, V., W, P.B., Dong, Y., Peters, S.G. & Gajic, O. 2010. Informatics Infrastructure for Syndrome Surveillance, Decision Support, Reporting, and Modeling of Critical Illness. In *Mayo Clinic Proceedings*. 247–254.
- Hoffman, S. & Podgurski, A. 2013a. Big bad data: law, public health, and biomedical databases. *The Journal of law, medicine & ethics : a journal of the American Society of Law, Medicine & Ethics*, 41 Suppl 1: 56–60. <http://www.ncbi.nlm.nih.gov/pubmed/23590742>.
- Hoffman, S. & Podgurski, A. 2011. Meaningful Use and Certification of Health Information Technology: What about Safety? *Journal of Law, Medicine and Ethics*, 39: 77–80.
- Hoffman, S. & Podgurski, A. 2013b. The Use and Misuse of Biomedical Data : Is Bigger Really Better ? *American journal of law and medicine*, 39(4): 497–538.
- Hripcsak, G., Bloomrosen, M., FlatelyBrennan, P. & Chute, C. 2014. Health data use, stewardship, and governance: ongoing gaps and challenges: a report from AMIA's 2012 Health Policy Meeting. *Journal of the ...*, 21: 204–211. <http://jamia.bmj.com/content/21/2/204.short> 5 August 2014.
- Huijboom, N. & Broek, T. Van Den. 2011. Open data : an international comparison of strategies. *European Journal of ePractice*, 12(March/ April 2011): 1–13. www.epracticejournal.eu.
- Iqbal, U., Ho, C.-H., Li, Y.-C.J., Nguyen, P.-A., Jian, W.-S. & Wen, H.-C. 2013. The relationship between usage intention and adoption of electronic health records at primary care clinics. *Computer methods and programs in biomedicine*, 112(3): 731–7. <http://www.ncbi.nlm.nih.gov/pubmed/24091088>.
- Janssen, M., Charalabidis, Y. & Zuiderwijk, A. 2012. Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, 29(4): 258–268. <http://www.tandfonline.com/doi/full/10.1080/10580530.2012.716740>.
- Jewett, D. 2014. 7 Tips to Succeed with Big Data in 2014.
- Kankanhalli, A., Hahn, J., Tan, S. & Gao, G. 2016. Big data and analytics in healthcare: Introduction to the special section. *Information Systems Frontiers*: 1–3.
- Kearney, A. 2014. *Rethinking Personal Data : A New Lens for Strengthening Trust*. <http://www.weforum.org/reports/rethinking-personal-data-new-lens-strengthening-trust>.
- Keen, J., Calinescu, R., Paige, R. & Rooksby, J. 2013. Big data + politics = open data: The case of health care data in England. *Policy & Internet*, 5(2): 228–243. <http://doi.wiley.com/10.1002/1944-2866.POI330>.
- Keen, J., Calinescu, R., Paige, R. & Rooksby, J. 2012. Big Health Data: Institutional and Technological Challenges. *Internet, Politics and Policy*: 1–19.
- Kim, M.I. & Johnson, K.B. 2002. Personal Health Records: Evaluation of Functionality and Utility. *Journal of the American Medical Informatics Association*, 9(2): 171–180.
- Klein, H.K. & Myers, M.D. 1999. A SET OF PRINCIPLES FOR CONDUCTING AND EVALUATING INTERPRETIVE FIELD STUDIES IN INFORMATION SYSTEMS. *MIS Quarterly*, 23(1): 67–94.
- Korpela, M. 2011. *ISD4D : A Holistic Information Systems Development Approach for Societal Development*.
- KPMG Africa Limited. 2012. *The State of Healthcare in Africa*. <http://www.kpmg.com/Africa/en/IssuesAndInsights/Articles-Publications/Documents/The-State-of-Healthcare-in-Africa.pdf>.
- Kruger, K. 2016. Big data: Where can data and analytics take healthcare in Africa? <http://www.sablog.kpmg.co.za/2016/05/big-data-can-data-analytics-take-healthcare-africa/> 25 July 2016.
- Laney, D. 2001. 3D Data Management: Controlling Data Volume, Velocity, and Variety. *Application Delivery Strategies*, 949(February 2001): 4.
- Laurie, G. & Sethi, N. 2011. *Information Governance of Use of Health-Related Data in Medical Research in Scotland : Current Practices and Future Scenarios*.

- Lazer, D., Kennedy, R., King, G. & Vespignani, A. 2014. Big data. The parable of Google Flu: traps in big data analysis. *Science (New York, N.Y.)*, 343(6176): 1203–5. <http://www.ncbi.nlm.nih.gov/pubmed/24626916> 10 July 2014.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. & Byers, A.H. 2011. Big data : The next frontier for innovation , competition , and productivity. , (June).
- Marchibroda, J. 2013. *A Policy Forum on the Use of Big Data in Health Care*.
- Markle Foundation. 2006. Connecting Americans to Their Health Care: A Common Framework for Networked Personal Health Information. <http://www.markle.org/sites/default/files/CF-Consumers-Full.pdf> (accessed 12 February 2015).
- Mata-lima, H., Alvino-borba, Andreilcy-O, Adilson-Pinhero, Mata-Lima, A. & Almeida, J.A. 2013. Impacts of Natural Disasters on Environmental and Socio- -Economic Systems: What Makes the Difference? *Ambiente & Sociedade*, 16: 45–64.
- Mediclinic. 2016. Our Hospitals. <http://www.mediclinic.co.za/hospitals> 21 September 2016.
- Menius Jr, J.A. & Rousculp, M.D. 2014. Growth in Health Care Data Causing an Evolution in the Pharmaceutical Industry. *NC Medical Journal*, 75(3): 188–190.
- Mind Tools. 2015. Affinity Diagram. http://www.mindtools.com/pages/article/newTMC_86.htm.
- Mosley, M. 2008. *DAMA-DMBOK Functional Framework*.
- Mxoli, A., Mostert-Phipps, N. & Gerber, M. 2014. Personal Health Records : Design considerations for the South African context.
- Netcare. 2016a. Divisions. <https://www.netcare.co.za/Who-We-Are/Group-at-a-glance/Divisions> 21 September 2016.
- Netcare. 2016b. You're in safe hands. <https://www.netcare.co.za/News-Hub/Articles/articleid/47> 21 September 2016.
- Ohm, P. 2010. Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization. *UCLA Law Review*, 57: 1701–1777.
- Ohno-Machado, L. 2014. NIH's Big Data to Knowledge initiative and the advancement of biomedical informatics. *Journal of the American Medical Informatics Association : JAMIA*, 21(2): 193. <http://www.ncbi.nlm.nih.gov/pubmed/24509598> 18 July 2014.
- Page. 2014. Google: 100,000 lives a year lost through fear of data-mining | Technology | [theguardian.com](http://www.theguardian.com/technology/2014/jun/26/google-healthcare-data-mining-larry-page). <http://www.theguardian.com/technology/2014/jun/26/google-healthcare-data-mining-larry-page> 27 June 2014.
- Pagliari, C., Shand, T. & Fisher, B. 2012. Embedding online patient record access in UK primary care : a survey of stakeholder experiences. : 1–7.
- Parliament of the Republic of South Africa. 2013. South Africa Protection Personal information Act, 2013. *National Gazettes, No 37067*, (10505): 1–148. http://www.gov.za/documents/download.php?f=204368%5Cnhttp://www.greengazette.co.za/notices/act-no-4-of-2013-protection-personal-information-act-2013_20131126-GGN-37067-00912.
- Pasquale, F. 2013. Grand Bargains for Big Data: The Emerging Law of Health Information. *Maryland Law Review*, 72(3): 682–772.
- Patton, M.Q. 2003a. Qualitative evaluation checklist. *Evaluation Checklists Project*, (September): 1–13. [http://www.wmich.edu/evalctr/archive_checklists/qec.pdf%5Cnhttp://dmeforpeace.org/sites/default/files/Patton_Qualitative Evaluation Checklist.pdf](http://www.wmich.edu/evalctr/archive_checklists/qec.pdf%5Cnhttp://dmeforpeace.org/sites/default/files/Patton_Qualitative%20Evaluation%20Checklist.pdf).
- Patton, M.Q. 2003b. *Qualitative evaluation checklist*. [http://www.wmich.edu/evalctr/archive_checklists/qec.pdf%5Cnhttp://dmeforpeace.org/sites/default/files/Patton_Qualitative Evaluation Checklist.pdf](http://www.wmich.edu/evalctr/archive_checklists/qec.pdf%5Cnhttp://dmeforpeace.org/sites/default/files/Patton_Qualitative%20Evaluation%20Checklist.pdf).
- Pear, R. 2007. Warnings Over Privacy of U.S. Health Network. *New York Times*. http://www.nytimes.com/2007/02/18/washington/18health.html?_r=0 5 November 2014.
- Pearlstein, S. 2011. Sunday Pearlstein: End of the 'anything goes' era of antitrust? *The Washington Post*. http://www.washingtonpost.com/blogs/wonkblog/post/sunday-pearlstein-end-of-the-anything-goes-era-of-antitrust/2011/12/11/gIQAEUcJmO_blog.html 26 February 2015.

- Peters, S.G. & Buntrock, J.D. 2014. Big data and the electronic health record. *The Journal of ambulatory care management*, 37(3): 206–10.
<http://www.ncbi.nlm.nih.gov/pubmed/24887521> 14 July 2014.
- Phelan, J.C., Link, B.G. & Tehranifar, P. 2010. Social Conditions as Fundamental Causes of Health Inequalities: Theory, Evidence, and Policy Implications. *Journal of Health and Social Behavior*, 51(S28).
- Pozzebon, M. 2003. Criteria for Conducting and Evaluating Critical Interpretive Research in the IS Field. *Cahier du GReSI no, (3): 1–27*.
<http://www2.hec.ca/gresi/documents/cahier0314.pdf>.
- Privacy Rights Clearinghouse. 2016. Identity Theft & Data Breaches.
<https://www.privacyrights.org/Identity-Theft-Data-Breaches> 19 September 2016.
- Raymond Lang. 2011. FUTURE CHALLENGES TO THE PROVISION OF HEALTH CARE IN THE 21 st CENTURY. In *Consortium of Institutes of Higher Education in Health and Rehabilitation in Europe Annual Conference*. Lisbon.
- van Riimenam, M. 2013. Why The 3V's Are Not Sufficient To Describe Big Data.
<https://datafloq.com/read/3vs-sufficient-describe-big-data/166> 24 May 2015.
- Saunders, B.M. & Tosey, P. 2012. The Layers of the Research Design. *Rapport*, (30): 58–59.
- Saunders, M., Lewis, P. & Thornhill, A. 2009. *Research methods for business students*.
- Schroeder, R. & Schroeder, R. 2016. Big data business models : Challenges and opportunities. : 1–15.
- Schultze, U. 2000. A Confessional Account of an Ethnography about Knowledge Work. *MIS Quarterly*, 24(1): 3–41.
- Segall, N., Saville, J.G., L'Engle, P., Carlson, B., Wright, M.C., Schulman, K. & Tchong, J.E. 2011. Usability evaluation of a personal health record. *AMIA ... Annual Symposium proceedings / AMIA Symposium*. *AMIA Symposium*, 2011: 1233–42.
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3243224&tool=pmcentrez&rendertype=abstract>.
- Shaikh, A.R., Prabhu Das, I., Vinson, C. a. & Spring, B. 2011. Cyberinfrastructure for consumer health. *American Journal of Preventive Medicine*, 40(5): S91–S96.
<http://dx.doi.org/10.1016/j.amepre.2011.02.012>.
- Sinsky, C.A., Beasley, J.W., Simmons, G.E. & Baron, R.J. 2014. Electronic Health Records : Design , Implementation , and Policy for Higher-Value Primary Care. *Annals of Internal Medicine*.
- South African Government. 2000. Promotion of access to information Act. *Government Gazette*, 416(20852): 1–45.
- South African Government. 2004. The National Health Act No 61 of 2003. *Government Gazette*, 469(26595): 1–94.
http://www.nsw.gov.au/sites/default/files/Government_Gazette_2_December.pdf#page=15.
- Statistics South Africa. 2011. *Use of health facilities and levels of selected health conditions in South Africa : Findings from the General Household Survey , 2011*.
- Studeny, J. & Coustasse, A. 2014. Personal Health Records : Is Rapid Adoption Hindering Interoperability ? *Perspectives in Health Information Management*.
- Tassi, R. 2009. Affinity diagram | Service Design Tools.
<http://www.servicedesigntools.org/tools/23> 27 February 2015.
- TechAmerica. 2014. Big Data. <http://www.techamerica.org/public-policy-advocacy/all-industry-priorities/big-data/> 1 September 2014.
- Tennessee Department of Health. 2016. HIPAA: Health Insurance Portability Accountability Act - TN.Gov. <http://www.tn.gov/health/topic/hipaa> 20 January 2016.
- Terhune, C. 2012. Patient data outage exposes risks of electronic medical records. *Orlando Sentinel*. http://articles.orlandosentinel.com/2012-09-10/news/hc-ls-patient-data-20120908_1_electronic-health-records-medical-records-patient-safety 26 November 2014.
- Terry, N. 2013. Protecting Patient Privacy in the Age of Big Data. *UMKC Law Review*, 81(2).

- Terry, N.P. 2008. Personal Health Records : Directing More Costs and Risks to Consumers ? *Drexel Law Review*, 1: 216–260.
- Thacker, S.B., Qualters, J.R. & Lee, L.M. 2012. Public Health Surveillance in the United States: Evolution and Challenges. *Morbidity and Mortality Weekly Report*, 61.
- Thilakanathan, D., Chen, S., Nepal, S., Calvo, R. & Alem, L. 2014. A platform for secure monitoring and sharing of generic health data in the Cloud. *Future Generation Computer Systems*, 35: 102–113.
<http://linkinghub.elsevier.com/retrieve/pii/S0167739X13001908> 17 July 2014.
- Torjman, S. 2005. What is Policy? *Caledon Insitute of Social Policy*, (September): 1–20.
- Tu, J. V, Willison, D.J., Silver, F.L., Fang, J., Richards, J. a, Laupacis, A. & Kapral, M.K. 2004. Impracticability of informed consent in the Registry of the Canadian Stroke Network. *The New England journal of medicine*, 350(14): 1414–1421.
- Uller, S.H.F., Rinkley, J.A.B. & Ornoch, P.E.T.A. 2007. Issues in Biomedical Research Data Management and Analysis : Needs and Barriers. *Journal of the American Medical Informatics Association*, 14(4): 478–488.
- University of California. 2014. *Personal Data for the Public Good: New Opportunities to Enrich Understanding of Individual and Population Health*.
- Vayena, E., Salathe, M., Madoff, L.C. & Brownstein, J.S. 2015. Ethical Challenges of Big Data in Public Health. *PLoS Computational Biology*, 11(2): 1–7.
- Vital, M. Issues in health information.
- Walsham, G. 2006. Doing interpretive research. *European Journal of Information Systems*, 15(3): 320–330.
- Wang, L., Ranjan, R., Kołodziej, J., Zomaya, A. & Alem, L. 2015. Software tools and techniques for big data computing in healthcare clouds. *Future Generation Computer Systems*, 43–44(November): 38–39.
- Western Cape Government Health. 2014. Healthcare 2030: The Road to Wellness.
<https://www.westerncape.gov.za/assets/departments/health/healthcare2030.pdf> 7 April 2016.
- Willie, M. & Matsebula, T. 2007. Private Hospitals. *South African Health Review*. 159–174.
- World Economic Forum. 2012. *Big Data , Big Impact : New Possibilities for International Development*.
- World Economic Forum. 2014. *Rethinking Personal Data : Trust and Context in User-Centred Data Ecosystems*.
- World Economic Forum. 2013. *Unlocking the Value of Personal Data : From Collection to Usage*.
- World Health Organisation. 2008. *Framework and standards for country health information systems*. <http://apps.who.int/iris/handle/10665/43872> 26 June 2014.
- World Health Organisation. 2012. *Health Systems in Africa*.
- World Health Organisation. 2015. WHO | The right to health.
<http://www.who.int/mediacentre/factsheets/fs323/en/> 28 July 2015.
- Wynia, M.K., Torres, G.W. & Lemieux, J. 2011. Many physicians are willing to use patients' electronic personal health records, but doctors differ by location, gender, and practice. *Health Affairs*, 30: 266–273.
- Yao, W., Chu, C.H. & Li, Z. 2010. The use of RFID in healthcare: Benefits and barriers. *Proceedings of 2010 IEEE International Conference on RFID-Technology and Applications, RFID-TA 2010*, (June): 128–134.
- Yin, R.K. 2013. Applications of case study research. *Applied Social Research Methods Series*, 34: 173.
- Yiu, C. 2012a. A Right to Data Fulfilling the promise of open public data in the UK. *Policy Exchange*, (March): 1–28.
- Yiu, C. 2012b. *The Big Data Opportunity*. London.
- Zuiderwijk, A. & Janssen, M. 2014a. Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly*, 31(1): 17–29.
<http://dx.doi.org/10.1016/j.giq.2013.04.003>.
- Zuiderwijk, A. & Janssen, M. 2014b. Open data policies , their implementation and impact : A

framework for comparison. *Government Information Quarterly*, 31(1): 17–29.
<http://dx.doi.org/10.1016/j.giq.2013.04.003>.

APPENDICES

A1 - Interview Themes

Theme A: Consensual Access VS Non-consensual Access to Patients Health Data

Theme B: Suppression of Data

Theme C: IP protection in health

Theme D: Regulations on information flow and use

Theme E: Anonymisation and Re-identification

Theme F: Personal Health Records (PHR) Use

A2 - Interview Questions

Interview questions (Approximate interview time per theme: 20min)

Section A Theme: Regulations on information flow and use

1. Does your organisation have any technical standards for Data Management and Integration to facilitate information flow and use?

Probe: When requests for information are made to your organisation, what processes are followed?

Probe: How is data being received from other entities prepared for integration into the mainstream data of your organisation?

2. Is your organisation engaged in recruiting and retaining personnel who are well versed on facilitating optimised information flow and use?

Probe: How is your organisation enforcing that it has personnel focused on data research (i.e. mining, analytics and sharing)?

Probe: What roles of personnel currently in your organisation are related to data research?

Probe: Has any training of staff been facilitated by your organisation, aimed at improving their data research skills?

3. How would you describe the policies and regulations that govern information flow between your organisation and other organisations?

Probe: Can you recall an incident when you had difficulties accessing or sharing information with an organisation above yours?

Probe: Can you recall an incident when you had trouble accessing or sharing information with an organisation below yours?

4. What could be said about your organisation's approach to the concept of big data?

Probe: Is there any explicit mention of the words "big data" in any of your organisations data management policies and or strategies?

Probe: Can your organisations data management repositories be described as an architecturally single data common?

Probe: What is your organisation's spending (budget) on data management and utilisation projects?

Section B Theme: Consensual Access VS Non-consensual Access to Patients Health Data

5. How do current data policies in your organisation address consensual access to patient health data?

Probe: Do you think the current policy measures on consensual access to patient data are adequate?

Probe: What can be done to address the issue of consensual access to patients private data?

6. What experiences do you have around consensual access or non-consensual access of patient information?

Probe: Would your job tasks be easier with or without any of these two?

Probe: Can you recall an incident when your work activities were hampered because consensual access to patient data had to be sought?

7. In your opinion, how do you feel about the current position of your organisation's policies on access to patient data secondary uses?

Probe: Which is better, consensual access or non-consensual access?

Probe: What could be the dangers of non-consensual access of patient data over consensual access?

Section C Theme: Suppression of Data

8. What is your understanding of the concept of an open data policy?

Probe: Do you know of any health organisation which is engaged in data suppression through the use of policies?

Probe: Can suppression of data have any benefits over an open access data policy?

9. To what extent do data policies in your organisation promote information access or suppress information access?

Probe: Why have you given that answer?

Probe: Can you recall an instance where data suppression hindered the progress of your work activities?

Probe: Can you recall an instance where your ability to access certain information was granted to the benefit of your tasks or another person's interests?

10. Do you know of any other organisations or entities that your organisation is actively sharing information with?

Probe: Was this sharing initiated by some data policy?

Probe: What benefits have these collaborations yielded?

11. In the context of healthcare patient information, which is better, an open data policy vs. suppression of data?

Probe: What benefits could there be of having an open data policy vs. data suppression?

Probe: What benefits could there be of data suppression over having an open data policy?

12. Are there any policies that seek to discourage data suppression?

Probe: Are there any policies that promote meaningful use of patient data?

Section D Theme: IP protection in health

13. Can any of the policies that govern your current organisation's data access and use, be described as protecting the intellectual property (IP) of either your organisation or a member of your organisation?

Probe: Is there any policy that does recognise the intellectual property over any healthcare related product?

14. What impact can IP protection policies in healthcare have on data utilisation?

Probe: Suppose a colleague discovers something that could improve the way healthcare is administered at the same time having a potential to make him rich.

Would you support his IP claims, which would effectively hinder using his new discoveries without his consent and or paying for them?

Section E Theme: Anonymisation and Re-identification

15. Does your organisation engage in any form of information anonymisation of patient data before it is used for secondary purposes?

Probe: Which patient data elements e.g. name, gender, should not be publicly revealed to unauthorised people?

Probe: What are the processes involved in hiding these data elements?

16. How important is data anonymisation to your organisation?

Probe: Can you recall an incident where use of patient identifiable data became a problem afterwards?

17. What criteria does your organisation use to determine what data elements of patient data should be hidden when using data for secondary uses?

Probe: What are the standards that determine if an information element is sensitive or not sensitive to be part of data for secondary use?

18. How does your organisation ensure that anonymisation of patient data does not render it totally useless for secondary uses such as analytics?

19. Under the current methods being used for anonymisation, how easy could it be to re-identify the owners that data?

Section F Theme: Personal Health Records (PHR) Use

20. What plans have been put in place to capacitate the use of PHR as part of the research data pools of your organisation?

Probe: Have any measures been put in place that promote increased user/patient participation in patient health data creation?

Probe: Is there any patient-generated data that is currently being used by your organisation to improve the quality of their healthcare services?

Probe: Does the organisation have policies that are directly concerned with promoting PHRs use as part of their big data initiatives?

21. What issues have been faced in integrating PHR into the mainstream data of the organisation?

Probe: How different are the standards (i.e. quality, meta-data, content) of PHRs you are exposed to, to those of data used by your organisation?

Probe: Can you recall an incident when some information on PHRs could not be used, and what the reason was?

22. What is the current position of your organisation as to who the owner of PHR data is?

Probe: Can a patient put measures to restrict your organisation from accessing their patient records?

Probe: Is it a requirement to seek the permission of a patient before their data can be used for secondary purposes?

23. Are there any other external systems that you are currently sharing PHRs with?

Probe: How interoperable are current PHR systems?

Probe: Are there adequate policies that govern PHR systems development to ensure that there is increased operability between PHR systems and other data management systems?

A3 – Interviewees List

Interviewees	Organization	Role
WCDoH-IV	WC Government Department of Health: Information Directorate	Director
WCDoH-CN	WC Government Department of Health: Information Directorate	Assistant Director: Emergency medical services, Provincial hospitals, National hospitals, Forensic pathological services
WCDoH-SR	WC Government Department of Health: Information Directorate	Assistant Director: HR Engineering, Laundry services
WCDoH-NF	WC Government Department of Health: Information Directorate	Assistant Director: Business Intelligence
WCDoH-FB	WC Government Department of Health: Information Directorate	Business Units
WCDoH-LS	WC Government Department of Health: Information Directorate	Assistant Director: Knowledge management
WCDoH-SB	WC Government Department of Health: Information Directorate	Assistant Director: Retail Services
NHO-KV	Netcare Limited	Head: SAP healthcare solutions

NHO-TS	Netcare Limited	Manager: Business Services GRC
NHO-DN	Netcare Limited	Governance Manager
NHO-LLR	Netcare Limited	POPI representative
NBI-VE	Netcare Limited	Manager: Business Intelligence
NSBR-DM	Netcare Limited	SAP BA reporting
CSIR-MH	CSIR	

A4 – Interview Transcripts are Available and Stored as Part of Research Data on the Institutional Research Repository.

A5 – Co-Design Sessions Plan

Co-Design Sessions Plan (1 hour 30 mins)

Warm-up (10 mins)

1. Introduce myself and explain my research topic and interests.
2. Icebreaker: Ask participants to introduce themselves and describe themselves as a manager.
3. Explain the ideologies underpinning the need for having the sessions.

First session (Problem Identification) Materials: paper notes, markers (20 min)

1. Group participants into groups of 3.
2. Participants to respond to the following questions in their groups: answers to be written on the paper notes.
 - What are your current challenges about your health?
 - Why fears do you have about patient information?
 - What would you advise your local healthcare authority about managing healthcare information?
 - What challenges might you face if you were the guardian of your own health information?

The answers should be one-word answers, where impossible, not more than three words.

3. Switch the groups so that each group has the set of notes not written by them. They should then formulate as many sentences as possible from the words that they have (which have been generated by another group).

4. Record these sentences as they present the issues that they groups have identified.
Actually, these are the worries that they personally have.

Second session (Brainstorming and problem solving) Materials: big paper sheets, sticky notes, prestick, markers. (20 mins)

1. Develop as many solutions to the issues identified in the previous session.
They are to be pasted on the sheets which will be set up on the wall.
2. Ask the participants to group related solutions into themes.
3. Jot down extra views and thoughts on their personal information management.

Third session (Personal Health Record Design) Materials: big paper sheets, prestick, markers. (20 mins)

1. Articulate design principles that ensure all relevant data elements especially the important ones are included in captured data.
2. Ask the participants in their groups to design a system of records that ensure their involvement in their health information management.
This solution should address the issues identified in the first and second session.