CAUSAL MECHANISMS THAT ENABLE INSTITUTIONALISATION OF OPEN GOVERNMENT DATA IN KENYA

A Doctoral Thesis

By

Mungai, Paul Wando

MNGPAU002

Presented for the

Degree of Doctor of Philosophy

In the Department of Information Systems

University of Cape Town

Ph.D Research Supervisor

Prof. Jean-Paul van Belle

Department of Information Systems

University of Cape Town

March 2017

DECLARATION

DOCTORAL DEGREES BOARD
UNIVERSITY OF CAPE TOWN
Cape Town, South Africa

PhD Thesis Title: Causal mechanisms that enable institutionalisation of open government data in Kenya

I, Paul Mungai hereby;

(a) grant the University of Cape Town free license to reproduce the above thesis in whole or in part, for the purpose of research;

(b) declare that:
the above thesis is my own unaided work, both in concept and execution, and that apart from the normal guidance from my supervisor, I have received no assistance. Neither the substance nor any part of the above thesis has been submitted in the past, or is being, or is to be submitted for a degree at this University or at any other university.

I am now presenting the thesis for examination for the degree of PhD.

Signature: Signed by candidate

Date: 11th March, 2017

ACKNOWLEDGEMENTS

This Ph.D was a long but exciting journey, thanks to the many great souls who conspired in helping me achieve this dream. This includes the University of Cape Town community, Strathmore University @iLabAfrica team (Alma mater), the Africa Open Data Collaboratives (AODC) team, and my family to whom I hold so dear. I will follow this order in giving my acknowledgments. First, I want to thank my Ph.D supervisor Prof. Jean-Paul van Belle for his unrelenting guidance and support. We did lots of interesting things together including going for hikes, travelling together (the travel bug is catching up with me), chatting over glasses of Irish coffee, and consulting on open data initiatives. Second, I want to thank the Information Systems (IS) Ph.D community, including the IS professors who have continuously provided input to my study both formally and informally. Third, I want to thank the critical realism group at the IS department. This includes Dr. Chidi Ononiwu, Laban Bangui, Pitso Tsibolane, and Yasser Buchana. Honestly, I had little understanding of Critical Realism (CR) at the onset, but through the debates and chats we held occasionally, my understanding of this philosophy has gradually improved.

I am also grateful to Brendon Wolff-Piggott and Dr. Mphatso Nyemba-Mudenda for shedding some light and sharing resources on critical realism. Special shoutout to the UCT GoodFellas group, namely Gordon Nana Kwesi, Gwamaka Mwalemba, Laban Bagui, Mohammed Alhassan, Musa Chemisto, Samwel Mwapwele, and Yasser Buchana. On numerous occasions, we broke bread together (foodies) as we watched soccer matches, national elections, and exchanged ideas on numerous social (media) uprisings. Thanks for helping me refine my ideologies and inner convictions.

Fourth, I want to thank Dr. Joseph Sevilla, who is the founder and director of @iLabAfrica. I feel indebted for your assistance in helping to establish connections with major actors in the Kenya open data space. This includes people from the World Bank, IBM, Kenya National Bureau of Statistics, ICT Authority, and Code for Kenya. I am also grateful for providing me with a workstation and internet @iLabAfrica during all my visits to Kenya as I carried out field work. Special thanks to Lorna Mutegi and Dennis Gichuki both from @iLabAfrica for providing additional support, and research input drawn from your experience in the open data space.

Fifth, I want to thank the Africa Open Data Collaboratives (AODC) that consists of Rose Katherine Townsend, Winnie Kamau, Dr. Jeanne Holm, Dr. Verena Luise Knippel, and Dr. Umaru Bah. Thanks for your continuous feedback on my research, assistance in obtaining travel grants to open data events, and creating research opportunities related to my study.

I am also grateful to my soulmate, family and friends for your unmeasurable love and fervent prayers. Those calls, emails, voice notes, delicious meals, and hearty conversations kept me going. This is a collective accomplishment. I would not have made it without You!

ABSTRACT

Open Government Data (OGD) has become a topic of prominence during the last decade. However, most governments have not realized the desired outcomes from OGD, which implies that the envisaged value streams have not been realized. This study defines three objectives that will help address this shortcoming. First, it seeks to identify the causal mechanisms that lead to effective institutionalization and sustainability of OGD initiatives in a developing country context. Second, it seeks to identify the social, economic, cultural, political structures and components that describe the OGD context. Third, it seeks to identify the underlying context-mechanism-outcome (CMO) configurations in the Kenya Open Data Initiative (KODI). The guiding philosophy for this qualitative study is critical realism, which is implemented using Pawson & Tilley's realist evaluation model. Data is obtained through observation of open data events, semi-structured interviews and documentary materials from websites and policy documents. Fereday & Muir-Cochrane's five-stage thematic analysis model is applied in conducting data analysis. Three main contributions arise from this study. The first contribution is the open data institutionalization analysis guide. This study collates several institutionalization concepts from literature with the aim of developing a lens for analyzing OGD initiatives. The second contribution is the identification of supporting mechanisms, including a description of the current CMO configurations. The resulting case study provides an in-depth account of KODI between 2011 and 2016. It describes how the supporting mechanisms manifest, their role and interconnectedness. They include; law and policy, demand, awareness, buy-in and ownership, advocacy, and planning, coordination and capacity building. This will assist policy makers in understanding the current setup, identifying gaps, and establishing or supporting existing support structures and mechanisms. The third contribution is related to scarcity of empirical work based on critical realism in the field of information systems. This research will act as a reference point for future IS research, in determining how critical realism can be applied to conduct similar studies.

PUBLICATIONS

Peer Reviewed Conference Proceedings

Two conference papers were published during the course of this PhD research.

1. Mungai P., Van Belle J.P. & Sevilla J. (2016). Mechanisms impacting the Kenya Open Data Initiative. Intl Conf on Information Resources Management (Conf-IRM), Cape Town.
2. Mungai, P. & Van Belle, J.P. (2015). Candidate mechanisms that enable institutionalization of Open Data Initiatives. *Proceedings of the 9th IDIA conference, IDIA2015, Nungwi, Zanzibar. ISBN: 978-0-620-68395-1.*

Research Contracts

Two research contracts related to the Kenya open data initiative were awarded in 2016.

1. Mungai, P. & Van Belle, J.P "*Understanding the structures and mechanisms that foster stewardship of open development in Kenya* based on Reilly and Alperin's framework *Public Engagement in Open Development: A Knowledge Stewardship Approach*". This contract was awarded by Strengthening Information Society Research Capacity Alliance (SIRCA) and funded by International Development Research Centre (IDRC).
2. Mungai, P. "*Kenya Data Report providing an assessment of the data for development ecosystem in Kenya"*. This contract was awarded by IDRC/UNECA/UNDP/Web Foundation partnership.

The first contract seeks to make an empirical contribution to the theoretical framework on public engagement in open development. This will be published as a chapter in the Open Development Book by SIRCA in 2017. The second contract contributes to the 2016 Africa Data Revolution Report, which is yet to be published by UNDP.

TABLE OF CONTENTS

1. INTRODUCTION 9

1.1 BACKGROUND OF THE RESEARCH 9

1.2 PROBLEM STATEMENT 10

1.3 RESEARCH QUESTIONS 11

1.4 CONTRIBUTION OF THE STUDY 12

1.4.1 Theoretical contribution 12

1.4.2 Practical contribution 12

1.5 RESEARCH METHODS OVERVIEW 12

1.6 THESIS STRUCTURE 14

1.7 CONCLUSION 15

2. OGD DEFINITION, CHALLENGES AND BENEFITS 16

2.1 DEFINITIONS 16

2.1.1 Open data 16

2.1.2 Open government data (OGD) 17

2.2 OGD CHALLENGES 17

2.3 OGD BENEFITS 19

2.4 CONCLUSION 20

3. CRITICAL REALISM ASSUMPTIONS, MODELS AND CASE STUDY METHOD ADOPTION 22

3.1 POSITIVISM 22

3.2 INTERPRETIVISM 23

3.3 CRITICAL RESEARCH 24

3.4 CRITICAL REALISM 25

3.4.1 The concept of causality 26

3.4.2 Ontological assumptions of critical realism 28

3.4.3 Scientific Modes of Inference and Reasoning 32

3.5 MODELS FOR AN EXPLANATORY SOCIAL SCIENCE 35

3.5.1 Popper-Hempel explanatory model 35

3.5.2 DREI 36

3.5.3 RRRE 36

3.5.4 Margaret Archer's morphogenetic approach 36

3.5.5 Pawson & Tilley's realist evaluation model 40

3.5.6 Danemark's Explanatory model 43

3.5.7 Preferred critical realism model 45

3.6 CASE STUDY DESIGN FROM A CRITICAL REALISM PERSPECTIVE 46

3.7 CASE STUDY METHODOLOGICAL PRINCIPLES 47

3.7.1 Explication of Events 47

3.7.2 Explication of Structure and Context 48

3.7.3 Retroduction 48

3.7.4 Empirical Corroboration 48

3.7.5 Triangulation/Multimethods 48

3.8 CONCLUSION 49

4. INSTITUTION, INSTITUTIONALIZATION AND INSTITUTIONAL THEORY 50
4.1 INSTITUTION 50
4.2 INSTITUTIONALIZATION 51
4.2.1 Habitualization 53
4.2.2 Objectification 53
4.2.3 Sedimentation 54
4.2.4 Institutionalization theory propositions 55
4.3 INSTITUTIONAL THEORY 56
4.3.1 Institutional assumptions 57
4.3.2 Institutional isomorphism 57
4.3.3 Theories of institutional theory 58
4.3.4 Deinstitutionalization 59
4.3.5 Criticism of institutional theory 60
4.4 COMPATIBILITY OF CRITICAL REALISM WITH INSTITUTIONAL THEORY 61
4.5 OPEN DATA INSTITUTIONALIZATION CASE ANALYSIS GUIDE 61
4.6 CONCLUSION 63

5. RESEARCH METHODOLOGY 64
5.1 RESEARCH DESIGN 64
5.2 CASE STUDY DESIGN COMPONENTS 65
5.2.1 Study questions 65
5.2.2 Study propositions 65
5.2.3 Unit of analysis 66
5.2.4 Logic linking the data to the propositions 66
5.2.5 Criteria for interpreting findings 67
5.3 CONCLUSION 68

6. PHASE 1 - HYPOTHESIS GENERATION 69
6.1 CASE STUDY INSTITUTIONALIZATION ANALYSIS 69
6.2 UNDERSTANDING INFORMATION SYSTEMS CMOS FROM LITERATURE 81
6.2.1 Generative Mechanisms of OGD 81
6.2.2 Strategic Information Systems Planning Using Critical Realism 82
6.2.3 Generative Mechanisms for Innovation in Information Infrastructures 84
6.2.4 Mechanisms Supporting Quantity, Accuracy and Traceability of Information 84
6.3 OGD ECONOMICS 86
6.4 OGD CRITIQUE 86
6.5 PROPOSED OGD CMOS FROM LITERATURE 88
6.5.1 Context 88
6.5.2 Enabling Mechanisms 89
6.5.3 Outcomes 90
6.6 PROPOSED CMOS AFFECTING KODI INSTITUTIONALIZATION 91
6.7 CONCLUSION 93

7. PHASE 2 - DATA COLLECTION 94
7.1 SAMPLING PROCEDURES 94
7.2 DATA COLLECTION METHODS AND TOOLS 95

7.2.1 Semi-structured interviews *95*
7.2.2 Document review *97*
7.3 CONCLUSION 98
8. PHASE 3 - DATA ANALYSIS 99
8.1 DEVELOPING THE CODE MANUAL 99
8.2: TESTING THE RELIABILITY OF CODES 101
8.3: SUMMARIZING DATA AND IDENTIFYING PRELIMINARY THEMES 102
8.4: APPLYING TEMPLATE OF CODES AND ADDITIONAL CODING 104
8.5: CONNECTING THE CODES AND IDENTIFYING THEMES 112
8.5.1 Definition of open data *112*
8.5.2 Benefits of open data *113*
8.5.3 Stakeholders or partners *113*
8.5.4 Roles of ICTA data fellows *114*
8.5.5 Open data curation tools *115*
8.5.6 Challenges faced by KODI data fellows *115*
8.5.7 Law and policy *116*
8.5.8 OGD sourcing, curation and publication workflow *117*
8.5.9 Measures for ensuring timeliness in data release *118*
8.5.10 Challenges in obtaining open government data *119*
8.5.11 Sensitive information and correctness *119*
8.5.12 Capacity, training and technical support *121*
8.5.13 KODI financial support *122*
8.5.14 Culture *123*
8.5.15 KODI Innovation *124*
8.5.16 Demand *125*
8.5.17 Open data champion and buy-in from stakeholders *127*
8.5.18 Awareness and stakeholder engagement *128*
8.5.19 Conceptualization and implementation of KODI *130*
8.6 CORROBORATING AND LEGITIMATING CODE THEMES 132
8.6.1 Demand *132*
8.6.2 Law and policy *133*
8.6.3 Planning, coordination and capacity building *133*
8.6.4 Awareness, buy-in and ownership *134*
8.6.5 Advocacy *135*
8.7 CONCLUSION 136
9. PHASE 4 - THEORY TESTING 139
9.1 CMOS AFFECTING KODI INSTITUTIONALIZATION 140
9.2 HYPOTHESIS TESTING 142
9.2.1 Demand mechanism *142*
9.2.2 Law and policy mechanism *143*
9.2.3 Planning, coordination and capacity building mechanism *143*
9.2.4 Awareness, buy-in and ownership mechanism *143*
9.2.5 Advocacy mechanism *144*

9.3 EXPLANATION OF KODI CMOS BASED ON INSTITUTIONALIZATION THEORY 144
9.3.1 Demand mechanism *145*
9.3.2 Law and policy mechanism *145*
9.3.3 Planning, coordination and capacity building mechanism *146*
9.3.4 Awareness, buy-in and ownership mechanism *147*
9.3.5 Advocacy mechanism *148*
9.4. KODI MECHANISMS INTERCONNECTEDNESS 149
9.5. CONCLUSION 150
10. KODI CASE STUDY 151
10.1 INTRODUCTION 151
10.2 KODI STRATEGIES AND SUPPORTING STRUCTURES 152
10.2.1 Re-launch of the KODI Portal *152*
10.2.2 Data fellows programme *153*
10.2.4 Request a dataset *154*
10.2.5 Data release form *155*
10.2.6 Awareness and capacity building workshops *156*
10.2.7 Internal capacity building *156*
10.2.8 Funding *157*
10.2.9 Official vs crowdsourced data *157*
10.2.10 Expanded Partnerships *158*
10.3 KODI CHALLENGES 158
10.3.1 Bureaucracy *158*
10.3.2 Complexity of data *160*
10.3.3 Lack of structures to support KODI *160*
10.3.4 Sensitive data and state interference *161*
10.3.5 Lack of awareness *161*
10.3.6 Low innovation *162*
10.4 CONCLUSION 163
11. CONCLUSION AND RECOMMENDATION 166
11.1 SUMMARY OF FINDINGS 166
11.2 QUALITY CONCERNS 167
11.3 CONTRIBUTION 169
11.3.1 Theoretical contribution *169*
11.3.2 Practical contribution *170*
11.4 RECOMMENDATIONS FOR FUTURE RESEARCH 171
12. REFERENCES 173
13. APPENDIX 1: SEMI-STRUCTURED INTERVIEW GUIDES 181
A. DATA FELLOWS INTERVIEW QUESTIONS 181
B. GOVERNMENT AGENCY STAFF INTERVIEW QUESTIONS (MANAGEMENT) 182
C. GOVERNMENT AGENCY STAFF INTERVIEW QUESTIONS (NON-MANAGEMENT) 183
D. CSO AND PRIVATE SECTOR INTERVIEW QUESTIONS (USERS) 184
E. ETHICAL CLEARANCE 185
F. CONSENT FORM 186

G. SAMPLE INTERVIEW TRANSCRIPT 188

1. Introduction

This introduction chapter provides the background to the research, and includes a historical background of the Kenya Open Data Initiative (KODI). This is followed by a description of the problem statement, research questions, current knowledge gaps and the intended impact of this research. This chapter also provides a detailed description of the contribution that this study seeks to make, and the research methods it applies with the aim of realizing its objective. It concludes with a brief outline of the various chapters in this dissertation.

1.1 Background of the research

For years, several governments have provided open government data (OGD) with the aim of increasing transparency, improving accountability, satisfying legal obligations, improving public-government participation and collaboration, harnessing government responsiveness and democratic control, increasing public awareness of government programmes and activities, and promoting innovation, efficiency and effectiveness in government services (Böhm et al., 2012; Hoxha & Brahaj, 2011; Robinson, Harlan, Zeller, & Felten, 2009; Shadbolt, Hara, Berners-lee, Gibbins, & Glaser, 2007; Ubaldi, 2013).

OGD comprises of the following types of datasets: business information; registers, patent and trademark information; public tender records; geographic information; legal information; meteorological information; statistical data on economics, employment, health, population, and public administration; and transport information (Ubaldi, 2013). Unfortunately, government sometimes releases processed data in the name of OGD, which then limits the types of insights that a user can derive from OGD. This reduces demand as users do not find much value from processed datasets. Data is a service by itself, and does not need to be processed to qualify (M. Janssen, Charalabidis, & Zuiderwijk, 2012). This partly explains why most of the governments that have attempted to offer this service have not succeeded in creating new value streams (Shadbolt et al., 2007).

Most of the governments that have institutionalized OGD – though not necessarily successfully are found in Europe, North America and Australia (K. Janssen, 2011; Shadbolt et al., 2007; Ubaldi, 2013; Yu & Robinson, 2012). There are a few initiatives in Africa, more particularly

Kenya and Morocco, though these only began in 2011 and have had varying challenges. This could be explained from a stakeholder perspective, since Morocco's initiative was purely government driven, while Kenya had a great push from the World Bank, which not only provided substantial financial support, but also technical advisory, which helped in shaping how the initiative was implemented.

This suggests the need to factor in the contextual differences as this could have an impact on the processes and outcomes (Heeks, 2002). For instance, from a cultural and political perspective, the value system may vary and thus, priorities and expectations between one country and another are likely to differ. These variations could be addressed by identifying the underlying causal mechanisms that affect OGD institutionalization. Once these mechanisms are understood, it is possible to improve the process, or introduce them in a similar context.

1.2 Problem statement

In 2011, Former President Mwai Kibaki launched the Kenya Open Data Initiative (KODI), in a public event attended by over two thousand participants. The initiative aimed at providing access to government information, which it views as a national asset that should be shared with the aim of increasing transparency, unlocking social and economic value, and building Government 2.0 (Provost, 2011).

However, KODI has encountered numerous challenges including data hugging – where custodians of Open Government Data (OGD) fail to release data, lack of demand and interest from the public, lack of political will from government, lack of adequate capacity in KODI, and low awareness about open data among government officials and the general public (Kags, 2014).

These challenges are inter-related. For instance, no datasets were uploaded to the KODI portal between the year 2012 and 2013. This was partly because there were no technical staff at KODI to upload datasets, and that government management failed to fill these positions soon after they were vacated. This demonstrates lack of adequate capacity, and lack of political will from government, which is linked to government bureaucracy in decision making. These problems arise due to lack of proper structures, which would need to be institutionalized if the initiative was to become sustainable. The following section identifies several research questions, which are

aimed at assisting KODI to become more effective in achieving its goals and objectives sustainably.

1.3 Research Questions

This study aimed at identifying the causal mechanisms that lead to effective institutionalization and sustainability of OGD initiatives in a developing country context. For this study, institutionalization refers the process of establishing OGD within government by creating structures that guarantee its sustainability. Using critical realism, it would be possible to explain how the expected outcomes or goals can be achieved given an appropriate context.

This study also sought to identify the social, economic, cultural, and political structures and components that describe the OGD context. For instance, it identified structures that are introduced and the processes or events that emerge from OGD institutionalization. This study also helped in identifying the possible context-mechanism-outcome configurations from KODI. This is important since it motivate government agencies to establish the required structures and activate the necessary causal mechanisms.

The main research question is stated below.

- What are the causal mechanisms that affect institutionalization and sustainability of KODI?

This question requires an understanding of the structures that affect the institutionalization process, which could either be positive or negative. It also requires an understanding of the emerging outcomes, which includes their associated context and mechanisms. This understanding helps in determining the various context-mechanism-outcome configurations that affect the institutionalization process. This results in an in-depth causal explanation of the mechanisms that affect proper institutionalization of KODI. The following sub-questions will help in arriving at this understanding.

- What are the structures that affect the institutionalization of KODI?
- What are the context-mechanism-outcome configurations that emerge from KODI?

1.4 Contribution of the study

This section describes the theoretical and practical contributions of this study to the OGD phenomenon.

1.4.1 Theoretical contribution

Four theoretical contributions were identified, including identification of mechanisms that enable OGD institutionalization, a more detailed description of OGD institutionalization from a developing country context, application of critical realism in the information systems field, and a scientific account of the Kenya open data initiative.

1.4.2 Practical contribution

The beneficiaries of this study include government, non-governmental organisations and the general public. Government should implement the findings with the aim of creating a more enabling environment through policy, and conducting periodic evaluations with the aim of identifying gaps and opportunities in OGD services. Non-governmental organisations and the general public should also use this information in OGD use and advocacy.

1.5 Research methods overview

This study applies the case study methodology to develop an in-depth description of the KODI phenomenon. Data was collected using qualitative methods, including semi-structured interviews, observations and document review (Dube & Pare, 2003). This approach is effective for social science, which is problem and not methodology driven, allowing the researcher to apply the method that is most suited in answering the research question at hand (Flyvbjerg, 2006).

This study aims at identifying the context and causal mechanisms that enable proper institutionalization and sustainability of KODI. Further to this, it seeks to identify the outcomes that emerge from the possible configurations of the contexts and mechanisms. To achieve this, critical realism was preferred as it helps in understanding both the objective reality and the subjective interpretations of the people involved in the process. Chapter three compares critical realism with the positivist and interpretivist approach and explains their limitations in realizing this study's objective.

Using critical realism as the underlying paradigm, this study applies Pawson & Tilley's realist evaluation model. This model comprises of four stages, namely, hypothesis, data collection, data analysis, and theory testing (Pawson & Tilley, 2004). A description of these stages is provided below.

The hypothesis stage starts with the formation of candidate theories, followed by a set of hypotheses based on the candidate theories (Pawson & Tilley, 2004). Candidate theories are developed through induction by conducting systematic search and review of literature to identify the characteristics and outcomes of OGD implementation (Pawson & Tilley, 2004; Ranmuthugala et al., 2011). The search and review process entails explication/logical analysis of events, structure and context, from which experiences are abstracted to obtain detailed aspects of the events, and to identify the components of structure and variations of contextual influences, and potential mechanisms (Bhaskar, 2008; Smith, 2006; Wynn & Williams, 2012). The candidate theories act as input to the deductive process of formulating hypotheses. This process involves rephrasing the Context, Mechanisms & Outcomes (CMO) configurations using a suitable theory (Ranmuthugala et al., 2011). A theory on institutionalization was identified to assist in rephrasing CMO configurations.

Data collection stage uses the findings of the hypothesis stage to create the data collection instruments. Once the data collection instruments have been formulated, data sources are identified, followed by data collection, with the aim of testing the preliminary theories identified in the hypotheses stage (Pawson & Tilley, 2004).

Data analysis is the third stage, which is a retroductive process of making observations using the data collected in stage two to test the hypotheses formed in stage one (Ranmuthugala et al., 2011). Retroduction involves proposing and testing the existence of several hypothetical causal mechanisms together with their underlying elements and contextual elements (Easton, 2010; Wynn & Williams, 2012). This stage helps in identifying the CMO configurations that occur with regularity (Ranmuthugala et al., 2011).

Theory testing is the last stage, which is also a retroductive process of empirical corroboration that results in programme specifications (Pawson & Tilley, 2004; Ranmuthugala et al., 2011). Empirical corroboration refers to the process of reviewing, validating and refining the proposed theories and potential CMO configurations using empirical observations from the data analysis stage (Easton, 2010; Popper, 2014; Ranmuthugala et al., 2011; Wynn & Williams, 2012).

Stage	**Activity**	**Logical Reasoning**	**Principles, Theory & Methods**
Stage 1: Hypothesis	Develop candidate theories	Induction	- Explication of events - Explication of structure and experiences - Institutionalization theory
Stage 1: Hypothesis	Generate hypotheses	Deduction	- Institutionalization theory
Stage 3: Data analysis	Make observations	Retroduction/Abduction	- Retroduction - Triangulation/multimethods - Institutionalization theory
Stage 4: Theory testing	Formulate programme specifications	Retroduction/Abduction	- Empirical corroboration - Institutionalization theory
Research Paradigm: Critical Realism			

Table 1.1 Research framework

The table above gives a summary of the various stages in this study, the underlying logical reasoning and principles, all of which will be discussed in detail from chapter five onwards.

1.6 Thesis Structure

This thesis consists of eleven chapters. The next four chapters consists of literature review, a discussion on critical realism, a description of institutionalization theory, and the research methodology. The chapter on critical realism describes its underlying philosophical assumptions and compares this to positivism and interpretivism, provides a justification on its suitability for this study, describes the four models of critical realism including a justification to use realist evaluation model, and a description of how case study approach is applied in a critical realism study.

The chapter on institutionalization provides a detailed description of what an institution comprises of and what institutionalization entails. This is followed by a discussion of institutional theory and its compatibility to critical realism. The findings in these discussions help in developing a framework that guides in analyzing open data institutionalization.

The next four chapters consist of each of the four stages of Pawson & Tilley's evaluation methodology. Based on the outcome of the realist evaluation, a case study is provided, which provides an in-depth discussion of KODI institutionalization. This thesis ends with a conclusion chapter, which includes recommendation for further research. Each chapter begins with an introduction of what it entails, including its underlying assumptions, and projected outcomes.

1.7 Conclusion

This chapter provides an introduction of the open government data concept, and how it is being institutionalized in Kenya. This helps in identifying the gaps in the Kenyan context, leading to a justification of the need to conduct a study that identifies the mechanisms that are affecting (positively or negatively) the proper institutionalization of KODI. This discussion is followed by a description of the theoretical and practical contributions that are envisaged in this study. It concludes with a discussion of the preferred methodology, which includes the justification for this choice and a description of the various stages and methods. The next chapter provides a literature review on open data.

2. OGD Definition, Challenges and Benefits

This chapter provides the definitions, challenges and benefits of OGD. Ideally, it would have included a detailed literature review of the OGD phenomenon except that this study follows Pawson & Tilley's realist evaluation model, which requires literature review to be included in the first stage on hypothesis. The literature acts as input for developing the initial propositions and candidate theories (Pawson & Tilley, 2004). As a result, a detailed literature review is provided in chapter 6, where the hypothesis stage is discussed.

2.1 Definitions

Definitions of Open data and Open Government Data (OGD) are provided below.

2.1.1 Open data

Data is described as open if it contains the following characteristics; no licensing restrictions, zero cost of access, ease of access, reuse capability and no administrative overheads (Hoxha & Brahaj, 2011; Shadbolt et al., 2007; Ubaldi, 2013). These characteristics are all related to access of data. In this context, data is described as "the lowest level of abstraction from which information and then knowledge are derived" (Ubaldi, 2013). This implies that access is provided on unprocessed data - which gains value when converted into services of public value (M. Janssen et al., 2012).

Open Data could then be defined as "any sets of data which can be reused with no restrictions by any form of licensing or patents, data that are well structured and can be easily accessed and reused" by various actors (Hoxha & Brahaj, 2011), including state and non-state actors. Government is the primary provider of data and is expected to be accountable and objective in providing this service in the interest of the entire society. Other data sources include the private sector, civil society organisations, and non-governmental organisations. They process open data and present the information and knowledge derived using formats that are accessible by the target community (Ubaldi, 2013).

The aim of open data is to foster transparency, improve accountability, satisfy legal obligations, improve public-government participation and collaboration, foster responsiveness and democratic control, increase public awareness of government programmes and activities, deliver

citizen self-designed public services, and foster innovation, efficiency and effectiveness in government services (Böhm et al., 2012; Hoxha & Brahaj, 2011; Parycek, Höchtl, & Ginner, 2014; Robinson et al., 2009; Shadbolt et al., 2007; Tinati, Carr, Halford, & Pope, 2012; Ubaldi, 2013). For this to be achieved, government must inevitably revise restrictions on data access when institutionalizing OGD (M. Janssen et al., 2012; Ubaldi, 2013).

2.1.2 Open government data (OGD)

This is data that is made accessible to the public by government without use-restrictions. OGD has no value in itself, except when it is offered as a service to the public (M. Janssen et al., 2012). Most governments that have attempted to offer this service have not succeeded in presenting it in ways that result in new value streams (Shadbolt et al., 2007). In addition, reliability of open data has been reduced following delays in publishing open data. This implies that it is still a concept in its infancy stage and governments are still using trial and error methods to effectively institutionalize OGD.

OGD results from a democratic government and interestingly, helps in furthering democracy. It is in a democratic government that citizens are able to request for broader freedom of information (Meijer, 2012). If granted, citizens make use of this information to influence public policy and generate new forms of public value (Dawes, 2005; Shadbolt et al., 2007). However, governments often experience uncertainty since it is difficult to determine how citizens will interpret and use the data (Meijer, 2012).

Values that are crucial to proper OGD institutionalization comprises of openness of operations, public scrutiny, clear policies, democracy, privacy, and efficiency, supportive legal framework (Huijboom & Broek, 2011; Meijer, 2012; Ubaldi, 2013). Governments and their agencies uphold these values differently, which affects how OGD gets institutionalized and consequently, the OGD outcome.

2.2 OGD Challenges

Besides the said values, OGD faces several challenges. These include: disclosure policies which limits OGD provisioning; copyright which creates contention on who owns government data; poor data quality and management practices which increases the cost of converting the data in

machine readable format; enormous and discrete nature of government data that requires extra effort and cost when transforming it to OGD; finding a dedicated government agency that solicits datasets from other government agencies; increasing public interest and awareness of OGD that includes public servants, citizens and the private sector (Ubaldi, 2013).

In some countries, the very government agents who are meant to assist in OGD provisioning become the stumbling blocks through resistance to change. In Cameroon for instance, government officials refused to use the e-government system whose main aim was to improve transparency and accountability (Heeks, 2005). In other cases, government agents "consider public information their own property and not of the citizen" leading to resistance in releasing information/documents (Meijer, 2012). An example of this was observed in Vienna, the capital city of Austria. The city is decentralized and comprises of several heads of departments who report to the city directorate on administrative matters. The city directorate issued a regulation on OGD, which requires all departments to release open data. However, like many other regulations issued by the city directorate, this directive was not adhered to, and departments retained control on what data to release and in what granularity (Parycek et al., 2014). This implies that internal interests affect OGD provisioning.

Another challenge is lack of governance structures that specifically address OGD, which implies that e-Government structures are not ideal for OGD. This is based on the vivid and fast changing nature of OGD, which may not be supported by the existing structures, such as the approval processes. Austria could be emulated in this regard, where a sleek governance structure was formulated to address this. Following approval, it was cascaded downwards to the provinces, cities and municipalities (Parycek et al., 2014).

Another challenge to OGD provisioning are the factors that complicate access and integration. This follows the complex nature of open data given its size, schematic heterogeneity, quality variations and lack of consistency (Böhm et al., 2012; Hoxha & Brahaj, 2011). This results from the fact that government has multiple agencies, which follow different standards of data presentation and also, the fact that these agencies produce different types of data, which call for

different presentation styles creating a challenge for uniformity. There is also the lack of meta-data, which would assist in describing the data (Hoxha & Brahaj, 2011).

Related to the complexity of open data, another challenge arises when government attempts to structure and publish processed data (Robinson et al., 2009). This often arises when the complexity that comes with the heterogeneous nature of open data is ignored (Janssen et al., 2012). It is difficult to develop sites that address the needs of all citizens. This is partly because there is lack of insight on their perspectives and needs (Janssen et al., 2012). To address this, government should focus on developing infrastructure that is capable of presenting the underlying data in open, structured and machine-readable format. This should not imply that private entities will understand, interpret and present this data correctly the first time, but it is believed that they are more capable of exploring more approaches faster and find solutions faster than the government would. This is partly because private entities deal with specific data sets while government would have to do this for all the data sets. They should also use open standards such as RSS (Really Simple Syndication) to notify users whenever new data is made available (Robinson et al., 2009).

There is need to determine whether the data in question is fit for use. Data cannot be universally fit for all users since requirements vary. This implies that users need to have access to good data descriptors that guide them in deciding on the appropriateness of the data in question. This calls for the provisioning of metadata – data that describes data. Noting the importance of metadata, we can deduce that "good quality metadata is as important as the quality of the data itself"(Dawes & Helbig, 2010; Ubaldi, 2013).

2.3 OGD Benefits

Despite these challenges, OGD presents numerous benefits including fostering transparency, improving accountability, satisfying legal obligations, improving public-government participation and collaboration, increasing responsiveness and democratic control, improving public awareness of government programmes and activities, and fostering innovation, efficiency and effectiveness in government services (Böhm et al., 2012; Hoxha & Brahaj, 2011; Robinson et al., 2009; Shadbolt et al., 2007; Ubaldi, 2013).

Citizens can help actualize these benefits by providing insights to government using the availed OGD. They can also use the OGD to provide new services that are aimed at value addition. Overall, one of the main objectives/outcome is to improve decision making of both government and the citizens, which can be achieved when the right data is availed (Dawes, 2005; Shadbolt et al., 2007; Ubaldi, 2013).

Another factor affecting the success of the OGD provisioning is the top-down political culture where the state operates a monopoly on data by structuring the data with little (if any) consultation from the citizens and homogenizing its interactions with the citizens. The problem with this approach is that they often do not succeed in presenting OGD to citizens effectively, leading to a loss in potential value streams (Shadbolt et al., 2007). A strong argument on this has been made advising governments to concentrate less on web sites and more on technical infrastructures that present OGD in structured formats. Citizens and organisations (both for and not for profit) can then analyse, format and present OGD in multiple ways – which then attracts more viewers leaving a greater percentage of the population enlightened and thus increasing transparency (Robinson et al., 2009).

It is also important to address the digital divide. This ensures that citizens are able to understand and use technology. This can be achieved if the issues related to accessibility and technology literacy are addressed. This is through adaptive technologies that cater for all including those with disabilities and functionality. This must be able to address most, if not all user needs (Bertot, Jaeger, Grimes, & Hansen, 2010).

There is also need to create an enabling environment, one that protects the privacy of the government agents, and creates open access using mechanisms that are accessible to majority of the Citizens within the country in question (Ubaldi, 2013).

2.4 Conclusion

Law and policy are the enablers of openness, though as was demonstrated in the case of Cameroon and city of Vienna, goodwill from government is key in realizing openness. This study will help in understanding the context-mechanism-outcome configurations that enable effective institutionalization of open data initiatives. This will start with an in-depth literature

review in chapter six, where the first phase of the realist evaluation model is described, resulting in theory based on the literature. The next chapter provides an in-depth discussion of critical realism assumptions, models including realist evaluation, and how case study is applied in a study based on critical realism philosophical approach.

3. Critical realism assumptions, models and case study method adoption

This chapter describes various research paradigms including positivism, interpretivism, critical research, and critical realism. It also compares these paradigms with critical realism, and provides a justification on the choice of using critical realism for this study over the rest. In describing the rationale used in selecting the critical realism approach, various options of operationalizing critical realism philosophy are reviewed, which include Margaret Archer's morphogenic approach, Pawson and Tilley's realistic evaluation method, and Danemark's explanatory model. The challenges of the preferred operationalization approach are also stated, including the approach taken to resolve their shortfalls. This chapter ends with a discussion of case study design within a critical realist framework.

These research paradigms are discussed below, including how critical realism creates a hybrid from some of their approaches. We also discuss critical research, a philosophy that may be confused with critical realism, since both share the word "critical" in their definition.

3.1 Positivism

This involves the humean notion of causality, which adopts the statistical approach to observe constant conjunction of events to find meaning (John Mingers, 2004; Smith, 2006; Wynn & Williams, 2012). It falls short of attempting to investigate causality of the underlying mechanisms that leads to the observed conjunctions (Bhaskar, 2008; Easton, 2010; John Mingers, 2004). The other challenge to this approach is the concept of closure, where open systems are assumed to have undifferentiated experience. The assumption is that the excluded factors or external mechanisms do not have significant effect, following the belief that they are impossible to measure, unknown and random. However, in practice, there is possibility at times that these excluded factors or external mechanisms have a significant effect outside the sample data (Bhaskar, 2008; John Mingers, 2004).

Besides these shortfalls, especially the lack of grip on the individual case, critical realism does not entirely abandon this statistical approach (Bhaskar, 2008; John Mingers, 2004). It finds it useful in identifying and explaining patterns within the data. Some of the more applicable

techniques it applies include factor analysis and path analysis. Once explanations are formulated, it makes use of corroboration and falsification to validate them (John Mingers, 2004).

In some cases, critical realism extends the use of the statistical approach by conducting quasi experiments within a closed environment, where influence can be induced on normally uncontrolled causal agents (Bhaskar, 2008; John Mingers, 2004). This follows the principles of substance (things) and causality, which are interdependent and complementary, where things remain normal (persistence) until they are acted upon by a causal agent. In the case of closure, critical realism applies the same concept in a controlled environment, which guarantees that the excluded factors or external mechanisms do not affect the results (Bhaskar, 2008).

3.2 Interpretivism

Interpretivism is based on the soft systems methodology, which involves understanding the subjective meanings that actors assign to a given phenomenon within a particular context (John Mingers, 2004; Wynn & Williams, 2012). It focuses more on epistemology while paying no regard to ontology, therefore rejecting the causal powers of the natural and social worlds. This sounds antithetical to critical realism since it regards both ontology (reality) and epistemology (social interaction) by upholding the existence and impact of the subject under study (John Mingers, 2004; Smith, 2006).

Interpretivism views the outcome of epistemology - ideas, concepts, meanings and categories to be less real than objects, and is strongly relativist, accepting all emerging viewpoints as equally valid (John Mingers, 2004). Critical realism regards them as social products with social properties, that emerge from - but cannot be reduced to the real world because they have causal effect on both the real (ontological) and social (epistemological) structures that constitute them (Mingers, 2004; Smith, 2006).

On relativism, critical realism distinguishes between epistemic and judgmental relativism. On judgmental relativism, when the purpose is clearly defined, people are able to publicly discuss their claims about reality as they think, and evaluate the emerging arguments rationally - despite having varying worldviews and beliefs, to arrive at judgments about what reality is objectively

like (Easton, 2010; John Mingers, 2004). On epistemic relativism (external realism), critical realism is of the notion that once ideas are expressed, they seize to be wholly subjective and separate from our own interpretations/constructions of them, becoming intransitive and open for discussion, investigation and judgment by others. This view is not shared by interpretivism since it is strongly relativist and views all ideas as being equally valid without investigating their impact on the physical world, limiting itself to the social and ideational world (John Mingers, 2004; Smith, 2006).

3.3 Critical Research

In information systems, critical research focuses on social issues affecting information technology (Myers & Klein, 2011). It attempts to critically evaluate the phenomenon under investigation with the aim of finding ways of improving the encompassing social and ethical responsibilities (Myers & Klein, 2011; Orlikowski & Baroudi, 1991). Research is deemed critical if it seeks to uncover the taken-for-granted assumptions about an organization and the supporting information systems, with the aim of critiquing the status quo by unveiling the underlying structural conditions within social systems (Orlikowski & Baroudi, 1991). This philosophy is therefore similar to critical realism but different from the positivist and intepretivist research which, according to Orlikowski and Boroudi, "are content to predict or explain the status quo" (Myers & Klein, 2011; Orlikowski & Baroudi, 1991).

Several notions on the relationship between critical research and critical realism exist. They include; critical realism is the same as critical research, critical realism embodies critical research, and critical realism and critical research are commensurable. Each of these notions are explained and clarified below, which also helps explain why critical realism was preferred for this study. Critical research does not subscribe to any specific methods, analysis procedures or even data gathering techniques. Also, it seldom applies all the ideas and logic of critical realism. Further to this, critical research cannot be referred to or treated as a method. Finally, being critical while conducting research does not amount to critical research (Eriksson & Goldkuhl, 2013).

3.4 Critical Realism

Critical realism is based on the notion that events should be investigated at the level of generative mechanism that occur in the real domain, not at the level of constant conjunction of regular events since establishing a constant conjunctive relationship is not sufficient (Easton, 2010; John Mingers, 2002; Smith, 2006). These mechanisms could be likened to the connections between variables, from which outcomes emerge (Fox, 2009). Events are selected for investigation based on their ability to have causal effect on the world. This differs from empiricism that selects events based on perceptability - the notion that only that which can be perceived can exist (Easton, 2010; John Mingers, 2002).

Events can be investigated at either the empirical, actual or real domain. The difference in these domains is as follows: Empirical domain contains events that are observed or experienced; Actual domain contains events that do or do not occur, including those in the empirical domain; Real domain contains the whole of reality that includes mechanisms, events and experiences (John Mingers, 2002). Events in the real and actual domain may not be observable at all or even when they are, observers may understand them quite differently (Easton, 2010; John Mingers, 2002). This is likely to imply that events can only be observed in the empirical domain. However, this is not the case, but simply that events may not always be capable of being observed in the real or actual domain, thus creating a need for experimentation. Also, the conditions established by the observer during experimentation do not cause the results, which are dependent on causal laws at play (Easton, 2010; John Mingers, 2002). Following this understanding of the various domains, critical realism suggests using the empirical domain during investigation (Easton, 2010; John Mingers, 2002).

The empirical domain could be likened to the tip of an iceberg, where only a part is visible, and it is that which we observe. However, this should not imply that what is invisible is non-existent or unconnected to the visible (Easton, 2010). This analogy leads to a fundamental epistemological assumption in critical realism, that no observation is infallible (Easton, 2010; John Mingers, 2004). This follows the realization that, under the empirical domain, it is unlikely to make observations that will result in full understanding of the social situation in question. In addition, there is no definitive criteria to judge the “truth” of a particular explanation. Therefore,

there is need for the observer to collect sufficient data that will aid in distinguishing alternative explanations of the same or a similar social situation (Easton, 2010; Smith, 2006).

These explanations are created and presented causally through the language and procedures we use ordinarily (Easton, 2010; Ilkka, 1991). They result in knowledge whose "truth" value is determined by scientists through consensus. This causality in explanation resulted in the principle of causality, which requires the use of the "same causal idiom" as other sciences (Ilkka, 1991).

Unlike the positivist and interpretivist, critical realism acknowledges the role of subjective knowledge of social actors, including the existence of independent structures/mechanisms within a social setting that act as either constrains or enablers (John Mingers, 2004; Wynn & Williams, 2012). This perspective helps in providing a more detailed causal explanation of the phenomena in question since it includes both the actors interpretations and the structures and mechanisms at play (John Mingers, 2002; Wynn & Williams, 2012). Once these interpretations are expressed, they seize to be wholly subjective, and become intransitive making them available for investigation, debate or judgment by others (Mingers, 2002).

Critical realism has three main benefits to information systems research. First, it helps in transcending a number of inconsistencies between stated philosophical assumptions and the actual practice of information systems research, under both positivism and interpretivism (Wynn & Williams, 2012). Second, it offers a way to address the rigor-relevance gap in management research following its approach to causal analysis through multi-method/triangulation and multilevel approaches (Wynn & Williams, 2012). This implies that critical realism is not limited to the case study approach and also that it can support several methods and approaches within a single study. Third, it assists in identifying connections between technology implementations and their outcomes (Wynn & Williams, 2012).

3.4.1 The concept of causality

In formulating knowledge claims about reality, scientific research is guided by three main factors; ontology, epistemology, and methodology (Bhaskar, 2008; Easton, 2010; John Mingers, 2002; Wynn & Williams, 2012). Ontology refers to the nature of reality (objects, properties,

facts, world), which, according to critical realism, is independent of human interaction (Easton, 2010; Ilkka, 1991; Wynn & Williams, 2012). Epistemology refers to the evidence-based assessment and justification of knowledge claims, which comprise of beliefs, perception and rationality (Ilkka, 1991; Wynn & Williams, 2012). Methodology refers to the process or procedures by which knowledge is created (Wynn & Williams, 2012).

In recent years, critical realism has gained interest from researchers in information systems (Easton, 2010; John Mingers, 2004; Walsham, 2006; Wynn & Williams, 2012). This follows its paradigmatic assumptions, which vary with those of positivism and interpretivism, thus creating new perspectives to research (Bhaskar, 2008; Easton, 2010; John Mingers, 2002; Smith, 2006; Wynn & Williams, 2012). The main variance is on the concept of causality where critical realism goes ahead to investigate the underlying mechanisms (Bhaskar, 2008; Easton, 2010; John Mingers, 2004).

Causality refers to the explanation about the relationship that exists between an action or thing (cause) and the outcome (effect) it generates, which is based on enduring entities that are either social, physical or conceptual (John Mingers, 2002; Wynn & Williams, 2012). These explanations are created and presented causally through the language and procedures we use ordinarily (Easton, 2010; Ilkka, 1991). It results in knowledge whose "truth" value is determined by scientists through consensus (Ilkka, 1991). Often, our knowledge of the factors and relationships that cause an event, determine our ability to explain that event/phenomenon (Wynn & Williams, 2012).

Causality has been adopted in two distinct ways. First, several competing explanations about a phenomenon are hypothesized and tested through a relationship between the conceptual entities, with the aim of finding the best explanation (Easton, 2010; Wynn & Williams, 2012). Second, explanations about a phenomenon are based on actors understanding and interpretation of their roles in a social setting, and how their subjective meanings are developed and sustained within that setting (Easton, 2010; Wynn & Williams, 2012). This explanation includes both the actor and the researcher's interpretation of meanings and intentionality, including the reciprocal influence of social action and context (Easton, 2010; Wynn & Williams, 2012).

Critical realism aims at providing accurate and empirically supported statements about causation, which elaborate how and why a certain phenomenon occurred, regardless of perceptability since having causal effect implies existence (John Mingers, 2002; Wynn & Williams, 2012). It does not base causality on regular occurrences, successions or correlation of events (John Mingers, 2002; Smith, 2006; Wynn & Williams, 2012). Instead, it bases causality on the underlying mechanisms that lead to the observed conjunctions or events within a given setting (John Mingers, 2002; Smith, 2006; Wynn & Williams, 2012).

3.4.2 Ontological assumptions of critical realism

Ontology is concerned with the nature of objects and whether their reality exists objectively or subjectively relative to humans (John Mingers, 2002; Wynn & Williams, 2012). From an ontological perspective, critical realism is founded on the following four assumptions: reality is intranstitive, exists independently of the conditions produced by humans, allowing them to access them (Bhaskar, 2008; John Mingers, 2002; Wynn & Williams, 2012); reality is stratified and involves stratification of structures, mechanisms, events, and experiences (Bhaskar, 2008; John Mingers, 2002; Wynn & Williams, 2012); power is a potentiality that may or may not be exercised, emerges from interactions between components, and is not reducible to the powers that formed it (Bhaskar, 2008; Wynn & Williams, 2012); and reality is an open system, where natural laws operate, and no constant conjunction of events prevail (Bhaskar, 2008; Wynn & Williams, 2012). These assumptions are further described in the following sub-sections.

3.4.2.1 Intransitive Reality

Critical realism accepts that the world and entities therein exist independent of human knowledge or subjective beliefs, and cannot be reduced to human perceptions or experiences (Bhaskar, 2008; John Mingers, 2002; Wynn & Williams, 2012). It also acknowledges that reality comprises of complex interactions and humans only experience a portion of it, following their limitation in apprehending, characterizing or even measuring it (John Mingers, 2002; Wynn & Williams, 2012).

Critical realism therefore views reality in two dimensions, namely, transitive and intransitive dimensions (John Mingers, 2002; Wynn & Williams, 2012). Transitive dimension (epistemology) involves the production of knowledge and beliefs of entities/objects - about their

causality, and is the work of humans (John Mingers, 2002; Wynn & Williams, 2012). This knowledge and belief is generated from reason and scientific research, and is subject to revision and reinterpretation (Wynn & Williams, 2012). Intransitive dimension (ontology) involves entities that form the world, and which operate independently of humans perceptions of them (John Mingers, 2002; Wynn & Williams, 2012). Bhaskar further stratifies reality into three domains, namely real, actual and empirical (Smith, 2006; Wynn & Williams, 2012).

3.4.2.2 Stratified Reality

Critical realism stratifies reality into three domains; real (what exists), actual (observable and non-observable events) and empirical domain (observable events) (Smith, 2006; Wynn & Williams, 2012). The real domain comprises of entities and structures of reality and the causal powers that control them, all of which exist independently (Smith, 2006; Wynn & Williams, 2012). The actual domain is a subset of the real domain, and involves events that occur when the causal powers are engaged on the related entities and structures of reality, which may or may not be observable by humans (Smith, 2006; Wynn & Williams, 2012). The empirical domain is a subset of the actual domain, which is limited to events that are observable, perceivable, and or measurable by humans (Smith, 2006; Wynn & Williams, 2012).

Positivism and interpretivism do not subscribe to the notion of stratified reality (John Mingers, 2004; Wynn & Williams, 2012). Positivism assumes a flat reality by adopting the humean inductive approach which observes reality from the point of constant conjunction of cause and effect, with little regard to the enjoining mechanisms (John Mingers, 2004; Wynn & Williams, 2012). Interpretivism subscribes to the soft systems methodology and holds the position that reality is socially constructed and can only be observed and understood by analyzing the actors' meaning and actions (John Mingers, 2004; Wynn & Williams, 2012). Critical realism contrasts to these positions by asserting that general elements exist - including mechanisms and structures, for which we have limited knowledge following the complexity of accessing them through the various levels of stratification (Wynn & Williams, 2012). There are two other general elements namely events and experiences. A definition of each of these elements is provided below.

a) Structure

Critical realism defines structure as a set of internally related objects or practices which are ontologically decoupled from the events they produce (Bhaskar, 2008; Smith, 2006; Wynn & Williams, 2012). Structures are also referred to as generative mechanisms as they give rise to certain causal powers, tendencies, or ways of acting (John Mingers, 2002). A structure cannot be reduced to the entities that constitutes it, just like water (H2O) cannot be reduced to either hydrogen or oxygen (Smith, 2006; Wynn & Williams, 2012). This implies that the structure acquires distinct properties from its entities, which results to a new identity (Wynn & Williams, 2012).

Information systems research operates in a social-technical environment, which is bound by several interacting structures (Wynn & Williams, 2012). These structures are of two types, namely, social and physical structures. Social structures affect social activities through agents, which in-turn reproduces and transforms them (John Mingers, 2002; Wynn & Williams, 2012). The occurrence of social activities acts as proof to the existence of social structures, though unobservable at times (John Mingers, 2002). Physical structures vary from social structures by having independence from agents who may not be able to interpret or observe activities within the structure, which protects them from being transformed by social activities, and thus enduring across periods of time (Smith, 2006; Wynn & Williams, 2012).

b) Mechanisms

Critical realism defines mechanisms as techniques of acting on entities, which are categorized into causal powers and tendencies (Smith, 2006; Wynn & Williams, 2012). They could be likened to the connections between variables, from which outcomes emerge (Fox, 2009). Causal powers arise from the nature of entities, which have inherent "dispositions, capacities, and potentials to do certain things, and not others", and are not necessarily attributable to human actors (Easton, 2010; Wynn & Williams, 2012). Tendencies describe specific actions that are typical of a given class, species, or type of thing within a given mechanism (Bhaskar, 2008; Wynn & Williams, 2012). Since tendencies are typical, they may or may not happen, which implies that they are just a plausible course of action, and thus may never be realized depending on how mechanisms are enacted (Bhaskar, 2008; Wynn & Williams, 2012).

c) Events

Critical realism defines an event as an action or occurrence that arises from the enactment of some causal powers or tendencies, which remains ontologically distinct from the generative structures and mechanisms (Easton, 2010; Wynn & Williams, 2012). However, in certain cases the desired change may not occur following the counteracting effects on other mechanism(s), which could be regarded as a counteracting tendency (Easton, 2010; Wynn & Williams, 2012). In other cases, the activation of one mechanism may have exacerbating effect(s) on other mechanism(s), which may change the result of the actual event, but which may also be counteracted by another mechanism (Wynn & Williams, 2012).

d) Experiences

Critical realism defines experiences as observable events, either directly or indirectly through sensory perceptions or via sensory enhancers, and takes place in the empirical domain (Easton, 2010; Wynn & Williams, 2012). Observation is fallible, since what is observable is often a subset of the actual reality, making it impossible to gain a full understanding of the social context (Easton, 2010; Wynn & Williams, 2012). Research therefore entails assuming the existence of certain structures and mechanisms, followed by a quest to find direct observations that support the existing interpretations or provide alternatives (Wynn & Williams, 2012).

3.4.2.3 Power Emergence

Critical realism is based on the ontology that entities/objects are independent and irreducible to the mechanisms that form them (Easton, 2010; Wynn & Williams, 2012). Entities or objects are the building blocks of a critical realist explanation, and include, but not limited to; organizations, people, relationships, resources, and information systems (Easton, 2010). They contain unique properties, causal powers and liabilities that are determined by the structure of the mechanisms that form them (Easton, 2010; Wynn & Williams, 2012). Critical realism research seeks to identify the entities involved in a study, their relationships, and the powers and liabilities they possess (Easton, 2010).

3.4.2.4 Open Systems

Critical realism defines reality as an open system that is operated by natural laws, and no constant conjunction of events prevail (Bhaskar, 2008; Wynn & Williams, 2012). This implies that in an open system, neither the experimental establishment, nor the practical application of human knowledge can be sustained (Bhaskar, 2008). To overcome this limitation, closed systems are designed with the aim of controlling the contextual conditions and exogenous influences, and making visible emergent experiences which can be replicated (Wynn & Williams, 2012). Social systems rarely present accessible reality, that is controllable allowing for spontaneous closure, and as a result, becomes difficult to conduct closed system experiments (Easton, 2010; Wynn & Williams, 2012). This is mainly because the boundaries of social systems are fluid and permeable, which causes the critical realist to move beyond events to uncover their causal explanation (Smith, 2006; Wynn & Williams, 2012).

3.4.3 Scientific Modes of Inference and Reasoning

Critical realism accepts that reality, which comprises of the world and its entities exists independent of human knowledge or subjective beliefs, and is not limited to human perceptions or experiences (Bhaskar, 2008; John Mingers, 2002; Wynn & Williams, 2012). It also accepts that this reality comprises of complex interactions and humans only experience a portion of it, following their limitation in apprehending, characterizing or even measuring it (John Mingers, 2002; Wynn & Williams, 2012).

This complexity arises from the fact that reality has no voice of its own (Danermark, Ekstrom, Jakobsen, & Karlsson, 2002). This calls for scientific enquiry, which is not limited to observing, registering and reporting of reality. There is also need for reasoning, which is cognitive process that is deemed as a fundamental precondition for scientific knowledge formation. It entails human ability to analyze, abstract, relate, interpret and draw conclusions. According to the philosophers of science, reasoning comprises of four main features, namely; feeling, intuition, imagination and creativity (Danermark et al., 2002).

Reasoning is based on the concept of inference or thought operation. Thinking or thought operation is a prerequisite that helps us derive sense and meaning from our observations,

enabling us to interpret the particular in a context, with the aim of deriving conclusions about the general from observations of the individual. This is based on the concept of generality (Danermark et al., 2002). Inference is a way of reasoning aimed at finding answers to questions such as: "What does this mean? What follows from this? What must exist for this to be possible?" (Danermark et al., 2002).

There are four different modes of inference each with unique thought operation: deduction, induction, abduction and retroduction. Deduction comprises of two aspects, inference and logic. Deductive inference entails interpretations or conclusions that are arrived at by use of a strictly logical way from a certain premise. Deductive logic helps in examining the logical validity of a scientific argument (conclusion or interpretation) regardless of the research methods/traditions in use. However, it does not examine the reliability of these arguments. It is the opposite of induction since it's starting point is the conclusion, which is the end-point of induction (Danermark et al., 2002).

Deduction and Induction are both referred to as formal logic. However, there is a big variation in the knowledge that emerges and how it is derived. In induction, the conclusion does not necessarily follow from the premise, and contributes to new/additional knowledge beyond what logically follows from the premise. It is a move from the known aimed towards the unknown (Danermark et al., 2002). In this study, induction is applied in the sense of inductive logic. In social science the concept of inductive is also used to describe a certain form of research procedure.

Abduction is a mode of inference that places and interprets original ideas about a phenomenon in the frame of a new set of ideas - which have the form of conceptual framework or theory, resulting in something different, which is possibly a more developed or deeper conception of the phenomenon. It builds on creativity and imagination in forming associations, something that differentiates it from deduction and induction. This requires a creative reasoning mindset, which helps discern relations and connections that are not evident or obvious, that is being able to 'see something in something else'. These new ideas, meanings and connections are not infallible.

Also, it is not possible to test the ultimate truth of such ideas, since there are no ultimately true theories or rules to test them (Danermark et al., 2002).

Retroduction is a mode of inference that seeks to understand what is basically characteristic and constitutive of the structures and internal relations associated with an individual phenomenon. It is not a formalized mode of inference as is the case for deduction and induction. However, in terms of thought operation, it is similar to these formalized modes of inference and abduction, since it enables one to move from knowledge of one thing to something else (Danermark et al., 2002).

The validity of any of these modes of inference is determined by its structure. Scientific inference sometimes entails the use of strictly formalized rules for logical argument and argumentation. Formalization is manifested through the use of different models and symbolic language. These are substituted and/or complemented by the use of abstraction, imagination and creativity by the researcher (Danermark et al., 2002).

These modes of inference are complementary to each other in research practice. For instance, deduction, which is regarded as the only strictly logical mode of inference, provides universal guidelines for testing the validity of conclusions drawn by other modes, for instance, retroduction. Table 3.1 below helps in determining the variation of the four different modes of inference based on the work of Danermark et al. (2002).

	Deduction	**Induction**	**Abduction**	**Retroduction**
Fundamental structure/thought operations	To derive logically valid conclusions from given premises. To derive knowledge of individual phenomena from universal laws	From a number of observations to draw universally valid conclusions about a whole population. To see similarities in a number of observations and draw the conclusion that these similarities also apply to non-studied cases. From observed co-variants to draw conclusions about law-like relations.	To interpret and recontextualize individual phenomena within a conceptual framework or a set of ideas. To be able to understand something in a new way by observing and interpreting this something in a new conceptual framework.	From a description and analysis of concrete phenomena to reconstruct the basic conditions for these phenomena to be what they are (move from concrete to abstract and the reverse). By way of thought operations and counterfactual thinking to argue towards transfactual conditions.
Formal logic	Yes	Yes	Yes and No	No
Strict logical inference	Yes	No	No	No
The central issue	What are the logical conclusions of the premises?	What is the element common for a number of observed entities and is it true also of a larger population?	What meaning is given to something interpreted within a particular conceptual framework?	What qualities must exist for something to be possible?
Strength	Provides rules and	Provides guidance in	Provides guidance for the	Provides knowledge of

	guidance for logical derivations and investigations of the logical validity in all argument.	connection with empirical generalizations, and possibilities to calculate, in part, the precision of such generalizations.	interpretative processes by which we ascribe meaning to events in relation to a larger context.	transfactual conditions, structures and mechanisms that cannot be directly observed in the domain of the empirical.
Limitations	Deduction does not say anything new about reality beyond what is already in the premises. It is strictly analytical.	Inductive inference can never be either analytically or empirically certain = the internal limitations of induction. Induction is restricted to conclusions at the empirical level = the external limitations of induction.	There are no fixed criteria from which it is possible to assess in a definite way the validity of an abductive conclusion.	There are no fixed criteria from which it would be possible to assess in a definite way the validity of a retroductive conclusion.
Important quality on the part of the researcher	Logical reasoning ability	Ability to master statistical analysis	Creativity and imagination	Ability to abstract
Examples	If A then B A Thus: B	From an investigation of the attitude of a representative sample of Swedes, draw the conclusion that 30% of the Swedish population is in favour of the EU.	Karl Marx reinterpretation/ redescription of the history of humankind from the historical materialist view.	For a ritual to be just a ritual there must exist, *inter alia*, emotionally loaded symbols and common notions of inviolable/sacred values.

Table 3.1: Modes of inference

3.5 Models for an explanatory social science

The aim of social science is to explain social conditions. There are several models that could assist in this: Popper-Hempel explanatory model, DREI, RRRE, Margaret Archer's morphogenetic approach, Pawson & Tilley's realist evaluation model, and Danemark's explanatory model of critical realism.

3.5.1 Popper-Hempel explanatory model

This is a model whose central modes of inference is deduction and induction. This model's understanding of causality is based on David Hume's (Hume 1966) empirical definition, which assumes that causality is based on empirical regularities between events, and that it is these regularities that we observe to form causal explanations. These explanations require the knowledge of universal conformity to a law or at least law-like regularities. The structure of this model is based on what they refer to as explanans and explanandum (Danermark et al., 2002).

Explanans are the required conditions and comprise of Universal law(s), Framework condition(s) and triggering causes. The explandum is a description of what is to be explained. This model results in empirical inductive generalizations, which raise several compatibility issues between this model and social science. First, in social science, there is hardly any conformity to law and that the regularities that we may occasionally observe only occur within a certain probability.

This reduces reality to events and empirical observation. Second, the explanations derived from this model only describe law-like/statistical relations and fail to identify any causal mechanisms. Based on these issues, this model is not ideal for critical realism (Danermark et al., 2002).

3.5.2 DREI

To address the impossibility of true experimentation in social sciences, Bhaskar suggests the use of transcendental arguments. Following this, he developed the DREI model. It is defined as the **D**escription of law-like behaviour, **R**etroduction, which involves using analogies to create possible explanations of the observable behavior in the phenomena, **E**laboration and elimination of alternative explanations, and **I**dentification of the underlying causal mechanism(s) in an empirically controlled environment (J. Mingers, 2000; Steinmetz, 1998).

3.5.3 RRRE

RRREI model (**R**esolution, **R**edescription, **R**etrodiction, **E**limination) is another model by Bhaskar that is also based on abduction and retroduction, and which seeks to address the shortfalls of DREI (Bhaskar, 1998; Carlson, 2009, p. 64). It makes use of the first three stages of DREI, but eliminates the last one (identification) since it is impossible in social science. It adds an important step to DREI, that of resolving complex events into its components through abstraction. This is necessary because there is often more than one mechanism producing an event, and also that events are overdetermined when we consider a nexus of events by the totality of causal mechanisms in an open system (Steinmetz, 1998).

This five stage process entails the following; (1) **R**esolution of a complex event into its components (causal analysis), (2) **R**edescription of component causes to make it relevant to certain theories, concepts or issues, (3) **R**etrodiction to possible (antecedent) causes of components via independently validated normic statements (postulation of hypothetical mechanisms or structures), and (4) **E**limination of alternative possible causes of components (Bhaskar, 1998; Carlson, 2009).

3.5.4 Margaret Archer's morphogenetic approach

The morphogenetic approach is a theory which aims at understanding emergence and analytical dualism between structure, culture and agency. It provides a bridge to resolve the methodological complexity between structure and agency, and culture and agency (analytical dualism - separates

social integration and system integration) without conflating them (non-conflationary theorizing). It consists of a three-stage cycle which involves structure, culture, and agency. It helps in understanding how new properties related to each of these emerge, interact and redefine each other over the course of time (Archer, 2007; Horrocks, 2009; Zeuner, 1999). This cycle involves, 1) structural or cultural conditioning, which is characterised by ideas/logical relations which are either contradictory or complementary 2) social or sociocultural interaction, characterised by material interests (use or escape from power) that causes change in interactions, and 3) social or cultural elaboration (Archer, 1995; Horrocks, 2009; Zeuner, 1999).

All the philosophical underpinnings of the morphogenetic approach are based on critical realism, except the first which is based on the three pillars of realist philosophy of social science. These pillars include; ontological stratification, epistemological relativism (relationalism), and judgmental rationality. Though realism provides a good framework for conceptualising structure, it does not say much about human agency (person) - who they are, what they care about and what causes them to act individually or collectively, a problem that is resolved by critical realism (Archer, 2007).

The morphogenetic approach shares the following assumptions with critical realism; i) Relationality - Reality is emergent and comprises of irreducible properties capable of exercising causal powers, ii) Structures are morphogenetic - structures are shaped by the interplay between their constituents, parts and human agency. This implies that society is open-ended and not "finalistic", iii) Value based rationality (*Wetrationalitat*), which is aimed at improving humanity iv) The relation between structure and agency presents a complexity that cannot be solved (transcended) through a stratified social ontology, as it consists of what Dahrendorf calls "vexatious fact of society", where we human agents shape it while it re-shapes us individually and collectively v) The problem of objectivity and subjectivity, which is similar to that of structure and agency as it deals with the relationship between structure and agency, in search of their causal powers from an ontological standpoint. Morphogenesis proposes that human reflexivity is what mediates between structure and agency. Reflexivity refers to the deliberations (internal) humans go through prior to acting (Archer, 2007).

In an attempt to further understand the assumptions of the morphogenetic approach, the following section compares the morphogenetic approach with Giddens' structuration theory, whose principles may appear similar, but are fundamentally different.

3.5.4.1 Structuration theory versus morphogenetic approach

Social structures are the product of rational actors, which once formulated end up shaping them influencing their interactions. Earlier theories gave more emphasis on structure and less emphasis on agency/rational actors. This disconnect is being resolved by two perspectives; morphogenetic approach which emerges from general systems theory in sociology, and structuration theory by Anthony Giddens. These perspectives both agree that agency and structure presuppose each other. However, they differ on how they theorize about the structuring and restructuring of social systems (Archer, 2010).

Structuration is concerned with the duality not dualism between structure and agency. This results to the notion of duality of structure which views structure as both a medium and outcome of social life (Giddens, 1979, p. 5). Structuration is premised on this notion and acknowledges that the constituting components are dynamic and ever changing, which implies that structuration will never reach development as it is always changing.

Despite this realization of structure, Giddens' structuration theory fails to reunite structure with human-action from within general systems theory. Structuration theory also disregards time, which is necessary for analytical history of systemic emergence (Archer, 2010).

The morphogenetic approach aims at connecting structure and agency and understanding the relationship between them. Morphogenesis is a process that entails the complex interactions that produce change in a systems form, structure or state. Its end product is structural elaboration, which varies from Giddens' social system, which is merely a "visible pattern" of recurrent social practices. Structure has properties which cannot be reduced to practices alone, though these are what generated them (Archer, 2010).

Morphogenesis suggests analytical dualism between structure and human action, since the emergent social-cultural system properties portray a disconnect between initial actions/interactions and their product - a complex system. Morphogenesis is also sequential as it also observes how the cycle between structural conditioning, social interaction and structural elaboration evolves. Morphogenesis has a different perspective to Giddens' structuration theory in the following three dichotomies: voluntarism and determinism, synchrony and diachrony, and individual and society (Archer, 2010). These will be discussed below;

a) The 'duality of structure' and voluntarism/determinism

Institutions are either causes of action (determinism) or embodiments of action (voluntarism). Structuration attempts at resolving the divide between voluntarism and determinism through one conceptual leap, the 'duality of structure'. In defining duality of structure, Giddens claims that structure is both a medium and an outcome. However, his concept fails in answering the when questions, particularly when there will be 'more determinism' (when actors are more transformative) and 'more voluntarism' (when actors are trapped into replication). It also fails to specify the strength of constraints, yet this is tackled adequately in sociology by appreciating that some properties engender more resistance to change than others (Archer, 2010). Another failure is on the understanding of the recursive effects of duality of structure, which he views as 'self-regulating properties' whereas others including Archer view them as 'emergent properties' (Archer, 2010).

Morphogenesis tackles the voluntarism/determinism divide differently. It seeks to analyse the strength/stringency of constraints and the degrees of freedom in varying structural contexts and for different social groups (Archer, 2010). It views voluntarism/determinism as one that oscillates between 1) hyperactivity of agency fueled by the volatility of society, and 2) rigid coherence of structural properties which are related to the recursiveness of social life (Archer, 2010).

b) 'Structuration' and synchrony/diachrony

Though Giddens agrees that structuration introduces temporality, which helps in resolving the divide between synchrony (static - chronic recursiveness) and diachrony (dynamic - total

transformation), he fails to investigate the interplay between structure and action yet the two closely presuppose each other (Archer, 2010). This is because he fails to acknowledge that the two work on different time intervals (Archer, 2010).

Morphogenesis proposes that structure and action operate on different time intervals based on two propositions: 1) structure predates action, which later transforms it, and 2) action predates structural elaboration (Archer, 2010). This proposal is supported by the five assumptions that the morphogenic approach shares with critical realism, and more particularly the second and fifth assumptions. The second assumption suggests that structures are morphogenic, to mean that structures are shaped by the interplay between their constituents, parts and human agency. The fifth assumption on the problem of objectivity and subjectivity supports the second proposition, which is mediated by reflexivity – the (internal) deliberations humans go through prior to acting (Archer, 2007).

c) Social systems and the individual/society dichotomy

The scope of structuration theory entails the study of the individual as a social product and the generation of society by human agency. He fails by conflating social integration with systemic integration when giving the parts-whole account where he argues that what integrates the individual into society is the same that integrates society (Archer, 2010).

Morphogenesis provides a different parts-whole account that the whole is impacted by the parts in two ways; 1) it emerges from them, and 2) it acts back upon them, though the force/impact takes time to detect since feedback takes time. Parts are also said to contain tension which affects the state of the whole (Archer, 2010).

3.5.5 Pawson & Tilley's realist evaluation model

Realist evaluation has an explanatory outcome, which helps describe "what works for whom, in what circumstances and in what respects, and how?" More specifically, it seeks to provide an understanding of how programmes bring change by probing the change apparatus within them. Programmes are an outcome of the foresight of policy-makers, and the imagination of practitioners and programme subjects/participants who operationalize and implement these policies into action. They are embedded within social systems, are open systems which cannot be

isolated or kept constant, and are self-transformational (Pawson & Tilley, 2004). They exist within pre-existing constants, which then constraints the choices available, and thus limiting the possibility of free will. This view should be applied consistently in realist evaluation research (Pawson & Tilley, 2004; Porter, 2015).

Realist evaluation seeks to understand the nature of these programmes, how they work, for whom, in what circumstances and the change that result from them. This is achieved by analysing the following linked concepts within a programme: mechanism, context, outcome pattern and context-mechanism-outcome (CMO) configuration (Pawson & Tilley, 2004).
Mechanisms describe what brings about change within a programme (Pawson & Tilley, 2004). In line with the view that reality is stratified (Bhaskar, 2008, p. 91), and in order to avoid being in contradistinction to this, it is important to make a clear distinction between structure and agency, which the realist evaluation method combines into the single entity of social mechanism (Porter, 2015).

Context describe the features of the conditions relevant to the working of the programme mechanisms (Pawson & Tilley, 2004). Context is not limited to location but is the contingent combination of interpersonal and social relationships, biological composition, technology, economy, demography, among other things where these mechanisms (causal powers) are embedded (Pawson & Tilley, 2004; Porter, 2015).

CMO pattern configuration consist of models/propositions that describe how programmes activate mechanisms including the conditions required for this change, the actors involved and the change that results which could be emergent events. These CMO configuration propositions/conjectures are developed and tested empirically (Pawson & Tilley, 2004).

However, programmes and the contexts in which they are placed, each contain their own mechanisms (Porter, 2015). Mechanisms are said to possess latent powers or structural transfactual causality, meaning that they endure regardless of whether they become manifest in open systems, and that they can be possessed unexercised, exercised unrealized or even realized unperceived. Mechanisms do not act in solitude, in that it is difficult to point a single mechanism

to an outcome or event, since these are usually the result of a combination of causal mechanisms (Bhaskar, 2008, pp. 1–9).

There are some mechanisms which will try to maintain the current state, as others seek to transform it towards the intended outcome or event. The former could be pre-existing/contextual mechanisms (social context which includes problem mechanisms that counteract programme mechanisms), while the latter could be programme mechanisms which are aimed at a certain desired outcome. In relation to this, there is also need to separate human agency from social mechanisms (Porter, 2015). This results in the modification of Pawson & Tilley's CMO configuration formula, which did not explicitly separate this. The adapted formula from Porter (2015) is as follows;

$$Contextual\ Mechanisms + Programme\ Mechanisms + Agency = Outcome$$

$$CM + PM + A=O$$

The major difference between this formula and the original formula by Pawson & Tilley is the separation of programme and social context mechanisms, and also the introduction of agency, which is not explicitly included in the realist evaluation CMO configuration.
The phases of Realist evaluation are described below;

3.5.5.1 Phase 1: Hypothesis generation

First, hypothesis generation that results in candidate CMO configurations/propositions. This mainly involves literature review and or initial interviews with programme architects or practitioners. The resulting information helps in identifying the CMO configurations within a programme. The resulting hypothesis/propositions seek to describe what works for whom and in what circumstances/conditions (Pawson & Tilley, 2004).

3.5.5.2 Phase 2: Data collection

Second, data collection on appropriate contexts, mechanisms, and outcomes. This involves collecting documents recommended by programme practitioners or architects, and conducting interviews with the various stakeholders. This process is guided by the outcome of the first phase, which could be referred to as the preliminary theories (Pawson & Tilley, 2004).

3.5.5.3 Phase 3: Data analysis

Third, data analysis which seeks to identify patterns and map these to initial theory. This stage involves subjecting the CMO configuration hypotheses/propositions to test using the data obtained in the second phase. The aim is to identify the various patterns of success and failure within and across a programme (Pawson & Tilley, 2004).

3.5.5.4 Phase 4: Theory Testing

Fourth, theory testing which aims at refining the understanding of CMO configurations using empirical findings. This aims at providing an assessment and interpretation of the analysis. This phase is often repeated several times with the aim of drawing closer to a more accurate explanation of the outcomes of a particular programme. Realist evaluation starts with a theory as illustrated in phase one, and ends in a theory, both of which seeks to describe what works, for whom, in what context and in what respect (Pawson & Tilley, 2004). However, it is important to note that testing of the observed mechanisms/theories is dependent on the entity's/researcher's capacity to bring about changes in reality. This is related to the intransitive nature of objects, where they continue to exist and act regardless of our knowledge about them (Bhaskar, 2008, pp. 11–14).

3.5.6 Danemark's Explanatory model

Explanatory model of critical realism is a model based on critical realism and whose central modes of inference is abduction and retroduction. It aims at identifying the underlying causal structures and mechanisms, and how these mechanisms cooperate under specific circumstances in enabling certain events and processes. It also applies deductive logic in analysing the scientific arguments that are formulated in the process. It consists of six stages which include: 1) Description 2) Analytical resolution 3) Abduction/theoretical redescription 4) Retroduction 5) Comparison between different theories and abstraction 6) Concretization and contextualization.

This model produces two types of knowledge about reality through abstraction and concretization, as each of these produces different type of knowledge. It moves from the concrete (stage 1), to the abstract (stage 2-5), and returns back to the concrete (stage 6). It is important to note that this is a guideline and not a template or prescription that must be followed as described.

The order of these stages can be switched and also, a researcher is at liberty to concentrate on certain stages than the rest (Danermark et al., 2002). These stages are described below.

The stages of an explanatory research based on critical realism

Stage 1: Description

This involves a description of events, which are often complex and composite. This description is enriched by the interpretations of the persons involved and how they describe the event/situation. In making these descriptions, there is need to makes use of both qualitative and quantitative methods (Danermark et al., 2002).

Stage 2: Analytical resolution

This involves identifying and separating the various components, aspects or dimensions of the phenomenon in question. This is followed by a selection of the components to be studied, since it is often impossible to study all components within a phenomenon (Danermark et al., 2002).

Stage 3: Abduction/theoretical redescription

This involves the interpretation and redescription of the different components/aspects using hypothetical conceptual frameworks and theories in search of the underlying structures and relations. This results in several and varying theoretical interpretations and explanations, which need to be compared and if possible, integrated together (Danermark et al., 2002).

Stage 4: Retroduction

This seeks to understand what fundamentally constitutes the structures and relations of the various components/aspects (X) identified and selected for study in stage 3. Examples of the questions asked: "How is X possible? What causal mechanisms are related to X?" (Danermark et al., 2002).

Stage 5: Comparison between different theories and abstraction

This involves reviewing the explanatory power of the mechanisms and structures described in stage 3 and 4. The aim is to identify the theory with the greater explanatory power if the theories under review are competitive. Otherwise, the theories complement each other since they do not have exactly the same focus. At times, this stage is combined with stage 4 (Danermark et al., 2002).

Stage 6: Concretization and contextualization

Concretization is the process of studying how different structures and mechanisms manifest themselves in concrete situations. This includes how they interact with other mechanisms at

different levels of encounter, but under specific conditions though allowing for accidental circumstances. The outcome contributes to explanations of concrete events and processes. These explanations need to separate between observations carried out under structural conditions and accidental circumstances (Danermark et al., 2002).

Table 3.2: Stages of Danemark's explanatory model

3.5.7 Preferred critical realism model

Application of Bhaskar's DREI model is a challenge in social sciences especially on the final stage that is based on experimentation. This is not possible in the social science context and presents a challenge in achieving Bhaskar's goal of a model that supports experimentation in social sciences (Collier, 1994, pp. 31–79). RRRE also faces a similar challenge to DREI, and both have been found to be complex and difficult to apply in social science (Danermark et al., 2002). This study found the similar complexity in Margaret Archer's morphogenetic approach, especially because there were insufficient publications in information systems that could have helped in resolving some of the complexities.

Popper & Hempel model doesn't fit a social science study and consequently this study, since it results in law-like empirical inductive generalizations, which can only be arrived at in an environment with observable regularities. In addition to this, it focuses on statistical relations instead of causal mechanisms which are important for social science (Danermark et al., 2002).

Danermark's Explanatory model would have applied for this study, except for the challenge arising from the fifth and the sixth stage. It was difficult to identify publications that applied this model, which would have helped in clarifying these stages. Stage five depends on hypothetical conceptual frameworks and theories that are formed in stage three. The challenge was in understanding how to form several theories from the same context, and also how to conduct stage six, which entails concretization and contextualization. It would be easier to have examples or social science research publications that describe how they studied the various structures and mechanisms at different levels of encounter, and how they were able to do this in specific conditions.

Pawson & Tilley's realist evaluation model was found to have the least complexities, and had several social science publications that mapped well with this study. Also, its assumptions blended well with Fereday & Muir-Cochrane's (2006) five-stage thematic analysis model, which was applied to implement the third phase of the realist evaluation model on data analysis.

3.6 Case study design from a critical realism perspective

In information systems, the main aim of critical realism is to formulate causal explanations that explain the way things act and how they are capable of doing so in a socio-technical context. To achieve this, case study method is preferred among many critical realism researchers (Easton, 2010; John Mingers, 2004; Smith, 2006; Wynn & Williams, 2012). Since this research aims at identifying causal mechanisms and providing in-depth causal explanations that will assist in formulating theory, the case study method is preferred. This is motivated by the following description of the case study method within the context of critical realism.

Case study research approach has been applied in the information systems discipline for more than two decades (Benbasat & Zmud, 1999; Dube & Pare, 2003; Easton, 2010; Flyvbjerg, 2006; Lee, 1989; Wynn & Williams, 2012). It focuses on studying a phenomenon within an organization and the relationships therein, which comprise of complex structures that are difficult to access, and which cannot be studied outside the context of occurrence (Dube & Pare, 2003; Easton, 2010). This follows its ability to study one or a small number of social entities or situations within a real-life context using multiple sources of data, where the boundaries between the phenomenon and the context may not be clearly evident (Wynn & Williams, 2012).

In order to tease out and disentangle these complexities, it starts by identifying the research questions, followed by the case selection criteria, which includes the boundary definition. The selected case comprises of a single or manageable number of entities to obtain data (Dube & Pare, 2003; Easton, 2010; Eisenhardt, 1989; Kvale, 1996).

The focus of data collection is individual actors, with the aim of understanding what produces change, or rather, what causes the events associated with the phenomenon to occur. To achieve this, while continuously seeking clarity on what was said by the actors, there is need to document all impressions that emerge including a detailed description of the data collection methods and

procedures (Dube & Pare, 2003; Easton, 2010; Eisenhardt, 1989; Kvale, 1996). Data is mainly collected through mixed methods/triangulation which often comprises of interviews, observation and document review (Dube & Pare, 2003; Wynn & Williams, 2012).

Once data is collected, critical realism requires the use of thoughtful causal language to describe observations, with the aim of fulfilling the objective of understanding things as they are (Easton, 2010). The case is written iteratively, giving a holistic description of the observed entities, which provides causal explanations about the phenomena in question (Dube & Pare, 2003; Easton, 2010; Eisenhardt, 1989; Flyvbjerg, 2006). Causal explanations should focus more on the cause and consequences of the problem, than the symptoms and frequency. These explanations - though long and complex, should retain diversity and not link them to theories, allowing the reader to make their own interpretation (Flyvbjerg, 2006). Once the case study is complete, it must be able to "stand on its own" by giving an in-depth and comprehensive description of a phenomenon (Easton, 2010).

3.7 Case study methodological principles

Wynn & Williams (2012) formulated five methodological principles for conducting critical realism research using explanatory case studies. These principles are not prescriptive but interdependent, providing a summary of useful insights and requirements that may assist information systems researchers to adopt critical realism more effectively. These principles include: explication of events, explication of structure and context, retroduction, empirical corroboration and triangulation/multimethods. Each of these principles is described below.

3.7.1 Explication of Events

This principle describes the need to identify the detailed and explicit aspects of the events under study through the abstraction of experiences, as the basis of analysis (Smith, 2006; Wynn & Williams, 2012). These detailed aspects include key actions and outcomes, the structural components, and sequence of events. They are aimed at developing a causal, transitive explanation of the complex socio-technical phenomenon under investigation (Bhaskar, 2008; Wynn & Williams, 2012).

3.7.2 Explication of Structure and Context

This principle helps in identifying analytically components of the structure, variations in contextual influences, and potentially activated mechanisms that are causally relevant in an open system (Smith, 2006; Wynn & Williams, 2012). This decomposition of structure into its constituent parts such as actors, rules and relationships, helps in identifying and describing the causal tendencies that generate events. Contextual influences include social, physical, artificial, or symbolic entities and the relationships among them. These entities are also referred to as structured entities and are similar to mechanisms since they have causal powers and liabilities to act, and are usually not directly observable, but are known from their artifacts and effects (Easton, 2010; Wynn & Williams, 2012).

3.7.3 Retroduction

Retroduction involves "moving back", and aims at identifying the conditions that must be true for an event to occur (Easton, 2010). Retroduction is a form of inference in which events are explained by postulating, identifying and verifying the causal mechanisms that are capable of generating them. It is the epistemological process that critical realists recognize (Bhaskar, 2008; John Mingers, 2002).

3.7.4 Empirical Corroboration

It is not possible to verify a theory as true or false, but instead as more or less probable, by assessing the tests it has withstood, a process known as corroboration. This principle seeks to use empirical evidence to improve and validate the proposed mechanisms with sufficient causal depth and explanatory power, thus enabling it to represent reality more adequately (Popper, 2014; Wynn & Williams, 2012). This is based on the ontological assumption that reality is stratified and independent of human intervention, the epistemological assumption that mechanisms may not be directly unobservable, and the potential of having multiple causal explanations about a phenomenon (Bhaskar, 2008; John Mingers, 2002; Smith, 2006; Wynn & Williams, 2012).

3.7.5 Triangulation/Multimethods

This principle expresses the need to adopt multiple approaches (data sources, theories, and methods) while performing critical realism research, and in particular, to support causal analysis (Wynn & Williams, 2012). Critical realism conjectures that reality exists independent of human

intervention or knowledge and can only be fallibly observed and accessed, and that science is capable of approaching this reality in an enlightened manner. It also conjectures that reality comprises of different types of structures, each comprising of unique properties, causal powers and tendencies. Following their uniqueness in nature, each structure calls for a unique means of causal analysis, which can be achieved through the application of different methods and perspectives (Wynn & Williams, 2012).

3.8 Conclusion

This chapter helped in describing the assumptions of critical realism, the various critical realism models, and how case study approach can be applied in a critical realism based research, as a tool for developing compelling causal explanations. Following a discussion of the various models in section 3.6.7 above, Pawson & Tilley's realist evaluation model was preferred for this study. The following chapter provides a detailed description of the framework that will be applied in identifying the causal mechanisms that affect institutionalization and sustainability of KODI.

4. Institution, Institutionalization and Institutional Theory

This chapter provides a detailed description of what an institution comprises of and what institutionalization entails. This is followed by a discussion of institutional theory and its compatibility to critical realism. The findings in these discussions help in developing a framework that guides in analyzing open data institutionalization.

4.1 Institution

An institution is defined as a social order or pattern that comprises of social needs, adaptive to external and internal pressures, and has attained a certain state or property (Jepperson, 1991; Selznick, 2011). This social order or pattern comprises of organised and established procedures, which represent standardised interaction sequences or procedures (Jepperson, 1991). These standardised procedures are driven by institutional forces, which transform individual interests and desires. These forces help in creating an environment for action and influence certain behaviors to either persist through actions that reinforce existing conventions or change through actions that alter them.

This implies that institutions are reproduced through the daily activities of individuals (Powell & Colyvas, 2008; Scott, 2008a). As individuals carry out their daily activities, they at times encounter puzzles or anomalies. If the environment allows, they go ahead to problematize these puzzles or anomalies by theorizing, which involves formulation of questions and development of answers to these questions. Once theories are formulated, the participating individuals ascribe meaning to these theories and, in so doing, develop and reproduce taken-for-granted understandings (Powell & Colyvas, 2008).

The institution is characterized by a non-monolithic environment, which is varied and conflicted, vague or weak boundaries that allow alternative logics to permeate and support divergent models of behavior, and elements (regulative, normative, and culture-cognitive) that are not often aligned and at times undermine each other's effects. Regulative elements are characterised by rule setting, monitoring, and sanctioning activities. Normative elements are characterized by definition of standards and norms, which are prescriptive, evaluative and obligatory. Cultural-cognitive elements are characterized by shared ideas, concepts and meanings about social reality

(Scott, 2005, 2008a). This implies that there are many institutional elements at play within an organisation, and it is imperative for the institutional scholar to determine what elements are at play within the context in question, and the extent to which they reinforce or undercut each other (Scott, 2008a).

Institutions are inevitably subjected to external (exogenous) and internal (endogenous) change. Exogenous change arises from economic or social factors that disrupt the existing rules and understandings. Endogenous change emanates from "gaps or mismatches between more macro systems and micro activities in response to local circumstances, inconsistencies existing between institutional elements or competing frameworks, and persisting poor performance levels in relation to expectations" (Dacin, Goodstein, & Scott, 2002; Sewell, 1992).

These institutions comprise of actors who are affected by institutional demands. It is expected that they might not passively comply to these demands, but could respond using any of the following strategies including compromise, avoidance, defiance and manipulation (Scott, 2005).

4.2 Institutionalization

Institutionalization is a process that occurs to an organisation over time leaving behind a distinctive history of people and ways it has adapted to its environment. It is the carrier of structure, which comprises of values such as group integrity and as a result, is regarded to some degree to be an end in itself (Selznick, 2011). It could be as simple as a network relation between rational actors, which constrains them, and also provides opportunities for their operations.

These networks are created by rational actors, and their history and external determinants determines their level of institutionalization, which results in considerable *path dependence* (Meyer, 2006). It is characterized by reciprocal typification (social construction based on standard assumptions) of habitualized actions by various types of actors. The outcome of this process is the institution (Tolbert & Zucker, 1996). This implies that institutionalization is a core process in the creation and perpetuation of enduring institutions. Habitualized action is a behavior that is formed empirically and adopted by an actor(s) to solve certain recurring problems. Once an action is habitualized, it requires minimal decision-making effort to execute (Tolbert & Zucker, 1996). Reciprocal typification involves forming generalized/shared meanings

and definitions of the habitualized actions by the various actors at play (Schutz, 1967, 1971). This process of generalizing meanings and definition of the various actions being habitualized is referred to as 'objectification', and is viewed as one of the key components of institutionalization (Zucker, 1977).

Institutions are formed through three sequential processes of habitualization, objectification, and sedimentation. These three processes are described as the components of institutionalization. Figure 4.1 below illustrates the actions involved between these processes.

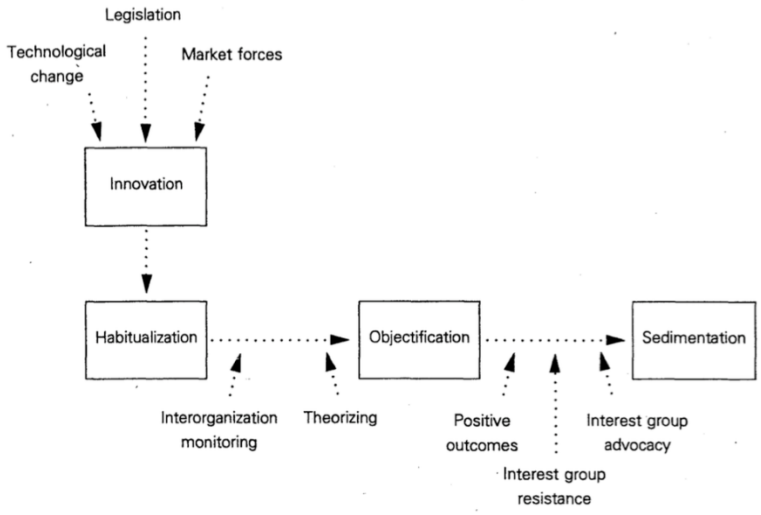

Figure 4.1: Component processes of institutionalization (Tolbert & Zucker, 1996)

In relation to adoption the structures, events and the associated relationships mentioned above are said to undergo three phases, first one being the early phase of partial acceptance, the second one being the middle phase of rapid diffusion and wider acceptance, and the last one being the phase of saturation and complete legitimation.

The following three sections give a more detailed description of this sequential model, which are also referred to as pre-institutionalization, semi-institutionalization and full-institutionalization respectively. These sections also discuss the fundamental dimensions of institutionalization, which include theorization, diffusion, exteriority and retention. The fourth section provides several propositions that would help ensure full-institutionalization of structures operating within or between organizations.

4.2.1 Habitualization

Habitualization is the process of developing replicable problem-solving behaviors, which become internalised and can be evoked with minimal decision making effort by the actor(s) who respond to particular stimuli. This process involves innovation of new structural arrangements, and the formalization of these arrangements through organisation policies and procedures. This process is also refered to as the pre-institutionalization stage (Tolbert & Zucker, 1996), and is regarded to as the early phase of partial acceptance (Zucker, 1987).

This phase is characterised by a low number of adopters whose organisations are possibly similar, interconnected and face similar circumstances. However, despite these commonalities, these adopters are likely to implement these structures differently. In the case of non-adopters, knowledge about the operation and purpose of the structures is quite limited, which is partly as a result of limited or lack of frequent interaction with the adopters (Tolbert & Zucker, 1996).

4.2.2 Objectification

Objectification is a process that entails development of shared social meanings attached to these behaviours, which follows a move towards a more permanent and widespread status through the development of some degree of social consensus among organizational decision-makers. These shared social meanings assist in the replication of actions to other contexts different from where they were designed originally. This stage is also referred to as semi-institutionalization (Tolbert & Zucker, 1996), and is regarded to as the middle phase of rapid diffusion and wider acceptance (Zucker, 1987). This consensus among the decision makers arises from two enabling mechanisms namely risk assessment and champions who are charged with generating awareness and developing theories.

Risk assessment can be done in various ways including the use of evidence obtained from print media or observations with the aim of assessing the risks involved in adopting a new structure. The outcomes of competitors implementing the structures in question greatly determine the adoption decision. This observation of other organisations' behaviour after adoption is important as it helps decision makers conduct the cost benefit analysis and make more informed decisions. This approach is less costly and involves fewer social resources (Tolbert & Zucker, 1996).

Champions are actors whose role is to generate awareness about a consistent pattern of problems that result in dissatisfaction or failing that has been observed in a number of organizations. This is achieved by creating a generic organizational problem, and the identification of the category of organisational actors affected by the problem. They are also charged with the role of developing theories that provide a diagnosis of the identified problems, which should be compatible with a particular structure that ensures general cognitive and normative legitimacy (Tolbert & Zucker, 1996).

The objectification (semi-institutionalization) phase is characterized by a shift in diffusion from simple imitation to a more normative base, and a decline in the variance in the form that structures take in different organizations (Tolbert & Zucker, 1996).

4.2.3 Sedimentation

Sedimentation is a process through which actions acquire the quality of exteriority (Tolbert & Zucker, 1996). Exteriority relates to the historical continuity of typifications, and refers to the degree to which typifications are 'experienced as possessing a reality of their own, a reality that confronts the individual as an external and coercive fact' (Zucker, 1977). This continuity is manifested by the reaction of new actors, who treat them as 'social givens', as they have no knowledge of the origin of such typifications (Berger & Luckmann, 1967; Tolbert, 1988). This process is also referred to as the full-institutionalization stage (Tolbert & Zucker, 1996), and is regarded to as the phase of saturation and complete legitimation (Zucker, 1987).

It is characterised by survival of structures across generations of organizational members (perpetuation), and complete spread or diffusion of structures across appropriate adopters (actors) (Tolbert & Zucker, 1996). Diffusion and retention of structures is affected by the existence of a set of actors who are adversely affected by the structures and who are in a position to resist them. It is also affected by the absence of demonstrable results associated with a structure, when the link between structure and intended outcomes is quite weak. However, such structures are still adopted following the theorization and promotion of champions who argue that the new structures are more promising and that the costs associated with the change are relatively low (Tolbert & Zucker, 1996). Therefore, for a structure to be fully institutionalized, there is need to have minimal resistance by affected actors, continued support and promotion by

champions and advocacy groups, and a clear and positive correlation or link with the desired outcomes (Tolbert & Zucker, 1996).

4.2.4 Institutionalization theory propositions

There are several propositions that have been derived from literature, regarding the proposed impact of certain actions on institutionalization. They are as follows;

- An increase in the degree of objectification and exteriority of an action results in an increase in the degree of institutionalization (Zucker, 1977).
- An increase in the degree of institutionalization results in an increase in the degree of action transmission, action maintenance, and resistance of that action to change (Zucker, 1977).
- An increase in the institutionalization of a routine results in an increase in the readiness and ease of transmission of such a routine to new employees (Nelson & Winter, 1982; Tolbert, 1988).
- The level of institutionalization varies depending on the configuration of a set of sequential processes that constitute the sequential model of institutionalization (Nelson & Winter, 1982; Tolbert, 1988).
- The broader the range of organizations for which a given structure is theorized, the more challenging it is to provide convincing evidence of a structure's effectiveness, which results in a lower level of institutionalization (Tolbert & Zucker, 1996).
- The higher the number of champions or champion groups, the less the resistance to change by affected actors, which results in a higher level of institutionalization (Tolbert & Zucker, 1996).
- The higher the investment costs in adopting change, the less the resistance to change, which results in a higher degree of institutionalization (Tolbert & Zucker, 1996).
- The stronger the incentives to maintain a structure, the higher the degree of institutionalization (Tolbert & Zucker, 1996).

These propositions are useful in understanding what impacts institutionalization and how full-institutionalization can be realized.

4.3 Institutional theory

Institution refers to "both the organizations and the rules used to structure patterns of interaction within and across organizations" (Ostrom, 2007, p. 22). This implies that focus is beyond the formal organisations which are at times characterised by bureaucracies, and includes interactions between people within an organisation and those in other organisations, with the aim of understanding the rules and strategies that govern their behaviour. It is these rules and strategies that impact on policies, which are either public, private or both (Kraft & Scott R. Furlong, 2015).

Institutional theory is a theoretical posture that examines the formation of processes and mechanisms through which structures, rules, schemas and routines become engrained as authoritative guidelines, resulting in stabilised social behaviour within an institution (Kraft & Scott R. Furlong, 2015; Scott, 2004, pp. 408–414). It also studies how such systems deteriorate and collapse (deinstitutionalization), and how their remnants affect successive systems. Change often arises from the collective effort of disadvantaged actors who challenge existing systems and truths. It also arises when current boundaries are breached, allowing for ideas and actors from one domain to flow to another (Scott, 2004, pp. 408–414). It also seeks to understand the realist 'mechanisms' through which local structures conform to wider models. Related to this, institutionalists also seek to understand the linkages that powerful interested actors place between policy and practice (Scott, 2008b). The field of macro-social research is shifting towards an enquiry on the origins of the institutional models involved within organisations (Scott, 2008b).

DiMaggio (1982) defines the process of forming an institution or structuration as one that comprises of the following stages: increasing participation among rational actors within an organisational field, emergence of inter-organizational structures of domination and patterns of coalition, increased information load which these organisations must contend, and development of mutual awareness among actors within an organisational field of their involvement in a common enterprise (DiMaggio & Powell, 1983).

Rational actors are influenced by the institutional models they are subjected to. This is despite the variation in resource capacity and culture between organisations. Examples include schools

and hospitals whose structures across the world seem to share standard models. Professional bodies and associations seem to further this cause. People are eager to join these bodies and associations as actors but unfortunately, they seem to be more focused in being actors than in acting which implies that there are many who do not have sound knowledge, understanding and experience required for members of these bodies or professions. The modern world has placed enormous premium on actorhood (Scott, 2008b).

4.3.1 Institutional assumptions

Institutions shape and govern social behaviour. This is supported by four main assumptions that institutions: 1) are governance structures comprising of rules and authoritative guidelines, 2) require legitimacy for survival which results from individual, group or organisation conformance, 3) suffer from inertia implying that they tend to resist change, 4) are shaped by their history since past structures constrain and guide new arrangements (Scott, 2004, pp. 408–414).

4.3.2 Institutional isomorphism

Related to these assumptions is the realization that the more rational actors try to change their organisations, the more similar they become like other organisations in the same field. This outcome is referred to as bureaucratization, which is change that results in similar organisations but not necessarily more efficient ones. According to Giddens (1979), bureaucracy emerges out of structuration of organisational fields. It emerges from three isomorphic processes: coercive, mimetic and normative. These processes form the theory of institutional isomorphism, which attempts to explain homogeneity within organisations of the same field that are commonly coupled with lack of innovation and frustration of power (DiMaggio & Powell, 1983).

The three isomorphic processes are also referred to as the mechanisms of institutional isomorphic change. *Coercive isomorphism* arises from formal and informal external pressures to an organisation either from other organisations to which there is a dependency or from the society in which they function. These pressures are facilitated by common legal environment, which could be politically formulated, and which tend to result in homogeneous organisations within a particular field. *Mimetic isomorphism* emanates from a coercive authority and is characterised by imitation, especially in areas that have a lot of uncertainties. A good example is

technology or ambiguous goals which results in some organisations modeling themselves on others. *Normative isomorphism* emerges primarily from professionalization. Larson (1977) and Collins (1979) refer professionalization as a collective effort by people within an occupation to define the conditions and methods of their work in an effort to control the creation of producers, in an effort to create legitimacy of their autonomy (DiMaggio & Powell, 1983).

Meyer (1979) and Fennell (1980) identify two distinct types of isomorphism: competitive and institutional. *Competitive isomorphism* is based on the assumption of a free and open market and partially explains bureaucratization according to Weber. However, it fails to fully explain the modern world of organisations. Kanter (1972) introduced *institutional isomorphism* with an attempt to explain the forces organisations or communities engage with the outside world in an attempt to gain social and economic fitness (DiMaggio & Powell, 1983).

4.3.3 Theories of institutional theory

Institutional theory comprises of three schools of thought namely, rational-choice theory, normative theory and cultural-cognitive theory. They are separated by different aspects of governance structures (Scott, 2004, pp. 408–414). However, they agree on one common thing, that society comprises of interested purposive, and often rational actors (Meyer, 2006).

The *Rational-choice theory* aims to understand how institutions supporting collective action are designed and constructed. It is founded on the atomist view which explains social behavior through individual preferences and choices. It seeks to explain why individuals rationally design systems to constrain their own behavior, and why they decide to rationally pursue their goals through institutions rather than individually (DiMaggio & Powell, 1983; Scott, 2004, pp. 408–414).

The *normative theory* aims to understand how values and commitments generated in interaction influence regimes that are regarded as formal and official. It explains how shared norms and values result in prescriptive, evaluative and obligatory rules (Scott, 2004, pp. 408–414). A norm is a rule with some binding authority over rational actors, but only inasmuch as these actors continue to support it. Some of these norms may have been created by the predecessors, and still have binding authority over the present actors regardless of their support (Meyer, 2006). The

theory of institutionalization by Philip Selzenick emerges from this theory, but focuses specifically on the process through which organisations, which are created as technical systems eventually become infused with their own values. Jerome Karabel, Charles Perrow, and Arthur Stinchcombe have helped in developing on Philip's theory (Scott, 2004, pp. 408–414).

Cultural-cognitive theory emphasises the value of shared assumptions and beliefs and how the emergent social identities reinforce social order. This new reality arises as individuals interact among each other as they create and share interpretations of what is going on at micro and macro levels. At macro level, individuals collectively create shared symbols (language) and shared understandings of their environment, which then shapes their understandings and cognitive processes, and is passed along to new entrants to the group. It holds the view that rationality if often a cultural construction. Rules vary across industrial sectors and face different pressures (coercive, normative, and mimetic) in embracing these rules. Conformance results in legitimacy, access to resources, and ultimately survival (Scott, 2004, pp. 408–414).

4.3.4 Deinstitutionalization

Deinstitutionalization involves deterioration and collapse of existing norms as a result of pressure by rational actors (Dacin et al., 2002). Pressure has resulted from man's understanding of the institutional bases of human activity, and the realization that they can rise above and control these structures, which were traditionally embedded in culture and history (Meyer, 2006). Oliver (1992) helped in identifying three types of pressure: functional, political and social.

Functional pressure emanates from perceived problems in performance level or utility of institutionalized practices. This pressure can arise from competition for resources. *Political pressure* emanates from change of interests and power distribution. These are triggered by performance crisis, environmental changes among others that lead to questioning the legitimacy of a certain practice. *Social pressure* emanates from group differentiation, heterogeneous divergent or discordant beliefs and practices, and changes in laws or social expectations that affect the life of a practice (Dacin et al., 2002). These pressures do not automatically result in the breakdown of institutional norms. Instead, they start by initiating a change process which involves organizational actors interpreting, responding, theorizing, and legitimizing new or

existing actors. Theorizing involves describing the failings of the present norms and practices, and justification of the proposed ones (Dacin et al., 2002).

4.3.5 Criticism of institutional theory

In social science, there exists a tension between theories that focus on stability and order (rational), and those that focus on choice and innovation. Institutional theory supports the first school of thought. However, the normative and cultural-cognitive theorists within institutional theory differ from this standpoint. The rationalists argue that controlling structures are formed from self-interests and are dependent on human-agency/actors for them to function. However, the normative and cultural-cognitive theorists within institutional theory differ from this standpoint. They argue that actors are embedded within ongoing systems which shape their interests and restrict their roles. This view point is reconciled by Anthony Gidden's theory of structuration, where he states that all behaviour is governed by some ongoing structure, which is continuously being reproduced and altered by participant behaviour. Structures position some actors better in proposing new rules than others, which implies that agency is socially constructed (Scott, 2004, pp. 408–414).

Some of the criticism against this theory is that it mainly explains persistence and homogeneity of institutions. However, this argument has been challenged by the view that institutions are not uniform and change in character and potency over time (Dacin et al., 2002). Another criticism by Scott (2000) suggests that most studies applying institutional theory focus on a single organization. Though this results in great insights, there is a risk in omitting other changes that occur in the process, such as: contractual versus equity relations/interconnections between organizations, boundaries between organizations, emergence of new populations, field boundaries, and governance structures (Dacin et al., 2002).

Wiseman & Baker (2006) put forward another criticism against institutional theory, that it fails to suggest what happens the moment legitimate models and forms reach equilibrium - if they ever do. One possibility is that institutional forms within a particular field will compete amongst each other for legitimacy, though this competition will not last indefinitely. Another possibility is that context-driven decoupling will challenge the validity of some legitimate models in

fundamental ways, like de-emphasizing the individual role in some or emphasizing it in others (Wiseman & Baker, 2006).

4.4 Compatibility of critical realism with institutional theory

The compatibility of a particular theory with critical realism is determined by the research approach. The approach can either be focused on substantive causes of a particular phenomenon, or on its nature and underlying mechanisms. Critical realism is compatible with a theory that focuses on the latter (Sayer 1992, 2000, Fleetwood 2004).

Institutional theory is compatible with critical realism as it does not focus on the factors that describe a phenomenon, but goes deeper into understanding the underlying mechanisms that support an institution. In particular, it is based on the assumption that institutions shape and govern social behaviour. Following this, it seeks to understand how processes and mechanisms form and manifest within institutions, whose focus transcends formal organisations to include informal interactions within or between organisations. This helps in understanding the formation or collapse of the rules and strategies that govern the behaviour of rational actors within an institution (Kraft & Scott R. Furlong, 2015; Scott, 2004, pp. 408–414).

4.5 Open data institutionalization case analysis guide

A total of six out of eight concepts identified in the literature review on institutionalization were selected in forming the open data institutionalization case analysis guide presented in table 4.1 below. A detailed explanation of why two of these concepts were excluded is provided below.

Competitive isomorphism does not apply in this context as it focuses on for-profit organisations, yet OGD is not profit driven. It is also discredited for not being able to represent the modern world of organisations today (DiMaggio & Powell, 1983). Institutional isomorphism is similar to coercive isomorphism, though it only focuses on the forces an organisation encounters while contending with the outside world, and does not address internal forces (DiMaggio & Powell, 1983). Coercive isomorphism is like an advancement of institutional isomorphism since it considers both formal and informal pressures from either within or outside the organisation (DiMaggio & Powell, 1983).

In addition, analytical questions relating to each of these concepts were formulated based on the understanding of the concept and the associated assumptions from literature. These questions act as an institutionalization lens for reviewing case studies, and also formulating the research instruments that will be used during data collection and analysis.

This theoretical guide will be applied in stage four of Pawson and Tilley's model, which expects a theory to be used in testing the candidate mechanisms and hypotheses.

Concept	**Assumption**	**Analytical Questions**	**Article**
Path dependence	- Institutionalization is influenced by history & external determinants.	- What are the historical factors and external determinants that are affecting institutionalization?	(Meyer, 2006)
Reciprocal typification. Habitualization	- Actors develop replicable problem-solving behaviors that become internalized. This includes emergent structural arrangements like policies, procedures.	- What actions are being habitualized? - Which actors are involved? - Are there other organisations that have adopted similar structures?	(Tolbert & Zucker, 1996)
Objectification	- Actors formulate shared meanings and definitions (socially constructed). - There is mutual awareness among actors. - Requires risk assessment of proposed structures and champions to create awareness. - There is increased participation among actors. - There is increased information load.	- What are the shared meanings and definitions? - How many champions exist? - Is there an increase in participation/buy-in by actors? - Are the proposed changes/structures facing resistance? - What is the level of investment for this change? - How are the new structures being documented?	(Schutz, 1967, 1971) (Zucker, 1977) (DiMaggio & Powell, 1983)
Sedimentation	- Actions acquire a quality of exteriority characterized by historical continuity of typifications. New actors treat them as 'social givens'. - Diffusion and retention of structures is dependent on actors who will either uphold or reject them. - Structures of domination and patterns of coalition emerge.	- What policies and procedures have been fully integrated into the organization workflow? - What are the actors' perceptions about current structures? - What are the actors' recommendations? - What is the link between current structures and intended outcomes?	(Tolbert & Zucker, 1996) (DiMaggio & Powell, 1983)
Coercive isomorphism	The organization experiences formal and informal	- What external pressures may have contributed to this	(DiMaggio & Powell, 1983)

	pressures (functional, political, or social) from society or other organisations. Supported by common legal framework.	change? - What problems do the current practices face and how are actors dealing with them (functional pressure)? - Are agents having divergent or discordant beliefs and practices (social pressure)? - Is there a change in social expectations (social pressure)? - Are there any political influences either for or against this change?	(Dacin et al., 2002).
Mimetic isomorphism	Organisations imitate others in cases of uncertainty on how to implement certain structures.	- How is the organization imitating others as they learn?	

Table 4.1: Open Data Institutionalization Case Analysis Guide

4.6 Conclusion

This chapter resulted in a framework that assists in understanding the institutionalization process of an open data initiative. This consists of six concepts of institutionalization, including their underlying assumptions and the kinds of questions that a researcher needs to answer in determining whether they are made manifest in a particular context or not. This framework helped in guiding this study from an institutionalization perspective in stage one, three and four of Pawson and Tilley's realist evaluation model. A detailed description of this model is provided in the following chapter.

5. Research Methodology

This chapter describes the research methods adopted for this study, which includes data collection, analysis and interpretation. Critical realism is applied in shaping the ontological and epistemological structure of these methods. More specifically, it describes the various stages of explanatory research based on critical realism, and how each of these stages were applied in this study. This study applies the case-study design, and will provide the rationale used in selecting a single case study, including sampling and access issues. A description of the preferred data sources and methods of analysis will also be provided.

5.1 Research design

The decision to conduct a qualitative study over quantitative study follows the nature of the research question, which could not be answered using observations at the empirical level or variable/statistical-causal analysis (Danermark et al., 2002, p. 175). Qualitative study is based on methods that allow for deeper understanding of a context within its natural setting. It focuses on the case study approach, which could be single or several for a given research. The case can be a person, institution, community, event or a process. This research involves a single case study focusing on the institutionalization process of the open data initiative in Kenya. The qualitative approach is defined by the following characteristics: design of the case study, study of the case(s) in the natural setting, focus is on understanding, 'thickness', and theory generation (Danermark et al., 2002, p. 158).

Case study approach has the following possible applications: provide an explanation of causal relationships in real life scenarios that cannot be studied using the experimental approach, describe an intervention in its context, describe certain topics within an evaluation, act as a meta-evaluation which is a study of an evaluation study, and finally other motives which could be aimed at broad generalizations of evidence (Yin, 2002).

This chapter provides a detailed description of the case study design. The aim of this study focuses on the first two applications of case study, which is to provide an explanation of the causal relationships and a description of the of the Kenya Open Data initiative (KODI), with the aim of understanding the causal mechanisms that affect institutionalization and sustainability of

KODI. This will be achieved through the following case study components/steps prescribed by Yin (2002), which implies that the empirical study is guided by a predefined procedure. The steps include: definition of the study's questions, formation of its propositions (if any), identification of its unit(s) of analysis, reflection of the logic linking the data to the propositions, and defining the criteria for interpreting the findings (Yin, 2002). Each of these will be described in detail in the following section.

5.2 Case study design components

This section provides a detailed description of the steps undertaken in this case study.

5.2.1 Study questions

The section clarifies the nature of the study questions with the aim of confirming that the case study approach was an appropriate method (Yin, 2002, p. 22).

The main research question for this study reads as follows; What are the causal mechanisms that enable institutionalization and sustainability of KODI? It aims at understanding how institutionalization of OGD occurs within government and the structures and mechanisms that shape this process, and the outcome that results from this process. Critical realism assists in identifying the interconnection between context, mechanism, and outcome. This study consists of three additional sub-questions which seek to describe these three aspects. The questions are as follows;

- What are the structures that affect the institutionalization of KODI?
- What are the context-mechanism-outcome configurations that emerge from KODI?

Case study approach is said to be more suitable to what and how questions (Yin, 2002). The questions raised in this research can be answered by this approach following the observation that the mechanisms in question cannot be manipulated as an experiment, and that the phenomenon in question cannot be empirically observed and needs to be studied in its natural environment (Danermark et al., 2002, p. 175).

5.2.2 Study propositions

Propositions are suggestions/opinions on where to look for relevant evidence following a research question. They are not a key requirement for case study research and some studies

prefer not to have any. There are several legitimate reasons that allow for exclusion. The main reason being that there are cases where the topic is the subject of exploration. As an alternative, it is recommended that the purpose of an exploratory study should be stated, including the criteria in which the success of the exploration will be evaluated against.

There are no propositions made in this study, since it is based on critical realism, which suggests a different methodological approach based on generalization, inference and explanation. The theme of generalization does not focus on empirical regularities but the structures that support social relations. Knowledge of these structures is achieved through a mode of inference/thought operation known as retroduction. There are other modes of inference, namely deduction, induction and abduction, which are complementary, and together form the procedure of scientific enquiry. This process involves reasoning, creativity, abstraction, and theoretical language which results in an understanding of the underlying meanings and structures in seemingly unambiguous and flat reality (Danermark et al., 2002, p. 113).

5.2.3 Unit of analysis

Defining what the case is was a major problem at the outset of case studies. A case can be an event, a programme, a process or an organizational change. The definition of the unit of analysis, and therefore the case is related to the study research questions (Yin, 2002, p. 23).

The unit of analysis for this single-case study was the Kenya Open Data Initiative (KODI). The embedded units of analysis were the various stakeholders that form part of the KODI ecosystem. These include government agencies who are the source of data, the Kenya Bureau of Statistics (KNBS) that supplies national statistics, donor agencies supporting the initiative financially and technically, civil society organizations which help in creating awareness or consumers of open data, and implementing partners such as Strathmore University who are outsourced to perform various tasks for the government agency.

5.2.4 Logic linking the data to the propositions

This involves identifying the data to be collected and the strategies that will assist in analyzing that data, with the aim of finding the best answer to the research question. This study collected

data from multiple sources of evidence within the unit of analysis, which helped in triangulating data and convergence of evidence.

Data was obtained from semi-structured interviews, document analysis, and media report review. These were grouped into primary and secondary source of data. Primary data comprised of semi-structured interviews and documents obtained from representatives at the, Kenya ICT Authority, Kenya Bureau of Statistics, government agencies and NGO's/Civil Society Organizations working on open data. Secondary data comprised of media report analysis where newspaper articles, and news feeds from television on these initiatives were reviewed. Most of the media houses publish some of the television news segments and documentaries via YouTube which makes it easy for referencing. The multiple data sources assisted in triangulating data and convergence of evidence.

5.2.5 Criteria for interpreting findings

This study was founded on the critical realist ontological and epistemological stance and followed the stages defined under the realist explanatory model for data collection and analysis. Pawson & Tilley's realist evaluation model was adopted as the preferred methodology for this study. This follows its ability to uncover the nature of programmes, how they work, for whom, in what circumstances and the change that result from them. It achieves this by analysing the configuration between context, mechanism, and outcome (CMO) (Pawson & Tilley, 2004).

Interpretation of findings, which constitutes abstraction was carried out in phase 3 of the realist evaluation model. Fereday & Muir-Cochrane's (2006) guideline to thematic analysis and code development was adopted for abstraction. It helped the researcher to identify, analyze and report on emerging themes or patterns. This guideline provides a systematic method for explaining how the themes were formulated from the raw data to uncover meanings related to institutionalization of the open data initiative (Fereday & Muir-Cochrane, 2006).

Theory testing, which constitutes phase four of the realist model was implemented with the aim of refining the understanding of CMO configurations. The theory formed in phase one was tested against the findings in phase three with the aim of validating the findings. This implies that realist evaluation starts with a theory (phase one), and ends in a theory (phase four) (Pawson &

Tilley, 2004). A detailed description of Pawson & Tilley's realist evaluation model is provided in chapter three.

5.3 Conclusion

This section helped in providing a detailed description of the preferred data collection, analysis and interpretation methods. This includes an explanation of how Pawson & Tilley's realist evaluation model was is applied in shaping the ontological and epistemological structure of these methods. It also provided the rationale used in selecting a single case study, and how this study addressed sampling and access issues. This was followed by a detailed description of the prefered data sources and methods of analysis.

The following sections will provide a detailed description of how Pawson & Tilley's realist evaluation model was applied in uncovering the underlying mechanisms that affect proper institutionalization of the KODI. Critical realism helps in defining the ontological and epistemological assumptions, which includes the formation of the interview questions, the process of analysis and interpretation of findings which concludes in phase four.

6. Phase 1 - Hypothesis Generation

This phase involved formation of hypothesis, and is characterized by candidate CMO configurations/propositions. To achieve this, a literature review on the OGD phenomenon was conducted, resulting in valuable insights that helped in developing propositions. Though this phase allows the researcher to conduct initial/preliminary interviews with programme architects or practitioners, the researcher preferred to conduct the interviews during the data collection phase. The resulting hypothesis/propositions helped in understanding what works for whom and in what circumstances/conditions for the cases that were reviewed, and would act as a lens that would help in identifying the structures and mechanisms affecting KODI institutionalization (Pawson & Tilley, 2004).

The following sections provide a detailed literature review, which is guided by institutionalization theory and critical realism. The findings are then used to develop CMO propositions. These propositions will serve as input for the data collection phase, as they guide in developing data collection instruments, and act as a lens for the analysis phase.

6.1 Case Study Institutionalization Analysis

The aim of this section is to identify the structures and mechanisms that affect the institutionalization of OGD in different countries through a review of nine case studies provided in Table 6.1 below. These case studies were published between the year 2010 and 2014, and represent seven different countries including the United Kingdom (2014), Austria (2014), Brazil (2012), India (2013), Qatar (2014), Netherlands (2012), U.S.A (2010), and Brazil (2014). One of these cases is a comparative study between U.S.A, UK and Netherlands that was conducted in 2013. This literature review was conducted in 2014, which explains why the review ends with 2014 publications. In addition, only Kenya and Morocco had embraced OGD and were in the process of implementing it. It would have been useful to include the case of Morocco, though there were no academic publications on Morocco at the time.

This review is guided by Table 4.1 institutionalization analysis guide, which acts as a lens in identifying the concepts and assumptions that affect institutionalization in a particular case study. In addition, this review will assist in identifying the stakeholders that drive the initiative, which

acts as a guide on who to reach out to in the Kenyan case. Following this, a summary of the identified structures and mechanisms will be provided in the next section. These will act as a guide for formulating the data collection instruments in the next phase.

Open Government Data: **Facilitating and Motivating Factors for Coping with Potential Barriers in the Brazilian Context** (Albano & Reinhard, 2014) *Brazil* Case Study Interpretive approach 18 semi-structured interviews and 12 unstructured interviews. Document review of the law dealing with public information access.	
Case Description	**Application to Institutional Theory**
This paper studied the perceptions of both producers (government agents) and users of OGD in Brazil with the aim of understanding how to improve demand and supply of OGD. On the demand side, the study sought to understand the possibility of a network of non-government agents, including the structures (legal, technical, etc) that would be required to support them. Brazil became a member of the Open Government Partnership (OGP) in 2011. Following this commitment, a new law was enacted in May 2012 which helped in regulating the constitutional right of citizens to access government data at all levels. According to the nationwide survey in 2010 on e-Government, these regulations face numerous implementation challenges especially at the municipal level. The Office of the Comptroller General (CGU) is charged with ensuring compliance to these regulations, and has conducted several awareness initiatives including the 2012 conference on open data (Albano & Reinhard, 2014). This was an interpretive study based on data from eighteen semi-structured interviews and twelve unstructured interviews. The interviewees consisted of government agents including municipalities, OGD software developers, academics, journalists, non-governmental organisations, civil society organisations, and a private for-profit company that makes use of open data. The interviews were aimed at understanding the respondent OGD activities, enablers, motivators, challenges and barriers in the use or creation of OGD. Additional data was sourced from the Brazilian legislation on OGD (Albano & Reinhard, 2014). Observed benefits and advantages for government include 1) political and social factors, where OGD is viewed as a channel towards transparency and public engagement/participation, 2) operational and technical aspects, where OGD is perceived as an enabler to optimization of internal processes which facilitates collaboration with various stakeholders, including other government agencies, 3) economic benefits were not emphasized much, though it was mentioned that creating a more collaborative environment, and providing incentives for innovation would eventually lead to economic gains (Albano & Reinhard, 2014). Observed benefits and advantages for non-government stakeholders, also referred to as users of OGD include 1) political and social factors with a focus on transparency and citizenship, 2) operational and technical factors, which includes provision of products and services geared towards greater integration between society and government (Albano & Reinhard, 2014).	There is an aspect of path dependence, since the existing culture seems to affect the implementation process (Meyer, 2006). This could explain the resistance and slow adoption at the municipal level. Unfortunately, the historical factors and external determinants leading to this are not explicitly stated (Albano & Reinhard, 2014). There is also an aspect of objectification. This is supported by the presence of a champion and the investment made towards the initiative (DiMaggio & Powell, 1983; Schutz, 1967, 1971; Zucker, 1977). The office of the comptroller general acted as the champion in this case, though there was resistance at the municipal level, which resulted in several challenges. The CGU also made significant investment on the initiative by funding several awareness initiatives including a conference in 2012 (Albano & Reinhard, 2014).

Observed inhibitors and barriers affecting uptake of OGD include 1) technical issues such as format and quality of data. Journalists, non-governmental organisations, scholars, and software developers can put pressure on government to improve on the quality of data and OGD related services 2) institutional factors including legacy structures and political issues, 3) legislation especially with regard to what constitutes private data/privacy, 3) resistance by government agents which is characterized by low interest in cooperating with these initiatives. Journalists, non-governmental organisations, and even scholars and software developers in the OGD agenda are government watchdogs. They can put pressure on government to enforce or formulate suitable legislation and improve on the quality of data and OGD related services. They can also create awareness among citizens who will help in creating more pressure on government using facts from OGD to improve service delivery (Albano & Reinhard, 2014). Observed facilitating and motivating factors per stakeholder include 1) Journalists: to generate greater OGD demand through information dissemination, building OGD capacity among peers, and create awareness among society on OGD, 2) public agencies and agents: to make data available by identifying interests of other government agents/agencies, and society. Also, by implementing existing regulations on public data provisioning, 3) software developers: to make practical applications that create and render information from OGD, 4) scholars: to promote interest in the subject, 5) private sector: to develop new products and services for commercial gain, these services could be intermediary solutions between government and citizens 6) Non-governmental organisations: to increase citizen participation in governance, and support government by providing technical tools and expertise (Albano & Reinhard, 2014).	
Information Strategies for Open Government: **Challenges and Prospects for Deriving Public Value from Government Transparency** (Dawes & Helbig, 2010) *U.S.A* Case study 35 interviews were conducted in five diverse counties in New York State. Official publications and websites were also reviewed.	
Case Description	**Application to Institutional Theory**
This case study was conducted in New York State between September 2004 and February 2005. It focuses on the value and use of land records and parcel data for various public and private purposes. It investigates the challenges in deriving social and economic value from OGD. Parcel data is a requirement by the property ownership laws and tax administration, and is collected at the municipality whenever property is sold or sub-divided. The municipality reports a part of this parcel data to the county government, who also report a subset of the data they receive to state (Dawes & Helbig, 2010). Thirty five interviews were conducted in five diverse counties in New York State. Official publications and websites by federal, state, local, nonprofit, and private sector organizations were also reviewed. The interviews were aimed at understanding the logical and purposive flow of data from the data collector to the user, who could be an organisation or individual. This also included understanding the costs involved, value of the data to the organisation, encountered issues and barriers (Dawes & Helbig, 2010). It was observed that there was no common definition of parcel data, and that users apply varying definitions. For instance, the planning department	The proposed strategies, if implemented would affect OGD policies, procedures, and outcomes. It also identifies some of the historical factors, external determinants and pressures that affected some countries on OGD (Dawes & Helbig, 2010). This applies more to the concept of coercive isomorphism (Dacin et al., 2002; DiMaggio & Powell, 1983).

and the real property tax administration capture different fields. The planning department captures parcel identification number, zoning code, actual uses, boundaries and physical characteristics. The real property tax administration captures parcel identification and location, structures within the parcel, parcel owner, and assessed value. The only common field is the parcel identification number (Dawes & Helbig, 2010). The various users, who comprise of individuals, public, private, and non-profit organisations make use of this data in varying ways. The following are some of the major uses/value of this data; 1) real property assessment and taxation by municipal assessors and county real property offices, and the state real property agency, 2) buying and selling of private land, 3) directing emergency response with regard to emergency routing, how to enter a property safely and who to notify, 4) transportation routing for instance creating bus routes or priority routes for snowplows in winter, 5) facilities siting by municipalities to plan for growth, or by companies wishing to relocate, 5) planning and prioritizing environmental initiatives, and 5) infrastructure management (Dawes & Helbig, 2010). Even though these users derive different value from parcel data, their interests coincide when it comes to accuracy, timeliness, and consistency of the data. This also includes ease of access in a variety of electronic data formats, and from one authoritative or trusted source. Most of them reported the need for better quality data including metadata as they spend considerable resources to verify, correct or integrate the data prior to use. Also, different countries use different technology to render their data, which adds to the complexity when one wants to compare or merge datasets (Dawes & Helbig, 2010). Data management is also a key challenge, since only a few municipalities have data management strategies that cater for the needs of external users, and in most places requests are handled in an ad-hoc manner. There are also no feedback mechanisms between data users and collectors, which would help in improving the quality of data. The data flow is in one direction and thus the effort made by users to improve the data does not feed back to the original sources. This becomes more complex when users expect periodic updates as they would have to match the older datasets and the new without overwriting the datasets they had improved on (Dawes & Helbig, 2010). These challenges call for improved protocols or policies on data management, data and technology standards, feedback mechanisms, and skills for government and other OGD stakeholders (Dawes & Helbig, 2010).	
Benefits, Adoption Barriers and Myths of Open Data and Open Government (M. Janssen et al., 2012) *Netherlands* Systems theory and Institutional theory. (myths & legitimacy). Nine interviews & a workshop were conducted to collect data.	
Case Description	**Application to Institutional Theory**
It correlates the publication of OGD with use and benefits. Affirms the proposition that open data helps in reinforcing existing structures instead of bringing change. Examples include the government publishing safe data only and hiding contentious data, budget cuts. Identifies the assumptions that exist when expecting government agencies to open up their data. Highlights the importance of feedback mechanisms in open systems, where government can interact with the citizenry. Not necessary for closed systems (M. Janssen et al., 2012). Benefits, barriers and myths of open data systems are identified through experiences shared by users of open data through interviews and	Though this is not a case study, it helps to highlight some of the aspects that affect the institutionalization process, including path dependence and coercive isomorphism. Path dependence is characterized by pre-existing culture which affects adoption of OGD. Some of the challenges listed include

workshops. Benefits are classified into three categories 1) political and social, 2) economic, and 3) operational and technical. Barriers are seen to affect either providers or users of data. Providers may resist following institutional level concerns including loss of revenue from private data. Users are affected by data complexity. Other barriers affecting both users and providers include legislation, data quality, and technology (M. Janssen et al., 2012). Findings suggest that government needs accept that traditional planning and control instruments will no longer work in the open system. This further suggests that the mechanistic system will be replaced by an evolutionary system which is self organising. Access to data is not sufficient, there is need to improve data quality, cultivate a culture of openness in government, and provision of resources for data manipulation and translation. There is also need for institutional measures for public participation (M. Janssen et al., 2012).	resistance to publish data for free as this would affect existing revenue streams, and also data complexity which may require additional skills, new ways of working and change of responsibilities for some. This may result in resistance if the terms of employment are not improved in relation to this (M. Janssen et al., 2012). Coercive isomorphism is characterized by the speculated evolutionary system that may disrupt the traditional planning and control instruments. This will result in change of culture, where government will be more open and inclusive, allowing for public participation. However, this process may be slowed down by legislation that is either missing or contradicts with the initiative, data quality which may require more resources and time to process, and availability of suitable technology and skills in the various government agencies to implement OGD (M. Janssen et al., 2012).
Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives (Ubaldi, 2013) *U.S.A, UK, Netherlands* Case study	
Case Description	**Application to Institutional Theory**
Identifies the preconditions necessary for an efficient and effective OGD implementation. This includes required principles, concepts and criteria. Beneficiaries of OGD are identified based on their role in the value chain for public value creation; 1) public sector agencies to improve service delivery, and increase transparency and accountability. New skills are required including data science and predictive analytics, 2) private sector to offer innovative value-added services, 3) academic institutions 4) Citizens to fund OGD initiatives (through taxes) and improve e-participation with government from more informed perspectives, and improved quality of life through smart disclosure of data such as that of flights by national airlines, and 5) civil society organisations (CSOs) to increase transparency and service delivery, including playing the intermediary role of helping to identify key datasets and lobbying for their publication (Ubaldi, 2013). In enabling fair competition with the private sector entities, government agencies need to use the same government data sources that the public access. The citizens, CSOs and private sector are also producing data through crowdsourcing and using it to make consultations with each other. Government needs to play the facilitator role (collaborator) to help create more real time data and share information. There is also need to create structures that enable open data (crowd or government sourced) to be	Coercive isomorphism may emerge as a result of advocacy and awareness campaigns by the CSOs, which is external pressure for the government resulting from a change in social expectations. These campaigns could also be politicized either for or against the OGD initiative, by either a push for reforms or a drive by the government to maintain the status quo. Path dependence may also emerge following a cultural shift on government service delivery. Citizens may also change their culture by adopting skills that enable them to interpret and consume OGD. Objectification may emerge from

available and accessible, and authorized for reuse and redistribution (Ubaldi, 2013). Impact of OGD can be measured by the level of ubiquitous engagement and participation. This results in new features, businesses, markets, competencies and tools (Ubaldi, 2013). Freedom of information (FOI) legislation has been developed and adopted by many member countries of the Organisation for Economic Co-operation and Development (OECD). FOI defines the requirements for accessing raw, un-manipulated datasets. This act calls for amendment of inconsistent laws such as the Official Secrets Act (Ubaldi, 2013). To avoid widening the digital divide, government and CSOs need to raise awareness of the privilege provided by this act among citizens to create a new culture that can interpret and consume OGD. Public authorities also need a cultural shift on service delivery, by understanding that OGD does not disempower or make them more vulnerable as some perceive. Also, that they are not the exclusive owners of data (Ubaldi, 2013). OGD exists within an ecosystem, which comprises of three sub-ecosystems 1) data producers ecosystem comprising actors such as public sector, academia, media and private sector 2) infomediaries ecosystem comprising actors such as media, developers, statisticians and CSOs, and 3) users ecosystem. A good example of a thriving ecosystem is the City of San Francisco, where there is a strong sense of community by citizens and activists (CSOs) (Ubaldi, 2013). In 2012, the United Kingdom national government published their first Open Data white paper, and government departments were required to use this as a baseline in publishing their own open data strategies. The main objective was to improve data accessibility and increase transparency. This counteracts disclosure policies that may have limited access and transparency (Ubaldi, 2013). Central/federal government needs to impose policies that are aimed at consistency and co-ordination, especially in an effort to try and harmonise from different government agencies. However, this is challenging if the process is not carried out collaboratively and the costs are bound to increase following poor data management practices. If successful, data would be more interoperable among different government agencies following standardization (Ubaldi, 2013). Commitment by government and its agencies is measured by the level of investment in acquiring needed skills and resources for collecting and curating data. This may require a change in the existing operational models, and additional budget to cater for the costs of curating data (Ubaldi, 2013).	standardization as various government agencies adopt shared meanings and processes in relation to OGD from a supply and even a demand perspective in the case where there is dependency between agencies. Legislation such as FOI and the Official Secrets Act may result in objectification and coercive isomorphism. These are structures that assist in ensuring sustainability and protect the right of the citizen. In the case where government fails to implement these laws, citizens with the help of CSOs may create pressure to government to enforce them.
Improving the transparency, openness and efficiency of e-government in Qatar in the era of open government data, and beyond (Al-kubaisi, 2014) *Qatar* Case study	
Case Description	**Application to Institutional Theory**
In 2000, the Qatar government started an e-Government initiative aimed at paperless government by converting manual services into electronic transactions. This started as a pilot in phase one, and following its success, full government support was granted to roll it out to all government ministries and public sector organisations. This involved the Ministry of Interior as service providers, Qatar National Bank as payment gateway, and Qatar Central Bank as the e-Government information systems host (Al-kubaisi, 2014).	Coercive isomorphism is evident following the decision by government to adopt OGD following instructions from the Gulf Cooperation Council (GCC) in an effort to remain competitive. Unfortunately, despite this decision, OGD is regarded by

In 2004, the Qatar Government established ictQATAR, which was charged with developing and managing the implementation of Qatar's ICT strategy. One of the outcomes was the e-Government site *Huukomi* portal.www.gov.qa, which provides detailed government information including laws, and also offers public services such as student school enrollment, traffic fine payment, visa application, tender information, and health card and licenses renewal (Al-kubaisi, 2014). Despite this array of services on the e-Government site, public uptake and participation remains low. This is contributed by 1) lack of public awareness, 2) lack of public trust since most of the content published is biased and leaves out contentious issues, 3) lack of understanding by government of current needs and perception of citizens, which would help in designing a more citizen-centric e-Government system (Al-kubaisi, 2014). In 2014, the government launched its OGD initiative following the approval of the Open Data Policy in an effort to address these shortfalls. The main objective was to synchronise information from the various government agencies. To achieve this multiple stakeholders were involved including; General Secretariat for Development Planning, the Statistics Authority, the Supreme Council of Health, the Supreme Education Council, the Supreme Council of Family Affairs, ictQATAR, Ministerial Cabinet and the Permanent Population Committee (Al-kubaisi, 2014). Qatar's Open Data Portal is maintained by the statistics authority and ictQATAR only acts as a facilitator. All data is contained within the portal without links to other government agencies datasets. Data is available in excel and pdf format, with little if any metadata. The portal lacks a participation mechanism, and the discussion forums and blogs available on the e-Government portal seem to have no effect, at least as of January 2014. It also lacks a collaboration mechanism, which implies that there is no channel to request additional datasets. Based on Tim-Berners Lee's open data rating system, Qatar scores two stars as data is available on the web, but only accessible through proprietary software, excel and pdf. Data provided in pdf is not readily reusable and poses a challenge to potential users. Economic data and fiscal information is also not available (Al-kubaisi, 2014). The key driving forces that led to the government signing up to the open data initiative include; 1) it is a global movement led by the OGP that is being emulated by member states including Qatar, 2) pressure from members of the Gulf Cooperation Council (GCC) to embrace OGD with the aim of attracting foreign investment, and also remain competitive among the GCC member states (Al-kubaisi, 2014). Approach used to initiate and manage OGD initiative. This initiative was led by ictQATAR, and involved formulation of the Open Data Policy in consultation with owners and stakeholders. The Statistics Authority, who own and manage the open data portal (Qalm), were granted the mandate to request for information from any government agency (Al-kubaisi, 2014). The challenges identified include; 1) poor understanding of the Open Data concept, and many government agencies still perceive it as a threat, 2) there is lack of a clear vision regarding ownership of the project and responsibility of the open data portal. Potential owners include Qatar Statistics Authority, Council of Ministries, or the Ministry of Information and Communication Technology. This should be resolved within the e-Government transformation strategy, 3) Lack of a legal framework addressing data protection, management and access. Qatar is yet formulate/adapt laws that support OGD, which includes the FOI act. 4) Poor co-operation between government agencies, which is aggravated by	government officials as a threat to their operations and lacks any benefit. This perception creates negative political pressure against the initiative. Related to coercive isomorphism, is the concept of sedimentation. The current government perceives OGD as unnecessary structure, and they fail to see the benefit they may accrue from it. This implies that the open data policy, though formulated, may not be enforced and institutionalized and the effort to synchronize information across government agencies may not bear much fruit. This calls for awareness and sensitization to help ensure objectification, as discussed in the next paragraph, which requires the various actors to have shared meanings, resulting in a common vision. Objectification is manifested through the open data policy, which seeks to create standards that will help in synchronizing information across government agencies. Several champions from eight different government agencies spearheaded this initiative. However, the vision to drive open data portal initiative based on the open data policy is lacking despite involvement by the eight agencies. This has resulted in resistance. The open data concept is not clear for many, which implies lack of shared meaning or understanding of the open data concept, which could explain the lack of vision. In addition to this, the variation in readiness towards open data among agencies, lack necessary policy and legislation, cultural barriers, and uncertainty about the value of OGD form additional barriers to objectification.

different levels of readiness, lack of necessary policy and legislation, cultural barriers, and uncertainty about the value of OGD. This calls for awareness programmes aimed at promoting OGD practices and nurturing a culture of openness. These challenges could be resolved through an OGD strategy that includes all stakeholders namely private sector entities, entrepreneurs, academics, Non-Government Organisations (NGOs), and the general public (Al-kubaisi, 2014). Perception of OGD among government agencies is that it lacks any benefit and is a threat to their operations. This calls for continuous education and awareness for all stakeholders including public and private sectors. Qatar is not yet a member of OGP (Al-kubaisi, 2014).	
Towards an Expanded and Integrated Open Government Data Agenda for India (Chattapadhyay, 2013) *India* Desktop research	
Case Description	**Application to Institutional Theory**
In March 2012, India approved the National Data Sharing and Accessibility Policy (NDSAP). It aligns itself with principle ten of the United Nations Declaration on Environment and Development on international citizens' rights declaration and section 4.2 of the 2005 Right to Information Act (Chattapadhyay, 2013). Government solicited public feedback during its formation. However, as is the case for comments submitted by Bangalore-based Centre for Internet and Society, this feedback was not incorporated in the final policy document (Chattapadhyay, 2013). All government agencies are required to adhere to NDSAP by making public all sharable and non-sensitive data. Certain acts such as the unit level census data prohibit categorise some data to be sensitive and therefore non-sharable. The NDSAP does not require for data to be shared under open license which limits re-use and re-distribution. The government has a central repository for this data data.gov.in, which is managed by the National Informatics Centre (NIC), which falls under the Ministry of Communication and Information Technology. The portal was developed with support from Office of Citizen Services and Innovative Technologies, General Services Administration, and the United State of America government (Chattapadhyay, 2013). The scope of NDSAP is limited to the central government, which creates the need for a policy that focuses on the twenty-nine states. There is also insufficient interoperability among government agencies, which results in lack of standardization of operations and data outputs. This results in cost and effort multiplication following redundancies and creates a challenge for sustainability. The right to information act supports NDSAP by emphasizing the need for proactive disclosure, and guidelines on how to operationalize and comply to proactive disclosure this in India was formulated through a memo to all government agencies. This requires them to list all datasets in their control, and indicate which ones are public or not. However, workflows for informational practices were yet to be formulated (Chattapadhyay, 2013). The NIC also developed the guideline document to guide in the curation process. This document was formed consultatively together with the government officers' in-charge of data curation and publication from each central government agency. It defines the standards including metadata requirements for each dataset and how to adhere to them while curating data (Chattapadhyay, 2013).	Coercive isomorphism is evident as India's decision to approve NDSAP is partly because of the need to comply with principle ten of the UN Declaration. A shift in social expectations could be supported by the fact that citizens provided their comments regarding the NDSAP. Objectification is also emerging following the formation of NDSAP, and the identification of NIC as the managing organ for the open data initiative. The awareness campaigns seek to create shared meaning across the various agencies, which results in increased participation and ownership of the OGD concept by the various agencies. Part of the proposed changes to the structure, is to extend NDSAP scope to not only focus on central government but also extend to the twenty-nine states in India. Regarding the level of investment, the NIC was funded to support this initiative and have been running awareness initiatives. There is some level of sedimentation through NDSAP, though it's scope is limited to central government. The right to information act was also enforced through a memo to all government agencies. However, informational practice workflows to implement this act at the agencies are yet to be

NIC also conducts awareness programmes targeting government agencies with the aim of creating a culture of proactive sharing, and also organising community outreach programmes to increase demand and use of data by the public (Chattapadhyay, 2013).	formulated.A procedure supporting sedimentation is the data curation guideline document developed by NIC.
Open Government Data in Brazil (Breitman, Salas, Saraiva, Gama, & Casanova, 2012) *Brazil* Case study	
Case Description	**Application to Institutional Theory**
Since September 2011, Brazil has been a member of the Open Government Partnership, and is committed to transparency and open data. This initiative was supported by the president which resulted in the launch of the national OGD portal. Prior to the OGD movement, there was an effort in 2009 by the Information Organizing Committee of the Presidency (COI) to create a reliable source of official data for the President and his advisors by aggregating all government data in digital format into a central information catalog. In 2010, following the success of the initiative, the catalog was made public and renamed to DadosGov, which is accessible via dados.gov.br. DadosGov hosts more than 1,300 historic datasets which date back to 2002 (Breitman et al., 2012). Following recommendation by COI, the data was classified into territorial (country, states, cities), temporal (year or month), and thematic subjects such as infrastructure, citizenship and social inclusion. Government agencies were required to submit their data in spreadsheet format, which was then processed and stored in relational databases and made available via open formats specifically XML, JSON, and Resource Description Foundation (RDF). The RDF format provides interoperability with the Linked Open Data (LOD) cloud. RDF comprises of extensible standardised vocabularies which define concepts and relationships within data, allowing for the creation of mashups and graphical visualisations. For instance, it was possible to compare Brazil and the United States of America using RDF datasets by comparing vocabularies defined in data.gov with similar concepts defined in DadosGov (Breitman et al., 2012). The W3C supported this initiative by providing sponsorship to ICT professionals in the public sector to receive training on OGD technologies. They also created an enabling environment for government officials to interact with the research community. This engagement resulted in the formation of a task group consisting of representatives from all sectors in the federal government to develop a strategy that would support the OGD adoption process (Breitman et al., 2012).	Coercive isomorphism emerges from the influence of OGP on the decision by Brazil to make a commitment towards transparency and openness. However, this was not the initial cause, as the COI had undertaken a similar initiative three years prior. This had political support from the presidency was critical in achieving success. There was also additional support from a skills perspective from the W3C, which increased the readiness of government to implement OGD. Objectification emerges from the directive by COI to government agencies to follow particular standards while publishing data. This implies shared meaning and definitions on what consists of open data, through the stated standards. COI is the main champion in this initiative. There is no mention of resistance by the government agencies to abide by the COI directive. Sedimentation, like objectification also emerges from the directive by COI, which provides the procedures to follow while curating and publishing data. The actors perceptions and recommendations about this procedure are not provided.
Open Government Data Implementation Evaluation (Parycek et al., 2014) *Austria*	
Case Description	**Application to Institutional Theory**
Austria's City of Vienna sought to evaluate its OGD strategy implementation from a return on investment perspective. It is very decentralized in terms of decision making, and though the city directorate may issue regulations, it remains the prerogative of the city department heads to support and implement. This is because they retain control over what data to publish to the public and in what granularity (Parycek et al.,	Path dependence emerges from the decentralized historical structure where city government heads retain the mandate to support and implement initiatives in their constituency, including open data.

2014). Austria's OGD project had political commitment (goodwill) resulting from an already successful tradition of federal co-operation in e-Government. Disclosure of government data was also supported by several existing laws. Though they don't have a proactive Freedom of Information (FOI) Act, there is a law on openness that requires government officials to provide data requested by the public, on condition that it does not adversely affect their regular work. However, Austria has an obligation of secrecy within the constitution, which undermines the effectiveness of the law on openness. There is need to move beyond political goodwill and formulate a legal framework that ensures that OGD is sustainable, following a legal guarantee (Parycek et al., 2014). Also, a group of people emerged from government, academia and civil society (Open3.at) with a common interest in OGD to form Cooperation OGD Austria. This gave birth to the competence centre of Open Government of Vienna, which consisted of several departments including the city ICT, public relations, GIS, economy and statistics, e-Government, city planning, and data protection departments. This is a virtual and non-formal entity that reports directly to the Chief Information Officer of Vienna, who is charged with implementing the OGD strategy. It became effective in providing guidance on various issues pertaining setting the OGD agenda, including data standards, licensing and identification of resources. Departments which were not active in this cooperation were found to have less knowledge on how to implement the Open Government strategy (Parycek et al., 2014). Austria's OGD portal data.wien.gv.at was launched in 2013 with a total of 1047 datasets from 21 different organisations. Departments directly upload their data on the portal and flag it as public or otherwise. However, marking data as public does not result in automatic publication as it needs to also comply with the guidelines of OGD such as the requirement to enrich the data with metadata. It was observed that this process is clearly defined and responsibilities clearly outlined, that they OGD portal was integrated with existing CMS systems, and that evaluation is a key part of the OGD process (Parycek et al., 2014). It was observed that the release of OGD did not affect the number of public requests by email or telephone. Also, staff highlighted that OGD created more work leading to an increase in labour costs and loss of revenue from data that was previously released at a fee. However, they maintained that data should be provided free of charge since its benefit is hard to measure. However, despite the challenge in measuring the tangible benefits of OGD, this knowledge was still deemed necessary as it would further persuade departmental heads to increase their efforts in releasing data to the public (Parycek et al., 2014). Some of the benefits they identified include 1) easier access following a more customer focused approach that stimulates and improves the value and efficiency of services 2) increased transparency where citizens are viewed as partners, and can assist in improving the process such as data quality by reporting errors they find in the data 3) reduction in administration costs, and 4) simplification of administrative procedures following the removal of some processes such as invoice generation and processing or user specific terms of service, 5) Some viewed OGD as a factor for innovation leading to new enterprises and business models, and one interviewee mentioned that they have increased a 5% increase in sales. However, no novel business models emerged from the interviews (Parycek et al., 2014).	The city has also historically successfully co-operated with the federal state on e-Government. Despite the lack of FOI act, there is a law that requires government officials to provide data upon request by the public in cases where this does not affect their regular operations. There is also a law on secrecy which at times stands in the way of openness. Sedimentation is also evident from the responsiveness of the city to provide data to the public, following the institutionalization of the law on openness, which has been incorporated within the city's departmental workflows. This also includes procedures on how to publish data including requirements for data to contain metadata. The staff at the city departments perceive OGD to having created additional work, increased labour cost and led to the loss of revenue from data that was previously provided at a fee. They also find it difficult to measure the value of OGD. However, despite these shortfalls, there was no resistance reported from the departments. Objectification also emerges especially from the shared understanding across departments on what constitutes open data, the requirements and the procedure to publish data to the public. The departmental heads are viewed to be the champions since decision making is decentralized and they decide what gets done. There has been a request to measure the value/benefits of OGD.

There were also unintended implications including 1) some departments used the OGD platform as a replacement for a missing enterprise information system, 2) departments used this platform to inspect/learn about other departments, or use the data from other departments for their own activities, 3) negative voices emerged with concerns about erroneous data and possible legal liability consequences, 4) more transparent processes will result in greater interest in politics as citizens can now actively contribute to the acceptance, stabilization and legitimation of policy decisions, 5) there was a call to release federal data such as commercial and land registers, though this falls outside the jurisdiction of the City of Vienna. This represents a demand for data, 6) lack of financial incentives explains lack of participation on OGD from some target groups including application developers and the research community, 7) city employees are motivated and value their work more as they see their output/data being used in various applications (Parycek et al., 2014).	
The strategic importance of information policy for the contemporary neoliberal state: The case of Open Government Data in the United Kingdom (Bates, 2014) *United Kingdom* Thematic analysis of 11 interviews, observational and policy documentation.	
Case Description	**Application of Institutional Theory**
This paper analyses the link between OGD and neoliberalism from a policy perspective in the UK government. It focuses on the development of the OGD information policy between 2010 and 2013, which is a period that was characterized with deepening crises and a drive by government to propel the neoliberal agenda forward. The findings suggest that the center-right coalition government used the OGD information policy as an instrument to support the cause of several controversial policies aimed at advancing the neoliberal agenda (Bates, 2014). Research methods involved 21 interviews. 11 of them involved OGD advocates, four of which who previously advised the UK government on open data or public sector information. The other 10 involved OGD UK policy makers from civil service, local government and private sector. This data was complimented by a series of observations of 13 OD events, content analysis of the okfn.org mailing list, and analysis of UK policy documentation published between 2009 and 2012 (Bates, 2014). This resulted in a theoretical framework that demonstrates the intersections between OGD policy and the major public policy initiatives of government between 2010 and 2013. These initiatives consisted of the transparency agenda, the open public services agenda, privatization, and economic growth strategy (Bates, 2014). UK public agencies produce several types of open data including mapping, meteorological, land use, public transport, company registration, government expenditure, and geo-location of local council services (Bates, 2014). In 2009, Sir Tim Berners Lee and Prof. Nigel Shadbolt, both of whom are successful technology innovators and passionate for openness/open data were appointed as information advisors to the UK government. This was triggered by the financial crisis and the parliamentary expenses scandal. Their first deliverable in January 2010 was the UK open data portal - data.gov.uk. Many civil society advocates view open data as a tool to fight neoliberalism and promote greater democracy (Bates, 2014).	Coercive isomorphism is evidenced by the drive by the center-right coalition government to support the OGD information policy as a means to another end. The civil society also used this as a tool to fight neoliberalism and promote greater democracy. This resulted in external pressure towards government. Government acted in consultation with a team of information advisors. Path dependence is evidenced by several historical factors including the financial crisis which was aggravated by expenditure scandal in the UK parliament, leading to a loss of trust by the electorate. The perception by some policy makers and the civil society that OGD can help resolve the crisis and rebuild trust assisted the institutionalization of OGD in the UK. Sedimentation is mainly evident through actors' perceptions about the current structures. Despite efforts to publish some datasets, some are of the view that this only creates new opportunities for the neo-liberalists since the commercial agencies can also

Open Government Data and the Reproduction of the Neoliberal State a) Transparency and trust formation in a neoliberal state In the UK, OGD was introduced at a time when the UK was facing economic crisis that was aggravated by the parliamentary expenditure scandal, leading a further drop in trust levels by the electorate. This resulted in a crisis for legitimacy, and explains why they supported the OGD agenda. Some policy makers publicly stated their concern for legitimacy and how OGD could address this crisis. Some OGD advocates viewed OGD as a response to social, economic and political objectives. However, others viewed it as a tool to create further breakdown of the political establishment, which reveals levels of distrust. Some of this was fueled by decisions such as UK invasion of Iraq in 2003, parliamentary expenses scandal, phone hacking scandal between 2011 and 2012. They also cited unethical relationship between media, politicians and the metropolitan police force. A disconnect emerged between the objectives of the OGD advocates and the state-based advocates, where the former aimed at empowering citizens beyond neoliberal confines, while the latter aimed at rebuilding trust within the confines of neoliberalism (Bates, 2014). b) Public services data and the open public services agenda. The transparency initiative in the UK is being developed in an environment characterised by public spending cuts and neoliberal marketisation. In 2011, the UK released several datasets including national health service data (Bates, 2014). c) Trading fund data and the provision for standardization. Privatisation is part of UK government's agenda. The UK government retains the position that trust funds should be charged (Bates, 2014). d) Financial market expansion: weather data and the exploitation of risk Noting that the UK government does not intend to open up all the trading data, there is need to identify the strategic importance of some datasets, and how these benefit certain political and economic actors. In 2011 high quality weather data and information was released. The beneficiaries include commercial agencies which result in the continuation of the neo-liberal state. This implies that some are after generating private wealth, which reduces the incentive for those intending to profit through products/solutions that are aimed at mitigating climate change (Bates, 2014). e) Information policy and the neoliberal state. Despite the growing interest in OGD, there is little literature/information that helps explain the technicalities and interrelations with other policy areas. OGD is a shift from an era of proprietization and commercialization to one of co-productive relationship between the citizens and the state. However, parties to neoliberalism are still finding ways to use OGD to continue their political and economic agenda. This implies that the information policy is being used as a tool to resolve the crisis of the neoliberal state (Bates, 2014).	access and derive value from OGD. One of the main recommendations is to identify the strategic importance of some datasets, and how these benefit certain political and economic actors, which could then assist in building a case to publish data that is currently considered confidential, as is the case for trading data.

Table 6.1: OGD Case Studies

Table 6.1 above provides a sample of the existing research on OGD in both emerging and developed countries. This includes a list of methods used for these studies. The aspects discussed by these articles include heterogeneous data management, data semantics aimed at creating new meanings, techniques for effective information use, accessibility challenges, OGD implementation, OGD policy, OGD initiative analysis/evaluation and finally, the role of culture

on transparency. In addition to this, two models were developed, including a model to study the complex dynamics of government transparency by Meijer (2012), and an analytical framework for an empirical analysis of OGD initiatives by Ubaldi (2013).

OGD research is not conclusive, which presents the need to study other aspects such as institutionalization and sustainability of OGD in the African context. The institutionalization analysis guide provided in table 4.1 helped in analyzing and identifying the institutionalization concepts that emerge from each of these case studies. The most common concepts are path dependence, objectification, sedimentation and coercive isomorphism. A similar process will be carried out once the KODI case description is complete. These concepts will help in identifying the causal structures and mechanisms that affect KODI.

6.2 Understanding Information Systems CMOs from Literature

The main aim of this section was to create an understanding of the kinds of mechanisms to expect in information systems, for which open data initiatives form part following their dependency on technology. This follows the observation that not much information sytems research has been conducted using critical realism, and as a result, it was found necessary to observe the mechanisms that have emerged from the existing studies. This would be a useful learning point since the context may be similar, from the point of view that OGD involves a technological component, and operates in similar a social context. Preference was given to research which is closely related to the OGD phenomenon.

This exercise helped in identifying the CMO configurations for each case, which included identification of the causal powers including conditions and pressures, real objects, and linked mechanisms, which refer to the relationship or interdependency between the mechanisms. This information is not made explicit in most cases, and was therefore a necessary exercise to help in creating an understanding of how to conduct similar research. Four cases were identified for this exercise and are presented in a tabular format in the following sections.

6.2.1 Generative Mechanisms of OGD

Jetzek et al. (2013) developed a critical realism based framework that seeks to explain how OGD generates value. In particular, the framework helps in identifying the various pathways to value

generation, and points out to the current tension between private/public and economic/social domains. The model helps in analyzing how value is generated from OGD (Jetzek et al., 2013). Table 6.2 below describes the enabling mechanisms, including the configuration of context, real objects, and causal powers that help in achieving certain outcomes.

Mechanisms and Context	**Real Objects** People/Systems	**Causal powers** Conditions/ Pressures	**Outcome**	**Linked Mechanisms** Dependencies, feedback loop
Efficiency - Creation of more effective methods of collection, management, distribution and use of data - automation	- Public sector - Private sector (beneficiaries of more efficient public service) - Computer applications that render and manipulate OGD to create value	- Vision of a more efficient government	- Generate economic value.	
Transparency - Create an open and transparent government	- Public sector - Private sector - Citizens	- Promise of openness - Technical connectivity and governance - Political leadership	- Reduced corruption - Reduce poverty	Innovation
Innovation - Supply data as a service	- Public sector - Private sector - Computer applications that render and manipulate OGD to create value	- Transformational effects resulting provision of OGD - Openness and technical availability of data	- Generate economic value. Examples: Netherlands - 400% increase in turnover for private sector re-users, 250% increase in high-end users, a rise in the use activity of re- users of 300% and an increase of over €35 million on corporate tax returns (deVries, 2012).	Transparency Efficiency
Participation - Crowd-sourcing activities. Engage the public to inform government solutions and decision-making.	- Public sector - Private sector - Citizens - Social practices	- Positive effects of scale where openness and sharing enable value generation drawing from a larger pool of resources - Social value. This could be what drives individuals and organizations to share their resources without direct monetary reimbursement. - Engagement platforms between government and citizens	- Improved citizenship and collaborative behavior through **crowdsourcing** activities. Examples: natural disaster incidents, such as hurricane Katrina and the earthquake in Haiti (Lee and Kwak, 2011).	Transparency

Table 6.2: Generative Mechanisms of Open Government Data

6.2.2 Strategic Information Systems Planning Using Critical Realism

Morton (2006) demonstrates how critical realism can assist in improving the understanding and finding solutions in information systems projects, which are characterized by complex

interactions and unpredictable outcomes. Supportive and unsupportive mechanisms are proposed as the basis of a causal explanation. These could assist in developing and implementing strategic IS plans (Morton, 2006). Some of the identified mechanisms may apply in KODI, and therefore act as input to the data collection phase, as they inform the researcher on what to seek for. They could also act as a lens in the analysis phase, while trying to generate an understanding of what the current case is, what works, and what strategies need to be transformed or introduced.

Mechanisms and Context	**Real Objects** People/Systems	**Causal powers** Conditions/ Pressures	**Outcome**	**Linked Mechanisms** Dependencies, feedback loop
Challenge Diagnosis - critical assessment of the organization's existing information systems policies, arrangements and quality.	- Planning manager	Social setting of the organization	Case for change challenging existing arrangements	
Integration - Integrate various systems - Modify systems to better meet the information needs of districts and the Board	- The Board - Divisional Manager Corporate Service Division (CSD) - CSD branch managers - System owners	- Inter-dependent and standardized approach to systems development	- Inter-connected CSD systems	
Approval and control - Approve funding - Introduction of the new project management methodology - Control of the CSSP budget by the planning manager - Control of all process activities	- The board - Planning manager	- Withhold or grant funds - Determine how funds will be dispersed.	- Implementation of the Corporate Systems Service Plan (CSSP)	Resistance
Resistance - Not complying with the project management methodology - Insisting on early allocation of project funds from the CSSP budget	- CSD branch managers - ITB Planning Manager - System owners - District managers	- The branch managers and their system owners – they resisted new conception of the organizational role of their systems by arguing that the consultant's understanding of their functions was inadequate. District managers - resisted the attempt to close down shadow systems by arguing that the proposed new central systems were unlikely to meet their needs.	Resist the attempt to control the way projects were funded and managed	Approval and Control
Fragmentation - In the absence of comprehensive systems that can meet a wide range of information needs different organizational entities try to develop their own systems.	- System owners - CSD branch managers - District managers	- Divergent interests of subunits of an organization. *Large firms may have coordination difficulties in instituting a comprehensive IS planning system and be splintered by divergent interests.* - Lack of a comprehensive	Districts and some service delivery divisional users of CSD systems were able to meet their needs through the ready availability of desktop computers to develop 'good enough'	

		system that can meet a wide range of information needs from the various organizational entities	local shadow systems.	

Table 6.3: Strategic Information Systems Planning CMOs

6.2.3 Generative Mechanisms for Innovation in Information Infrastructures

Bygstad (2010) applies critical realism while investigating innovation in information infrastructures. In particular, this study sought to understand how information infrastructures can generate enabling mechanisms for innovation of ICT based services. The analysis proposes two self-reinforcing mechanisms namely, the innovation mechanism – which results in a new service, and the service mechanism – that results in more users and profits (Bygstad, 2010). These mechanisms may apply in the KODI context, which also seeks to spur innovation from the data supplied by government agencies. Innovation could also emerge from the processes involved in supplying data, which would be marked by greater efficiency, reliability and data accuracy.

Mechanisms and Context	**Real Objects** People/Systems	**Causal powers** Conditions/ Pressures	**Outcome**	**Linked Mechanisms** Dependencies, feedback loop
Innovation reinforcement - Decide on which components to build/reuse - Decide on whether to buy or outsource - Build and implement services - Sign contracts with vendors - Develop integration specs - Extension of the information architecture following new actors and operations	-IT - Sales - Marketing - Expert consultants. - External vendors - IT architecture.	- A (limited) number of key persons with a thorough knowledge of the dynamics of the information infrastructure, including a strong technical knowledge of the IT architecture	- Space of possibilities - Ideas for new services - Assembly of new components - New services	
Service reinforcement - Conceptualize and design solutions - Integrate add-on services from partners	- IT - Sales - Marketing - Expert consultants - External vendors	- A successful and expandable information infrastructure	- Attract more partners - Provide more add-ons - Reinforce value for users - More users	Innovation reinforcement

Table 6.4: Generative mechanisms for innovation in information infrastructures

Innovation reinforcement was added as a linked mechanism to service reinforcement. This was not made explicit in the publication, though it is implied in the explanation. This demonstrates the need for this exercise, as it helps in amplyifying some explanations explicitly.

6.2.4 Mechanisms Supporting Quantity, Accuracy and Traceability of Information

Fox (2009) supports the argument that critical realism should be applied to information and communication technology (ICT) research. This is based on the argument that it assists in

understanding the mechanisms and structures that assist in the realization of the desired outcomes. Unfortunately, there are very few applications of critical realism in ICT research to support this argument, or to provide guidance in similar research. As an effort to resolve this, this study provides an additional example based on a case study on quantity analysis software, which is used to is used to extract data from computer-interpretable building information models (BIMs) (Fox, 2009). The main aim was to identify the mechanisms and structures that increase the quantity, accuracy and traceability of information. This is also a requirement in OGD, and the research on KODI initiative could benefit from the findings of this study, as the researcher is informed on how to develop data collection instruments focused on these aspects, and also analyze the data using these findings as a lens.

Mechanisms and Context	**Real Objects** People/Systems	**Causal powers** Conditions/ Pressures	**Outcome**	**Linked Mechanisms** Dependencies, feedback loop
Fast Information - Computerize data extraction - Computerize data manipulation - Separate link files	- Software vendors' personnel - BIM data analysis software - Users of BIM data analysis software	- Computerized data extraction - Computerized data manipulation - Separate link files - User perceptions about the software.	- Faster BIM quantity information at a lower lifecycle cost. - Immediate disbenefit – Lack of time for estimators to gain insights about the building. - Long-term disbenefit. Following aggregation of immediate disbenefit, there is a shortage in commercially astute estimators. - General disbenefit – Lack of competitive companies due to aggregation of long-term disbenefit.	
Accurate Information - Install and Configure the analysis software with the aim of eliminating random human errors	- BIM data analysis software - Software vendors' personnel - Users of BIM data analysis software	- Possible to extract from multiple sources and levels - No random errors – Automating the analysis process will eliminate human errors	Accurate BIM quantity information at a lower lifecycle cost	
Traceable Information - Create visual and data links to the BIM - Create links to archived data	- BIM data analysis software - Software vendors' personnel - Users of BIM data analysis software	- Visual links - Archive links - Data links	Traceable BIM quantity information at a lower lifecycle cost	
Reconfigurable Information - Format data in various forms to suit different users	- BIM data analysis software - Software vendors' personnel - Users of BIM data analysis software	- Multiple data models - Interconnected data models - Flexible formatting options	Reconfigurable BIM quantity information at a lower lifecycle cost	

Table 6.5: Mechanisms Supporting Quantity, Accuracy and Traceability of Information

6.3 OGD Economics

Section 2.3 helped in describing some of the envisaged benefits of OGD. Though this is necessary in understanding what OGD can do, it would be important to also understand the costs involved. The following list helps in describing the economics of OGD, with the aim of clarifying certain assumptions and providing insights on what to consider when planning for an OGD initiative.

- The production and curation costs need to be considered whenever conversations about the free nature of open data arise. For OGD, government needs to be made aware of the costs and resources that will be required to ensure sustainability of this initiative (Kitchin, 2014).
- In addition, the claim that state data has already been paid for through taxes is not as simple as desired. Government would need to prioritise its expenditure and secure budget for this initiative as a new cost centre. For instance, new roles would be introduced since the existing staff may not have the required skills or capacity to perform some of the duties linked to curation and publication of open data (Kitchin, 2014).
- Open data is disruptive, resulting in disruption of some of the existing business models. Resistance from affected parties need to be planned for, and considerations on how their income streams will be sustained should be made without reinforcing an existing digital divide (Kitchin, 2014).
- Open data has the potential to generate revenue and create new markets and economies (Open Knowledge Foundation, 2012). However, it is important to note that open data markets are both high-end and low-end. High-end markets refer to markets with a low number of customers who have to pay a substantial fee for services, whereas low-end markets refer to markets where a large number of customers pay are required to pay a significantly low fee for services. Open data services need to be designed with this in mind, bearing in mind that even the low-end markets may not break-even in-terms of revenue, and may have to rely on alternative revenue streams such as advertising (Kitchin, 2014).

6.4 OGD Critique

Not everyone believes in the said benefits of OGD. Concerns raised by critiques of OGD can be grouped intor three categories. The first concern is that OGD facilitates the neoliberalisation and marketisation of public services, the second concern is that OGD promotes a politics of the

benign and empowers the empowered, while the third concern is that OGD initiatives lack sustainability, utility, and usability (Kitchin, 2014).

a) Neoliberalisation and marketisation of public services

Political parties and businesses have disguised their capitalist interests with the transparency agenda. This is made manifest by their lack of support for the right to information movements and whistle blowers (Bates, 2012).

b) Politics of the benign and empowering the empowered

Most of the focus has been on the technical and economic aspects of OGD, geared towards access to OGD. This focus needs to be extended to include the politics of OGD, which involves considerations such as 'what does this data reveal?' or 'how is it being used and for whose interests was it generated or processed?' (Shah, 2013).

OGD is characterised by a high degree of social values and privilege. For instance, whose interests are represented, and whose interests are excluded within a dataset? (Johnson 2013). Citizens have differential privileges, including access to required technological tools and skills. This implies that only the privileged will be able to analyse, contextualise, and interpret OGD (Gurstein, 2011).

c) Sustainability, utility and usability

OGD debate and practice has paid more attention on the supply of data, and not much attention has been given to the question of sustainability, and how OGD demand can be harnessed (Kitchin, 2014).

The availed datasets are often low-hanging fruit, consisting of non-sensitive data with relatively low utility. Not much effort is made to release sensitive data, which requires management intervention and additional processing to ensure compliance with data protection laws, which also need to be revised accordingly (Chignard, 2013).

It has also been observed that the initial spark of interest in an OGD initiative is often shortlived. This drop has been attributed to the limitations of OGD, which emerge after users start to use it. They struggle with understanding how the availed datasets can be profitably analysed and used. This shortfall can be avoided by paying more attention to the demand-side, through provision of services and tools that facilitate use (Kitchin, 2014).

6.5 Proposed OGD CMOs from Literature

This section provides a summary of the features that define an ideal OGD context, a proposed list of enabling mechanisms, and the expected outcomes from an OGD initiative. The context, enabling mechanisms, and outcomes listed below are derived from literature review provided in table 6.3 on OGD case studies. These assisted in identifying the CMOs that may exist in the Kenyan context after analyzing the existing literature on KODI. KODI candidate CMOs are provided in section 6.4 below.

6.5.1 Context

- OGD initiatives are funded by public sector through taxes.
- Government is not interfered to act on behalf of society to manage data risks which include quality and errors.
- There is only one source of data for either the public or government agencies. This ensures that the public is not disadvantaged.
- Citizens are adopting crowdsourcing and thus becoming active data producers.
- Government is creating sustainable channels for citizen and civil society engagement.
- There exists a strong sense of community between government agencies, citizens and the private sector.
- There is collective learning and intelligence which is supported by emergence of more advanced features in the open data platform.
- OGD is readily accessible free of charge.
- All government agencies are working collectively as a team to produce OGD.
- There exists a common understanding of OGD across government agencies.
- The country has an effective legal framework that supports openness and transparency.
- The political leadership are in support of OGD.

- There are regional OGD initiatives which encourage member countries to develop and implement OGD policy.
- There is a clear outline of the responsibilities of each stakeholder and the process layout for each activity.
- There is a clear outline of the process involved for each activity especially curating of data and feedback channels with the citizens and private sector.
- There are plans and resources to provide structured education and training for government practitioners especially on how to curate data through guidelines, workshops and conferences.
- There are adequate and sustainable resources to support competitions, application development contests and boot camps targeting the software development community.

6.5.2 Enabling Mechanisms

1. **Law and Policy** – These should support publication of data, protect privacy of government agents, and regulate copyright and disclosure issues.
2. **Skill management** – This involves adequate and sustainable training programmes on how to curate and publish data, handover, maintain low turnover, clear roles and responsibilities, and collective learning. Red flags include; high staff turnover, poor handover, inadequate training, inadequate staff, lack of clear roles and responsibilities.
3. **Government commitment** – There is need for dedicated, qualified, and well equipped staff. There is also need for goodwill from the current government. Resistance could be evidenced from either of the following: Failure to appoint a dedicated government agency to solicit datasets from other government agencies, lack of government structures specific to OGD, goodwill.
4. **Efficiency** – This involves accurate and timely release of data, clear processes (what next), and clear roles for those involved in the OGD curation process.
5. **Data Quality** – Good data takes less time to convert to machine-readable format, meta-data provisioning, open-standards format. Poor quality is characterized by increase in time to convert to machine-readable format, and lack of uniform data standards across government agencies.

6. **Transparency and accountability** – This involves commitment by government not to withhold data, not to release tampered data, formulate clear policies on the publication and use of data, and also openness of operations. Security and confidentiality for government agents is critical despite the call for openness, and measures should be taken to safeguard this. Lack of transparency and accountability results to loss of trust and goodwill.
7. **Crowdsourcing** – This requires the public to become active data producers, since some data will have to be sourced from the public. This would help in governance once proper channels for participation and engagement are established.
8. **Stakeholder engagement** – This involves identifying and engaging with stakeholders and identify their needs.
9. **Communication** –This involves conducting awareness campaigns, identifying engagement platforms where the public, including views from developer community's views on OGD can be heard and addressed.
10. **Adequate & reliable resources** – There is need to adequately plan and allocate resources for infrastructure, technical support, training, and staff capacity. There is also need to allocate funds and expertise to support bootcamps, and other activities that may support innovation and create awareness. Resources must not be delayed/ inadequate/ inconsistent. Cost of data is inversely proportional to access.

6.5.3 Outcomes

1. **Reliable OGD –** Data that is timely, accurate, machine readable, and based on open standards.
2. **Democratic participation** – Citizens are able to derive information from OGD and use this to exercise their democratic right.
3. **Transparency and accountability** – This is characterized by improved government integrity, which can be measured by the level of trust by citizens towards government.
4. **Improved public service delivery** – Citizens have access to better services at reduced/affordable cost of access. It is characterized by reduced transactional costs within government, increased number and quality of public services, and use of information curated from the public to support policy and improve efficiency.
5. **New revenue models** – This results from an increase in innovation and competition within the private sector sparked by OGD generated insights.

6. **New public-private partnership models** – This is characterized by collaboration between government agencies and other agencies such as donors and researchers.

6.6 Proposed CMOs Affecting KODI Institutionalization

Following a review of literature in the previous sections within this chapter, the researcher identified five mechanisms that could assist in enabling institutionalization and sustainability of KODI. The process of identifying mechanisms remains a challenge. This study followed the approach suggested by Bygstad (2016), which starts by identifying the structural components of a mechanism (context, real objects, causal powers, outcome and linked mechanisms), which then helps in validating the appropriateness of the proposed mechanism.

This study proposed several mechanisms including law and policy enforcement, skills management, efficiency, data quality, and stakeholder engagement. These would help in resolving existing challenges such as the lack of timely release of data, scarcity of skills, inadequate resources, and high staff turnover.

A summarized description of the required context for each mechanism, including the people or systems required, and a proposal of the enabling causal powers is provided. This also includes a list of outcomes for each, which could help in evaluating the effectiveness of a particular mechanism, by determining whether those outcomes are met or if another mechanism is required to assist in achieving a particular outcome. This is important because it is at times hard to associate a particular outcome to a certain mechanism in isolation of the rest since the study is conducted in the real domain, and one cannot exclude certain variables as would be the case for research conducted on the empirical domain, which depends on the observable events (Bhaskar, 2008; Easton, 2010; John Mingers, 2002). These mechanisms will act as input in the next phase that involves formation of research instruments, and will serve as an initial lens during data collection. The fourth phase on theory testing will determine whether these were the necessary mechanisms or whether the data suggested otherwise.

Proposed Mechanisms and Context	**Real Objects** People/Systems	**Causal powers** Conditions/ Pressures	**Outcome**	**Linked Mechanisms** Dependencies, feedback loop
Law & policy enforcement - Identify, amend or create suitable laws and policies on data publication and use - Control publication and use of data using copyright laws and disclosure policies. No one should withhold data/release tampered data - Protect the privacy of government agents - Establish and implement a legal framework and policies on right of access to information, confidentiality, exceptions to openness, and intellectual property rights - Establish engagement platforms aimed at supporting citizen Participation in policing and law enforcement	- Data Publishers - Data Aggregators - Government - Citizens - Civil societies	- Copyright laws - Disclosure policies - Government commitment - Openness and transparency of government officials	- Publication of data - Confidence among government agents on privacy - Transparency and accountability	- Skill management
Skills management - Train staff on how to curate and publish data - Facilitate proper staff handover - Set and manage staff roles and responsibilities - Establish competitive channels for staff recruitment - Develop and issue best practice guides - Build understanding and capability amongst officials from across government departments and agencies	- Data Publishers - Government - Researchers - Infrastructure	- Low staff turnover - Comprehensive training - Staff motivation - Staff competency - Adequate infrastructure	- Efficient publication of OGD - Accurate data	
Efficiency - Timely publication of data - Clear definition and communication of processes, interconnectedness and actors involved - Proactive disclosure of OGD. This entails releasing data without waiting for specific data requests from the public. - Devolve decision-making. This allows communities a stronger say in choices of government programmes and services. It also supports proactive disclosure. - Acquire infrastructure - Develop a business case and financing model for curating OGD.	- Data Publishers - Infrastructure - Government	- Proactive disclosure policies - Devolved decision-making - Organizational cohesion – commitment from the entire organization. - Availability of infrastructure	- Reduced bureaucracy - Efficient data publication - Reliable data provisioning - Timely acquisition of resources	Skills management
Data quality - Provide meta-data for all data - Publish data using open-standards format - Process feedback from citizens - Obtain user perspective. This includes; availability and access, searchability, usability, understandability, quality, comparability and compatibility, metadata, and interactivity with data provider.	- Data Publishers - Infrastructure - Government - Engagement platform - Data Aggregators - App Developers - Citizens	- Inclusiveness of user needs - Comprehensive quality controls - Adequate skills and infrastructure - Public engagement. Engage the public to inform government solutions and decision-making.	- Quality data. Addresses current user needs. - Accurate and reliable data. - New revenue models. This is as a result of increased innovation and competition in the private sector	- Efficiency - Skills management
Stakeholder engagement - Identify and engage with the stakeholders and identify their needs - Organize conferences and bootcamps	- Data Publishers - Data Aggregators - App Developers - Infrastructure	- Collaboration among stakeholders - Shared control during engagement sessions.	- Better stakeholder buy-in. - Inclusiveness. Incorporate	

targeting data aggregators and app developers - Cooperate with independent volunteer and nongovernmental organisations - Interactive policy making - engage citizens in the process - Establish engagement platforms aimed at supporting citizen participation in policing and law enforcement - Conduct informative sessions aimed at: increasing public interest and preparedness; appreciating the value of crowd sourcing; changing the attitude of public officials on openness; ensuring stakeholder "buy in"	providers - Government - Citizens - Civil societies - Business representatives - Journalists	The organizers accommodate participant views	recommendations from the various stakeholders. - Crowdsourcing. Gather ideas from the various stakeholders that help inform decision making - New public-private partnership models - Democratic participation	

Table 6.6: Proposed CMOs Affecting KODI Institutionalization

6.7 Conclusion

This chapter has provided some insights on the structures and mechanisms that affect institutionalization of OGD in various countries. This analysis was conducted through the lens of table 4.1 institutionalization analysis guide, which provides a summary of the concepts of institutionalization and the research questions one needs to reflect on when conducting a case review. In addition to this, the researcher conducted a literature review of studies in Information Systems (IS) that were based on critical realism, to get an understanding of the kinds of mechanisms that are prevalent in IS, and whether some could apply in the KODI case. Following this, the researcher tried to define the ideal context, the necessary mechanisms, and the outcomes that one should expect in a thriving OGD environment. The findings in this exercise assisted in formulating a proposition of the CMOs that are likely to affect the KODI case. These findings helped in developing the data collection instruments and the analysis tools in the following stages. The following chapter explains the data collection phase.

7. Phase 2 - Data Collection

The aim of this phase is to obtain sufficient data to assist in understanding the context, and the underlying structures and mechanisms that affect the institutionalization of KODI. The research instruments will be formed based on the findings of the first phase on hypothesis, which are referred to as the preliminary theories (Pawson & Tilley, 2004). These theories were later tested and revised based on the collected data.

The research instruments consisted of different sets of semi-structured interview guides and codes for thematic analysis. Data was obtained from semi-structured interviews, document analysis, and media report review. These were grouped into primary and secondary sources of Data. Primary data comprised of semi-structured interviews and documents obtained from representatives at the, Kenya ICT Authority, Kenya Bureau of Statistics, government agencies and NGO's/Civil Society Organizations working on open data. Secondary data comprised of media report analysis including newspaper articles and videos on these initiatives were reviewed. The multiple data sources assisted in triangulating data and convergence of evidence.

The following sections provide a detailed description of the sampling procedures and the data collection methods and tools.

7.1 Sampling procedures

The method of purposive sampling was used to identify the sample group. Noting that this approach is based on non-probability sampling, the interview participants were identified based on their knowledge, relationships and expertise on KODI. These are people who either influence or are influenced by KODI as their work depends on its output (Freeman, deMarrais, Preissle, Roulston, & St. Pierre, 2007). The participants consisted of diverse categories of stakeholders including policy makers, non-governmental organizations, civil society, media, researchers, and technology implementing partners. Table 7.1 below demonstrates the different categories who were sampled, including the number of interviewees per category.

Stakeholder category	Semi-structured Interviews
Researchers	2
Government policymakers	6
County Government Official	1
KODI Team	5
KODI Data fellows	5
The World Bank	3
Civil Society Organizations (CSO)	7
@iLabAfrica – (Implementing partner)	2
Total	**31**

Table 7.1: Interview Sample per Category

7.2 Data collection methods and tools

The researcher explored the Kenya Open Data Initiative (KODI) from multiple data sources. Qualitative data was obtained from thirty-one face to face semi-structured interviews, and more than twelve documents sourced from either the stakeholders during interview sessions, the internet including blogs and websites managed by open data stakeholders, and newspaper publications from Kenyan media houses. The data collection process commenced after obtaining ethics clearance from the Commerce Faculty Ethics in Research Committee, University of Cape Town. The following section provides a detailed description of the process involving the semi-structured interviews and document review.

7.2.1 Semi-structured interviews

The interviews in this study were semi-structured and personal, except for one occasion when two members of the KODI team at the ICT Authority preferred to have a joint session, since their work was interrelated. Prior to conducting the interviews, the researcher obtained consent via email from each of the participants to be interviewed and audio-taped. There was one participant who did not want to be audio-taped, but allowed the researcher to take notes during the one-hour interview. Another participant allowed the researcher to audio-tape only the first half of the interview, and to take notes in the other half. All the participants were issued with the interview questions prior to the semi-structured interview. The interviews took place at a location choose by the participant. Once the interview was complete, the audio-recording was transcribed verbatim. In the case where there was no audio-recording, the researcher took notes during the

interview session. The documents and interview transcripts were stored on the researcher's computer and backed up securely on google drive.

There were different types of interview guides depending on the participant's role. These questions guided the interview in fulfilling the research objectives, and also helped in forming additional questions specific to the participant in question while still focusing on the research objectives. Some of the questions forming the semi-structured interview guide are provided in table 7.2 below, and the complete interview guides are provided in Appendix 1.

Government Agency Management Interview Guide
Section 1 Q1. What is the focus and objective of your work in relation to open data? ***Section 2*** Q1. What data is your institution mandated to publish to the public? ***Section 3*** Q1. Describe the workflow for curating and publishing data? ***Section 4*** Q1. What policies have been adapted within your institution to assist in implementing the initiative?
Data Fellows Interview Guide
Section 1 Q1. What was your role as a data fellow? ***Section 2*** Q1. Describe the workflow for curating and publishing data? ***Section 3*** Q1. What policies have been adapted within the institution you worked in to assist in implementing the initiative?

Table 7.2: Sample KODI Interview Guide

The face to face interviews ranged between forty five minutes and one hour and twenty minutes, and involved the following stakeholders: Dr. Bitange Ndemo, Former Permanent Secretary of Kenya's Ministry of Information and Communication (2005 to 2013); Ten Kenya ICT Authority (ICTA) staff members working on KODI, five of whom were temporary staff under the data fellows programme; Mr. Cleophas Kiio, Director for Information Communication and Technology at the Kenya National Bureau of Statistics; Three world bank staff including Mr. Robert Hunja who is the Director for Public Integrity and Openness, and also Governance Global Practice; Michelle Willmers who did a comparative research between Kenya and South Africa on open data institutionalization; two staff members at Strathmore University - iLab Africa; Assistant Director of ICT at the National Treasury; IT Officer County Government of Kiambu; Chief Francis Kariuki, the head of the Lanet-Umoja community; George Nengo, Senior Director

for Parliamentary Affairs at the Office of the Deputy President; Mr. Samuel Musumba, Programme Director at National Transport and Safety Authority; three Code for Kenya team including David Lemaiyan, Catherine Gicheru and Njambi Rono; Jessica Musila, Mzalendo's Executive director; Muchiri Nyaggah, Executive Director at the Local Development Research Institute (LDRI); Davis Adieno, former Capacity Development Manager at Development Initiative; and Open Institute founders, Jay Bhalla and Al Lags.

7.2.2 Document review

Table 7.3 below provides a list of documents that were reviewed in this study, with the aim of understanding how KODI is being institutionalized. To complement this, nine videos published on Vimeo by the Open Institute were also reviewed. The videos consisted of personal interviews with the pioneering KODI taskforce members, and provide a detailed description of how the initiative unfolded – including expectations and recommendations from each participant. The videos were transcribed verbatim.

Author	Document Title
Kenya National Bureau of Statistics (KNBS)	Concept note and mid-term reports from the data dissemination capabilities through research and attachment programme. This programme was in partnership with the World Bank as a donor and Strathmore University as the implementing partner.
Maya Gainer	Judicial Sector reforms in Kenya, 2011-2015
Nation Media Group	National newspaper articles on KODI
Commission on the Implementation of the Constitution (CIC)	Memorandum on the Freedom of Information and Data Protection bills by the CSOs to the Commission on the Implementation of the Constitution (CIC)
iHub	Research reports on KODI
Jessica Bayern	Impact of Open Data Initiatives in Kenya, Uganda and the Philippines
IODC	2015 International Open Data Conference (IODC) Report on enabling the data revolution, an international open data roadmap
KODI	Blogs articles
Francois van Shalkwyk, Michelle Willers and Tobias Schonwetter	Embedding Open Data Practice, developing indicators on the institutionalization of open data practice in two African governments (Kenya and South Africa)
Government of Kenya	Laws of Kenya - Official Secrets Act
Linet Kwamboka	2013 Kenya Open Data Initiative Report
Linet Kwamboka	Data Revolution framework for Africa

Table 7.3: List of Reviewed Documents

7.3 Conclusion

This chapter helped in describing the sampling procedure and data collection methods applied in this study. This was a qualitative study, which sourced data using semi-structured interviews and document analysis. A total of thirty-one interviews was conducted. This comprised of policy makers, non-governmental organizations, civil society, media, researchers, and technology implementing partners. A list of interview sample per category is provided in table 7.1 above, followed by a sample of the interview guide in table 7.2. Document analysis involved review of more than twelve documents sourced from multiple sources including stakeholders during interview sessions, the internet including blogs and websites managed by open data stakeholders, and newspaper publications from Kenyan media houses. A list of the documents reviewed in this study is provided in table 7.3 above. The following chapter provides a detailed description of the analysis phase.

8. Phase 3 - Data analysis

This phase seeks to identify patterns and map these to preliminary theory. It involves subjecting the programme CMO configuration hypotheses to test using the data obtained in the second phase. The aim is to identify the various patterns of success and failure within and across a programme (Pawson & Tilley, 2004).

To achieve this, thematic analysis was adopted as the method for qualitatively analysing data, which involves the identification and reporting of emerging themes or patterns (Braun & Clarke, 2006). The researcher derived themes from codes based on a five stage model developed by Fereday & Muir-Cochrane (2006). The model is an improvement to the work of Boyatzis (1998) and Crabtree & Miller (1999) who developed a guideline to thematic analysis and code development, and also a template approach to text analysis using codebooks. This is both an inductive and deductive process since the preliminary codebook will be formulated from literature (deduction) and modified using data (induction) (Crabtree & Miller, 1999). This process is described in the following section on thematic analysis.

Fereday & Muir-Cochrane's model consists of the following stages; developing the code manual, testing reliability of codes, summarizing data and identifying preliminary themes, applying template of codes and additional coding, connecting codes and identifying themes, and corroborating and legitimating code themes. A description of each of these stages is provided below, including an explanation of how the researcher applied them to this study.

8.1 Developing the code manual

This is a data management tool that assists in organizing portions of similar or related text and deriving new insights from the emerging patterns (Crabtree & Miller, 1999). This was developed prior to analysis and was based on the research questions and theoretical concepts derived from literature in chapter six. The findings from literature review determine appropriateness of a code, and suggest the phrase or words that should be used as a code. Table 8.1 below provides the findings of this exercise.

The code manual comprises of the label or name, definition of what the theme concerns, and description of how to know when the theme occurs (Boyatzis, 1998; Fereday & Muir-Cochrane, 2006).

#	Code	Definition	Description
1	Open data & big data	The aspect of open in data is that it is free to access and use without any restrictions or mechanisms of control. The aspect of big in this context applies to open data in very large or complex volumes of datasets, which require specialized tools to manipulate.	Mention of open/big data in relation to access, use or process.
2	Open data curation process	This is a process that involves extraction, annotation, publication, and maintenance of data from multiple sources.	Description of the curation process or strategies supporting this.
3	Completeness & Correctness of data	These are attributes of data. It needs to be released without delay, contain all the required attributes, and provide the correct values.	User perception on these attributes during interviews or media publications describing their perception about them.
4	Public & Private sector demand and engagement	This refers to the demand side of open data. This code seeks to identify patterns where the public and private sector are engaging with the published OGD, requesting for new datasets, and whether a feedback mechanism is in place and active.	Mention of activities involving public or private sector use or request for open data (demand), or engagement of either in open data activities.
5	Stakeholder buy-in to publish	This refers to the uptake of government agencies to curate and publish data, either on their platform or the KODI platform.	The willingness of government agency to publish data. This could be manifest in their commitment through the data release form, allocation of resources to curate and publish, or formation of internal policies supporting open data publication.
6	Impact of KODI	This seeks to collate all the impact related information.	Transformation or introduction of new government services as a result of KODI.
7	Collaboration & teamwork	This seeks to understand the strategies in place to support collaboration and teamwork among stakeholders. The focus was on agencies responsible for curating and publishing open data, and their relationship with ICTA.	Open data activities involving several stakeholders. Commitment to implement a policy or common action plans on open data.
8	Decentralization	This seeks to understand whether OGD affects power distribution. This follows the perception that withholding information creates power.	Devolution of power structures.
9	Equip & Educate	These two aspects identifies strategies and actions towards building capacity and providing resources to assist in data curation, management and visualization.	Training activities or resources provided or set apart for open data implementation.
10	Culture	This seeks to understand the ideas and customs that are related to KODI.	An account of new ways of working, buy-in or resistance to adopt open data, change of perception about open data.
11	KODI adoption process	This seeks to document all the strategies and activities involving KODI. The aim is to identify those strategies that have worked and those that haven't, including the resources and support involved in their implementation.	Open data activities, policies and strategies including dates on when they were formulated and implemented.
12	Public participation	This aspect relates to any efforts on KODI to consult with the public using two-way communication channels. The aim is to be more inclusive, which results in greater ownership/buy-in.	Public participation on open data activities.

13	Political influence on adoption	Political influence is the ability of individuals to impact on government decisions on policy implementation. This determines ownership and buy-in from government. This includes an understanding of the perceptions of government towards OGD.	This includes formation or support of legislation (law and policy) that affects OGD. Goodwill on OGD initiatives which could be manifested through public statements.
14	Privacy & Confidentiality Concerns	This refers to information that should not be divulged to the public in the interest of the state or to protect the safety of the government official responsible for that data. At times officials may refuse to publish data when they have insufficient knowledge of OGD.	Reaction by government agents on calls to publish government data. Data requests vs. published data including timeliness, reliability (frequently updated) and accuracy of data.
15	law & policy formulation & enforcement	This entails formation and implementation of laws and policies that affect OGD. Sometimes a law or policy could conflict with OGD implementation and may require an appeal.	Formation and implementation of laws and policy that either support or oppose OGD implementation.
16	Transparency & Accountability	These principles are mutually reinforcing and enable citizens to contribute to issues affecting them and provide an opportunity for them to influence decision-making and hold those in public office accountable. OGD assists in achieving this by empowering citizens with data for informed decision making.	Publication of open data as per data requests. Feedback mechanisms to engage with the citizens.
17	Ownership by government agents & agencies	This refers to commitment by government agencies and representatives to support the OGD cause. This includes availing of required resources, and formalising of OGD processes.	Reaction by government agents on calls to publish government data. Data requests vs. published data including timeliness, reliability (frequently updated) and accuracy of data. Measures to build capacity and provide resources for OGD implementation. Formation of law and policy in support of OGD.
18	Openness	This is a concept that is characterised by transparency of government operations, which is supported by the publication of OGD.	Willingness by government to provide data regarding its operation. This includes turnaround time to respond to request for datasets, and constant update of existing datasets.
19	Open Data ecosystem	This refers to an interconnected network of Open Data actors in Kenya.	OGD activities, which provide information on the stakeholders and their roles.
20	Open data policy	This refers to a set of principle or a defined course of action on how to implement open data, which leads to institutionalization after these principles and actions have been internalized.	Draft or published open data policies.
21	Domesticating development initiatives and policies	This refers to efforts to localize international or national policies to a specific context, and adjusting some of the strategies used to implement a particular initiative based on the dynamics of the context in question.	Strategies by government agencies based on national or international policies or initiatives.

Table 8.1: Code manual

8.2: Testing the reliability of codes

This aims at testing the applicability of the code to the data (Boyatzis, 1998). Five interview transcripts were selected as test pieces. Table 8.2 below provides an illustration of how the codes were applied to the transcripts using Atlas.ti as the tool for thematic analysis. Note: this is just a sample of how reliability tests were conducted on the codes identified in table 8.1 above.

Code	Data from transcript
KODI adoption process	So yeah, we put together the task team and we started meeting couple times a week, and it was fantastic, like a miracle actually, it was really great, we had fantastic support, especially from the tech community, we involved people from civil society as well, and then from within government we brought in a few people within government, from the ministries involved, from the ministry of planning and the Kenyan bureau of statistics, and as a team we started planning the launch, putting together the platform and we met our 3-weeks deadline.
Ownership by government agencies and agents	You know that know after we launched it, it worked for three years, now it is was sort of not very functional. We have not updated it, simply because many people opposed it, many people didn't want Open Data. Simply because people feared that when information is available, a lot of power comes from holding data. That's why most people in government did not like that we could just give…
workflow, law & policy formulation & enforcement	But they can't say that anymore because Section 35 of the Constitution now demands that government should provide data to the citizens. It is good even those who are in government opposing when they leave they would want to have information themselves. The problem has been that the law has not been operationalized.
Stakeholder engagement, buy-in & key drivers	So the world bank played a role in supporting with information and a bit of knowledge and we also supported with some of the data curation because we were regularly looking at Kenyan data. So things that were already in the public domains that we could help with we did, but it was really…I think that was part of what's special about it.
Leadership, Planning, Team Commitment, Capacity & Resources	KNBS challenges include Capacity and mandate. We could be the leading role but need but we need the ministry of communication and technology. We lack resources and funding. If we had funding, we could have setup a similar unit.

Table 8.2: Codes Reliability Test

Some of the codes, like the first two in table 8.2 above were not modified. However, others like the third one, which initially was labeled "law & policy formulation & enforcement" were modified to include the concept of workflow. This helped to capture instances where workflow supported by these structures emerged from the review. Another similar example is the second last code on table 8.2 which initially labeled "Stakeholder buy-in to publish", but was modified to include two other aspects, namely engagement, which is necessary in achieving stakeholder buy-in, and key drivers, which help in capturing instances where explanations on what drives stakeholders to engage emerges. New codes also emerged, especially in cases where the preliminary code manual did not fit an emerging theme, as is the case for the label on "Leadership, Planning, Team Commitment, Capacity & Resources", which aims at capturing the enabling attributes for OGD within government.

8.3: Summarizing data and identifying preliminary themes

This stage involved summarizing of raw data, which comprised of individual transcripts and documents under review. The outcome of this exercise is a change in the subconscious as the researcher goes through the process of listening, reading and summarizing each dataset, where

information is internalized as it is absorbed and analyzed subconsciously (Boyatzis, 1998). The researcher made an effort to summarize by noting key points from the interviews as they occurred, and also as he reviewed the interview transcripts and documents before and during the coding process.

This process of data analysis was not focused towards content analysis, and as a result, even single/individual comments were found to be equally useful, compared to those that appeared multiple times from several sources. This formed part of the preliminary analysis, which provided needed insight on the potential themes that could be found in the raw datasets (Fereday & Muir-Cochrane, 2006). Table 8.3 below provides an illustration of some of the key points made by the researcher per research question during the analysis phase. Headings or short phrases of the questions were formulated and used for this exercise. This was a necessary analytical step as it helped in bringing to the surface additional themes. This illustration is based on questions directed towards data fellows, who were part of the ICT Authority but deployed to various agencies to build their capacity in OGD management.

Interview topic	**Summary of responses**
Data fellow roles	Train agency staff on how to curate data. Curate and publish OGD. Communicate and coordinate ICTA process on OGD including signing of data release forms by the agency for each dataset.
OGD curation and publication workflow	Obtain buy-in from departments within an agency to release data. In some cases government officials asked for an email with a questionnaire on the data the data fellows needed. Once the data was curated by the data fellows, it was sent to an ICTA staff for verification. Once its verified, the data fellows present a data release form to the responsible government agency staff for signing. Once this commitment and agreement is obtained, the data is released to the public.
Quality assurance measures	Data fellow to send dataset to ICTA staff for verification. Prior to signing of the data release form, allow the agency to verify correctness of the dataset.
Measures for ensuring timeliness in data release	The current measure applies to existing datasets. The data release form includes commitment by the agency to release updates at specific times. ICTA staff use this to send reminders.
Challenges as a data fellow	They were perceived as auditors which created resistance. Some agencies were not aware of OGD. Some agencies were also not aware about the data fellows, leaving the introductory work to the fellows. Not all department in an agency cooperated in the OGD curation process. Time constraint - The data fellows programme ran for only six months, which was not sufficient for

	them to collect data from all departments within an agency. Some data was complex. Some agencies felt that protocol wasn't followed in the communication leading to resistance (bureaucracy/red tape).

Table 8.3: Data Summary

8.4: Applying template of codes and additional coding

This step involves applying the code manual to identify meaningful text from data (Crabtree & Miller, 1999). The transcripts and project documents were loaded into Atlas.ti, a qualitative data analysis software. The code manual was also loaded into the programme. These codes helped in guiding the analysis of the text. Table 8.4 below provides details of the preliminary codes, including code description and sample matching text from the interview transcripts for each.

Code	Description of code	Matching text
Public & Private sector demand and engagement	Mention of activities involving public or private sector use or request for open data (demand), or engagement of either in open data activities.	"I don't know if you saw in Smart Company today there is a not so nice article that has been written about KNBS? You can have a look at it, they are being criticized today. Is it about Open Data and their speeds or the way they release it? No, they are being criticized...I think it's...if you even go to the Nation website it's one of the top most popular articles on the side if you see it. They are criticizing KNBS because the data they are producing is wrong." "But we want to know whether this exposed data, until we find a mechanism of trying to identify the people who have used it. Safaricom could be using it to launch their new product, like I&M could be using it. There's somebody who has developed an App, he could be using it in the backend, I might be using that App, I might never know..." "I think these were some of the challenges outreach also from the developer point of view because once the portal was launched there was no outreach as far as to, how you can engage with the portal, how you can use the APIs that kind of thing."
Collaboration & teamwork	Open data activities involving several stakeholders. Commitment to implement a policy or common action plans on open data.	"That was the first time the Kenya ICT board had come in, and they are really good at mobilising the right people. We had our first taskforce meeting, I believe that Socrata is the platform that was recommended by the World Bank - DC, they were already using it for their own data portal. It was an out of the box solution. There was a bit of resistance from some local developers who said they could come up with a solution, but we could not take this chance. I mean, this was a Presidential directive. In hindsight, that was the right decision. I think that even now, it's a bit sad because developers could actually, six months down the line, developers could have proved the government wrong - unfortunately they haven't." "I cannot comfortably name particular organizations by name, because there are various reasons why we collaborate with various people. So giving five weight and leaving out the rest or something like that for me doesn't sound very fair. Because there is a reason why we collaborate with some, there is a reason why we might not be able to work with others. But I would just say that within the four institutions that I had mentioned that we do work with regularly with people and we do not tend to work with one." "Do not provide data exclusive for KODI. Support the Open Data Initiative. KNBS challenges include Capacity and Mandate. We could be the leading role but need but we need the ministry of communication and technology. We lack resources and funding. If we had funding, we could have setup a similar unit."
Equip & Educate	Training activities or resources provided or set apart for open data implementation.	"Now it's touching at the core that we can eliminate poverty simply by providing information, more resources. We need information, we need knowledge and the only way you can do it is to put it out there so that some people can do the analytics of it, whichever way they look at it." "There has also been a kind of move with the news that I've watched and the documentaries that have seen on TV, there has also been a move into infographics and things like that. I don't know if this is a direct result of Open Data, but I feel like data has

		been widely adopted in news segments and newspapers and media than it was before. We had a workshop with the media sometimes back and one of the people we brought along was some consultants who've worked with news and who've talked about using data...I don't know the exact word for it but using data to send out messages and it could be accidental but since then I've seen all these things happen."
KODI adoption process	Open data activities, policies and strategies including dates on when they were formulated and implemented.	"So what happened and I am sure you've heard it and you've seen it in the videos in 2011 is that these several datasets became available...Ndemo got commitments people to give him these datasets and allow him to make it public. So we had that opportunity to launch a portal, and as you will see that even with just those datasets even if we didn't add there were a lot of downloads, there was still somebody that was saying, we didn't have that, the census data available there." "Yes, it takes long but we are working towards it. I don't think...there was never a time here that we were like, Oh whenever is going to hire staff, so let's forget about it, that's not true. We were just going through the process to establish it as a project and then do the necessary proper mechanisms to get the staff here. So here we are, last year I did...two years ago we said, by the end of next year we will have the staff and we did Okay, maybe a bit towards the beginning of this year, but we got that."
Transparency & Accountability	Publication of open data as per data requests. Feedback mechanisms to engage with the citizens.	"The interesting thing is, and this also shows you where the country has come from. Is that at the launch, or before the launch, we had over 1700 people confirming that they were going to attend, and all these confirmations were done online. Normally, you wouldn't have expected this. It was a great relief when the PS called and said "it's a go". I would have loved to be a fly on the wall to see what the President was shown, or what the President actually said. But I believe that the case that was put forward by the PS was nothing about transparency or governance, it was all about what this can do for the youth, opening up this data, or what it can do for the economy. Government themselves would better understand how to govern citizens and provide public services. The power of Open data is that such businesses, individuals, developers, can actually add to the data, or can actually take from that data and produce economic value from it. It's not just about governance and transparency. It allows citizens a platform to engage with government, it allows government to engage with citizens better."
Ownership by government agents & agencies	Reaction by government agents on calls to publish government data. Data requests vs. published data including timeliness, reliability (frequently updated) and accuracy of data. Measures to build capacity and provide resources for OGD implementation. Formation of law and policy in support of OGD.	"Then there were stakeholders from...and they are obviously the World Bank was instrumental because they supported it in a number of ways including funding, including getting the right partner to build the portal in terms of Socrata. Then there were people like Google from a technical perspective, there was Kenya National Law Review Council, the guys who looked at the licensing. Then there were individuals as well, so myself, there was Athman, Mohamed was from the Trademark East Africa. Then from the Developer Community there were people like I-Hub, Erick Hersman, so you could see it was an all-inclusive process. Civil Society had also people like AfriCOG, so everybody was there to give in their input. I think if you look at the report for the case study that the World Bank took out there's a long list of people that were involved."
Openness	Willingness by government to provide data regarding its operation. This includes turnaround time to respond to request for datasets, and constant update of existing datasets.	"It is in our constitution that institutions should you give out data to the public -- public should know what they're doing. The data that was given, it was uploaded to the Open Data website on Kiambu, and I believe as at now they have some data in that site." "The hype is not the portal but conversion to openness." "If you read a lot about Open Data movement and the principals and what not, you will hear people talk about around government and data should be open by default. I don't agree with it, I do not want government data to be open by default, I want there to be a process that decides what becomes open and what doesn't. The likelihood for me just thinking about it, why? I don't want government data to be open by default just so that there can be a cheap supply of data in the place."
Open Data ecosystem	OGD activities, which provide information on the stakeholders and their roles.	"You heard my colleague Bill saying that even in the UK 20% of the datasets on the government portal has now been looked at by anybody. So the assumption that getting Open Data and publishing it then will get it use is now being challenged. Now people are moving to the narrative to how do we get to use this data, how do we get data that's more relevant to people, what are people looking for that we are either providing or not?". "But Kenya has the ICT community, it has the developers, it has people working with the NGOs that are working with different constituencies. Kenya can really take…it's a net benefit to the economy. If you start big industries you can build the ladle Open Data, but it's a matter of leadership. But I think when you lost PS. Ndemo you lost the guy who was really driving this".
Domesticating development initiatives and policies	Strategies by government agencies based on national or international policies or initiatives.	"But specifically around development issues. So our work now is at least based in organization line, it's focused on, we are looking at how, to what extent governments are being able to domesticate development initiatives or treaties that have development impact. So the current level of effort is on building the initial dataset for that."

Table 8.4: Coding based on Preliminary Coding Manual

During the coding process, the researcher found the need to modify some of the codes to capture certain themes and aspects that were emerging from the data being analysed. Table 8.5 below provides a detailed description of these changes.

Code	Code Change Description
Open data & big data	The aspect of government data was missing and was introduced through OGD, resulting in a new code: "Open data, OGD, & big data".
Open data curation process	The researcher found it necessary to include the various aspects involved in the curation process related to data and how it gets rendered to the user. The resulting code was "Open data source, formats, standards, curation & visualization"
Completeness & Correctness of data	This code was modified to include additional properties of open data resulting in "Availability, Timeliness, Completeness & Correctness of data" as the new code.
Stakeholder buy-in to publish	It was found necessary to include the aspect of engagement as this precedes buy-in. In addition to this, there was need to identify the key drivers resulting to buy-in. This resulted in the following code: "Stakeholder engagement, buy-in & key drivers"
Impact of KODI	In understanding impact, it was necessary to consider how accessible the data is, which could be determined by the visibility of the initiative. Following these aspects, the application of open data was observed, which helped determine the impact of KODI. This resulted in the following code: "Accessibility, visibility, application & impact of KODI"
Decentralization	The main aim was to capture the impact of devolution in Kenya. This code was removed since there was no significant information arising from the data.
Culture	This code was expanded to include historical legacy & value system.
Public participation	This code was replaced by 'Public & Private sector demand and engagement'. Participation could be likened to a programme that citizens have to go through. Engagement is more citizen driven allowing them to engage randomly through multiple engagement channels, including informal ones like social media.
Political influence on adoption	This code was replaced by "Political economy, capital, leadership, buy-in". Political economy refers to relations with government and the law, and the distribution of national resources. Political capital refers to the influence the open data champions have with the public and political leaders. Political leadership entails an understanding of the political processes and outcomes, which is necessary for leaders of the open data initiative. Political buy-in refers to acceptance by government leadership to the proposal to implement open data and their will to support it.
Privacy & Confidentiality Concerns	This code was modified to include the aspect of sensitivity resulting in a new code labeled "Sensitivity, Privacy & Confidentiality Concerns". Sensitivity refers to measures to control access of information in order to safeguard a perceived advantage in hoarding it, or maintaining a level of security, which would be lost if it was made public.
law & policy formulation & enforcement	This code was modified to include workflow, resulting in "Workflow, law & policy formulation & enforcement". This helped in capturing the processes involved when curating data, and understanding how the supporting laws and policies emerge and are applied in the process.
Open data policy	This code was replaced by "workflow, law & policy formulation & enforcement".

Table 8.5: Changes on Coding Manual

The modified codes were applied in the coding process as they emerged. The use of Atlas.ti helped in this process as code modifications also applied to previously coded text. This implies that modifications should not be fundamentally different from the preliminary code, otherwise a new code should be formed. Table 8.6 below provides a description of each code, and an illustration of how the modified codes were applied using extracts from the transcripts.

Code	Description of code	Matching text
open data, OGD, & big data	Mention of open data, open government data (OGD) or big data in relation to access, use or process.	"There is no data either and statistics is not data either, it is data, but for us the term data is now a composite word that deals with the stats, that deals with the raw data coming out of stuff that are doing, it's big data, it's now this one word that we are using interchangeably." "What is important is agencies being converted to understand the essence of open data. The data portal from 2011 should be used as an example of the value of opening up. It is not a must for data to sit in one portal. The hype is not the portal but conversion to openness."
open data source, formats, standards,	Description of the source, curation process or strategies supporting this.	"I feel that quality of data is going to improve only when we bring this data to one store because then both organizations start asking themselves why do we have discordant data, why don't we agree, where do you collect the data from, who…?" "You do a lot of, you consolidate the data and when our minister presents it to Parliament,

curation & visualisation		and it's approved, then it becomes public data, and I can now format it and send to Mr. Sifa to publish it, so there's that procedure which guarantees…" "So we provide this platform, the County Government can upload datasets to automatically handle all the visualizations. So they don't have to build capacity for visualizations, all they need to do is upload that CSV, and it does the rest of the magic"
Availability, Timeliness, Completeness & Correctness of data	User perception on these attributes during interviews or media publications describing their perception about them.	"So, it's periodical and we hope to improve that in future, we will be able to avail more data like now expenditure which can be done at whatever period." "Another instance is somebody who was trying to see…that's my classmate who was trying to ascertain the viability of MTR – Mobile Termination Rates and this guy is looking at the historical data like how far would we've come with the MTR up to the point where we're right now. He calls me up and he says, "Hey, do you know anybody at ICTA who can give me this data?" And I am like, "what data?" So he says MTR…actually crazy enough he didn't even know it's called MTR which probably means there's a need to be a change of name or there need to be clearly defined thing that somebody who is not either serving that industry can still be able to access that data." "So, if there is some data that's incomplete you will know and how comes there is certain data on some fruits that are missing for 2015 that were available in the data for 2014, for once you've already noted such anomaly or incompletion or inconsistency you would then raise a question, so that way you would send an email copy it to ICT Authority of course through the post person, so when they respond then they might say something like the data was never collected or it's there but was never really synthesized, or whatever they say."
Stakeholder engagement, buy-in & key drivers	The key-drivers that result in buy-in by government agencies to publish data. This could be manifest in their commitment through the data release form, allocation of resources to curate and publish, or formation of internal policies supporting open data publication.	"We tend not to…email is good but with my experience agencies and they don't take these emails seriously or something, we want such that it can be documented earlier, so even if we write to you and you don't correspond we do follow up and say, 'We've got this letter that was received by you and stamped on this date and this…' **Followup question:** So you send formal letters? **Answer:**Yes, we send formal letters." "With the coming of governors and [President] Uhuru's administration, that has certainly changed, so no matter how much we try to build the systems that were useful. They somehow never really grew in to what they could fully be or truly be because there was no demand. Its same for the Open Data, there's stakeholders and the supply side and it's very important that the key line ministries are on board, evidently they are not, maybe in word but certainly not in action." "You know that know after we launched it, it worked for three years, now it is was sort of not very functional. We have not updated it, simply because many people opposed it, many people didn't want Open Data. Simply because people feared that when information is available, a lot of power comes from holding data. That's why most people in government did not like that we could just give…"
Accessibility, visibility, application & impact of KODI	Transformation or introduction of new government services as a result of KODI. This includes an assessment of KODI accessibility, visibility and application strategies.	"Yeah, [LAUGH] one man army. I do Data Analytics stuff basically the entire focus of my position is to surface stories from the data, to make the data that we source at opendata.go.ke relatable to somebody who necessarily doesn't have a statistical background. So, I do this through maintaining a blog which is blog.opendata.go.ke and then I also do this by sending out periodical newsletters trying to tell people what we've done, which people we've met, what progress we have, the number of datasets that we have and something like new datasets of interest and things like that, a new blog post and stuff like that." "About the background of these people because the only other alternative places we could get information was the particular Parliamentarians CDF website, some of them had CDF websites which would give us this information about them. So, the main datasets we used from KODI were to do with health, to do with schools, why schools? Because schools in Kenya tend to double up or so as polling stations during election years, so we got the datasets for all the schools and then we got data on all the health centers. I think those were the main datasets we got from the portal". "I will tell you for free so many jobs in Nairobi right now are around Open Data for one reason or another. So it's me the fact that they created such beautiful jobs for so many people, that's amazing. If for no other reason that we gotten people out of their homes, off the streets, no tarmacking they are working on something. It has to do with government habit because someone took out time to fill forms and probably do the data collation like and that's amazing in itself. But then beyond that, it enables us in amazing Initiatives, whether on the policy side or tech side that there's chainings around these things."
Culture, Historical Legacy & Value System	An account of new ways of working, buy-in or resistance to adopt open data, change of perception about open data.	"Most of the data, you do realize and when you look in Open Data Kenya, if you look within our data available that we're able to handle it and manage it, it was at that basic level because we inherited data from the police, this is data that was kept, I would say for lack of a better word, the analogue way [laughs], its hard copies, files and those kinds of things. So, in tracing data 10, 12, 20 years ago it was so difficult, the files got lost, so once they give us the report it's like that is it. So, we are trying to reconstruct that data but still those gaps in there we are under-resourcing, over resourcing, inaccuracy [0:02:46], some data is left out, like I hear there was somebody is not in my role or an accident happened, there are those

		who die on the road, there are those who are injured, and yet some have died in hospitals, so we don't capture that data, that's lost"
Sensitivity, Privacy & Confidentiality Concerns	Identify concerns by government agents on calls to publish government data. Data requests vs. published data including timeliness, reliability (frequently updated) and accuracy of data.	"You know that know after we launched it, it worked for three years, now it is was sort of not very functional. We have not updated it, simply because many people opposed it, many people didn't want Open Data. Simply because people feared that when information is available, a lot of power comes from holding data. That's why most people in government did not like that we could just give…" "What is public is what is aggregated without bleaching confidentiality. Creation of that portal was based on the premise that government data is hidden. The survival of that portal is premised on the fact that other agencies data remain hidden – which is not the case. What they are doing is not opening data."
Workflow, law & policy formulation & enforcement	Formation and implementation of laws and policy that either support or oppose OGD implementation. This includes a description of how they shape the workflow.	"For the Open Data portal with the counties, we have agreed with the controller of budget that her team will send data in Excel tables to the Council of Governors to be able to put all together and upload. If it's a generated system made board that requires a lot of database for a few indicators. So much is here to handle and they are going to know what metadata is and you can post that as well [inaudible]." "Once we are done there will be...now we meet, we brief them on what we've done then if possible we can share with the management like the HELB we go to the management then after they've said Okay, it's fine...after we reach a consensus then now Okay that one can go to the public. We sign the Data Release Form and it goes to the portal. Then we come to another agency or that...if we maintain the relation with these agencies for updates for instance like at the HELB their dataset changes like every second, like as we are meeting today ten people have repaid their loans, ten people have been cleared, they have just detected somebody who has defaulted, they have sent some names to the credit bureau, and they have shortlisted them."

Table 8.6: Coding Based on Modified Codes

In addition to modification of preliminary codes, there were also new codes that emerged from the inductive process as the researcher internalized the text from the transcripts using the preliminary coding manual (Boyatzis, 1998). The code manual was updated as the codes emerged. A description of these codes, including sample matching text for each is provided in table 8.7 below.

Code	Description of code	Matching text
Bureaucracy	This refers to excessively red-tape administrative procedures that stifle the progress of the open data initiative.	"The Kenya ICT CEO is at the same level, because they are both agents, they are at the level of DG on the other side, so that guy has to send an email to this guy. If there's a meeting, both of them have to be there, you don't ask for a meeting and then you send a deputy or this guy is not there, that meeting is over." "So, after we have done the letter, in the letter we request for the meeting, the meeting is to tell them about the Open Data and in the course of that we need to know more about what they generate, then thereafter after the first meeting we write a letter again to them saying in the last meeting we held between Mr. A and Mrs. B from your side and Mr. A and Mr. B from our side we reached the following. So, we request that you authorize them to start sharing the data with us. That's done by CEO, it's signed by the CEO and sent back and then they communicate back...well, we even ask that they give us a primary contact person so that in case of anything we can follow-up and update and all that, so that we don't' go through the entire bureaucracy again. So after all this has been spelt out in the second, in their response letter then the engagement starts."
Data Journalism & story telling	Data journalism refers to the use of open data and infographics generated from open data to tell compelling stories. This code will capture instances where this is evident, including trainings and awareness efforts.	"These are conclusions you can come up to, because all these organizations are providing their data to the public, before you'd have to wait for somebody who you'd consider an authority figure to tell us this, but he looked at paradigms such as the number of houses constructed in those areas and all the areas in Kenya, major towns; Nairobi, Mombasa, Kisumu, Eldoret and Nakuru and then looked at the number of traffics of major airlines, and looked at a bunch of things. I can share with you a link to that you will be interested, you can go to blackorwa.com, it's called BLACKORWA." "What we are doing this year is something we are calling the significant number, so significant number is where we pick a dataset, do a blog post about it, but then try and point some numbers that we find very interesting. For example, this week's significant number is the number of road accidents in Kenya occur at 6 PM which is very insane

		because I am thinking everybody is stuck in traffic at 6 PM. Actually, not accidents but fatality like people die more at 6 PM, looking at Nairobi there's too much traffic and Nairobi contributes the bulk of road accidents in Kenya doing something about 50-60%."
Disabling forces	This captures structures and mechanisms that work against the realization of open data initiatives.	"So, there are some things that have already been institutionalized for a long time that will not change overnight because there is this kind of thing. But then change management really, happens for everyone. There is not going to be a time when someone will tell them there is open data and everyone will be – yaayaay lets go with this. No, there is going to be questions – why or for whom, so you have to manage that kind of process and that's exactly what we are trying to do now." "Like in the case of Kenya, you end up having a situation where, you cannot sustain the push for publishing data, because there is no policy involved. Government officials release data regularly because there's a written policy that tells them that they are supposed to release data. If you do not have that information, if you do not have that imperative, then are in general, you tend to not release the data."
Funding, Donors	This captures information on the source or the amount of funding.	"Actually the other thing that I was going to say about World Bank, I think also World Bank was already part of the Open Government Initiative at the beginning time when KODI was being formulated. I think they wanted to show an example of a sub-Saharan Africa example, and they happened to find a champion in Ndemo who were in line with their interest. So, I can tell you all the data that was on the site on KODI on the beginning was data that was funded by World Bank money. So, even if opportunities opened for the World Bank was they could have lobbied it, they could have talked to all the different official aid arms that were allied to also make their data visible on the portal, you know like the way all these governments which are part of the Open Government Partnership like the US government and the DFID UK." "So solving this coordination problem, and getting meaningful information out to the public, I think that's a very key role for the World Bank to support. One the lessons learnt is, you can't do reforms without resources. Voluntary work only last so long, so I think the big push will come once we have proper funding in place, and the World Bank is going to the board in march with additional financing request, with somewhere around 7million USD to support the open data initiative."
Innovation	This captures instances where new open data methods, ideas or products are either conceptualized, designed, developed or implemented.	"People think that it's a new discipline, and that there is a new science involved, there is not. Journalists have always worked with data. We just haven't worked with these massive amounts of data that are now available. So in the past it came to us in small samples or in static versions, as a pdf document or as a statement or even as a book. Now with the new technology that are available we are able to get real time data in phenomenally large samples." "There's also a push from developers and intermediaries, people want to work with the data to create products that can inform the people. Look at Ma3Route for example Ma3Route has right now 400,000 daily users -- 400,000 that's more than the population of people in Kisumu who access Facebook on a daily basis, that's insane."
Jurisdiction	This seeks to identify instances where official power to make legal decisions and judgements is required or applied. It helps in understanding the role and impact of certain stakeholders in the open data ecosystem.	"These are conversations that also comes up when you talk about Open Data and talk about KNBS, you talk about the issues of jurisdiction always come up. One of the problems people had with Socrata is that it's software that is hosted in the US jurisdiction. Despite the fact that this is data that's already open anyway it's not data that...we are dealing with sensitive data, people still...there is still a feeling, perception that our data sovereignty is being violated. It may not be an accurate assessment, like perceptions are powerful the moment you say these people are getting data from us to give the Americans, that changes the mood. It changes the mood for a lot of things and it's not just here it's a continental issue."
Leadership, Planning, Team Commitment, Capacity & Resources	This code combines several interrelated concepts. Leadership seeks to identify the leaders of open data initiative, and the strategies they apply in the process. Planning seeks to capture the goals and strategies driving various open data initiatives. Team commitment helps in identifying open data team dynamics, what brings and keeps them together. Capacity and resources is what enables the people involved to realize their goals.	"The other challenge that was there was obviously a funding challenge because what should have happened was and that again this was this was because of the time and the opportunity or more the opportunity and the time. The biggest challenge was not having the infrastructure in place to support it after it was launched. By infrastructure I mean the manpower in the ICT authority, there was nobody who was going to be responsible." "So, it has to be across, there's a...we want the structures, we want the data owners to know about it and we want the buyin right at the top. To get us started we had to get...it was fantastic to have a buyin right from the top, and it was so valuable and fantastic. But we want to work across all the government." "For me that's a bad sign, Linnet works two days a week in gov. The other three days a week she is CEO of Data Science, but they don't have a full-time person dealing with the Kenya Open Data or open government partnership issues, which signals the level of commitment that is there. It also tells you where their priorities are; it's not necessarily a bad thing. That says that you have to adjust your expectations about some of these things." "So the idea is that everything that's available for the public and somebody wants to give it to us certainly we'll publish everything. Now, at the moment we are not being overwhelmed."

Monitoring & evaluation	This code seeks to identify the processes that assist in managing open data tasks, improving performance and realizing preset goals on the open data initiative. It also identifies strategies used to assess whether a certain open data output matches preset goals.	"Also there had to be a report that was a follow-up to the work we've done, there had to be a monthly report showing what we have done and response from the clients who were in the same... it was structured report, the name of the host institution and the name of the programme fellow, so each institution had to do one report with Data Fellow, so we sit down together and file the report." "The challenges being able to access various datasets integrate because now they're still in silos, Treasury has got its own data, Health have their own data, Education its own data, so we need KODI to come, integrate that data and bring some meaningful insights." "Everybody is a stakeholder because when the constitution says that I have a right to that information. I grew up in a coffee growing area in Kisii. My mother got to a point where she uprooted the coffee because there was no value but at the international market coffee was being sold a lot of money. Nobody knew where the money went, that crisis is the one that led to people in Nyeri uprooting coffee. The governor did intervene by coming out of the Coffee Board because he found out most of the money got lost in between from the farmer to the market. Why didn't that data come out? What does it help the government in hiding such data? What does it help?"
Open data advocacy & awareness	Open data advocacy refers to the outspoken public support and recommendation of the open data initiative and enabling policies and laws. Awareness refers to comprehensive effort to help inform open data stakeholders including general public, private sector and government agencies on the value of open data, supportive laws and policies, access, and curation processes.	"So it's behind…it's not where we all hoped it would be. I think another issue that came up is that some people in the ICT community saw this as being someone else's initiative; they didn't see it as theirs. I think PS. Ndemo tried to involve many different people but you cannot involve everybody all the time. I think its people…people need to be a bit more open minded and say this is something for everybody." "Then for me I was a social media guy at NTSA so, what I would do if we have any engagements like workshop, I would post that on Facebook and Twitter, then you find from there people request, they say, 'It's good you have this, how comes you don't have that?' So, from there you're able to know. Then also guys at the host institution, there's a way they say, 'I think this is what the people need to know' So, probably that also helped." "We don't want a situation where the DG knows, the director of ICT knows, but others are not, because they are not looped into the communication. So they don't know who these people are and what they are intending to do. So that internal communication is very key, and in our second phase of this we will have to highlight this to them that please kindly ensure that your officers, the heads of departments within your agencies are aware of this."
Open Data Champion	This is a role characterized by an understanding of the open data ecosystem and curation processes, coupled with a strong political capital able to influence decisions among policy makers in various sectors including government and the political elite.	"Getting the census data was a very memorable moment, that was the point which we knew we were going to go for it. That was quickly followed by the next moment where PS Ndemo said "so, should I get the president to launch?" and I was like, "Yes of course. Get the president to launch". Then for the next meeting again on the next day he said, "So, I have got the president and I have lined it up and it's going to happen on the 28th of June, which was the original date". At the point we had the data, we had somebody who was willing to take the political wheel, and you could bypass a lot of internal bureaucracies if you have the president behind you launching it. I think that was kind-of key, if you don't have this political wheel, the champion in PS Ndemo, backed by somebody who is also going to make everybody else a little cautious along the way, the closer you get to launch date, the more and more people get nervous about what's happening, having that makes a big difference."
Political economy, capital, leadership, buy-in	Political economy refers to relations with government and the law, and the distribution of national resources. Political capital refers to the influence the open data champions have with the public and political leaders. Political leadership entails an understanding of the political processes and outcomes, which is necessary for leaders of the open data initiative. Political buy-in refers to acceptance by government leadership to the proposal to implement open data and their will to support it.	"One the first things that we did is PS Ndemo organized a breakfast meeting with key Permanent Secretaries and directors from agencies that are key holders of data with Kenya's National Bureau of Statistics, with the World Bank in a supporting role and our country director attended, a few members of our team attended. It was basically an initial sharing session about what is open data all about, what are some of the economic benefits of open data, because what we see around the world is, well I came into it from a bit of a governance prospective, open data actually has very strong rational from an economic prospective because its contributed to the growth of whole sectors of economies." "We need a champion at the cabinet level. It has got to be one of the two principals." "I think his [President] support was important, his endorsement was important and I think again he responded to PS. Ndemo and bringing Kenyan developers who can show how Kenya could make use of data and that Kenya could show Kenya well it's not such a threatening thing. It's really a thing about pushing forward with development and he responded positively to that. In that way I think he was important." "After the Code for Kenya Initiative, there was a slack, partly because there was a change of guard. The elections happened and new people came in, those were big parties in terms of playing politics, the ICT Board was disbanded into ICT authority. The main Champion for Kenya Open Data Initiative who was Dr. Bitange Ndemo who spoke widely about it was no longer PS. Now, Paul Kukubo who was the CEO of ICT Board left, went to Rwanda. So when we don't have the champions within then be the organization or we don't even to champion the product, it kind of we lose that…Yeah, one of the key people in that, Linet Kwamboka also left ICT Board authority. In this entire time the key champions once they left the thing started to fall apart."

Table 8.7: Coding Based on New Codes

Table 8.8 below demonstrates the groundedness and representativeness of each code. Groundedness refers to the total number of quotations linked to a particular code. This was generated through Atlas.ti document manager "export as table" utility. Representativeness is the proportion of documents that make reference to a particular code. Unfortunately, the "export as table" feature in document manager does not include representativeness. However, the researcher was able to obtain these values by using Atlas.ti document manager filter option to create a documents per code filter, which generated a count of all documents containing a particular code.

Code	Groundedness	Representativeness
Stakeholder engagement, buy-in & key drivers	233	32
Ownership by government agents & agencies	223	33
Leadership, Planning, Team Commitment, Capacity & Resources	220	31
KODI Adoption process	209	32
workflow, law & policy formulation & enforcement	190	30
Collaboration & teamwork	175	34
open data source, formats, standards, curation & visualization	169	28
Public & private sector demand & engagement	154	27
Disabling forces	152	28
Accessibility, visibility, application & impact of KODI	142	31
Open Data Advocacy & Awareness	119	29
Political economy, capital, leadership, buy-in	88	28
Open Data ecosystem	86	20
coordination & capacity building	82	20
Equip & Educate	76	19
Availability, Timeliness, Completeness & Correctness of data	74	18
Open Data Champion	69	24
Transparency & Accountability	47	19
Open data policy	46	18
Sensitivity, Privacy & Confidentiality Concerns	44	17
Innovation	37	14
Culture, Historical Legacy & Value System	32	10
Bureaucracy	25	9
Data Journalism & story telling	23	10
Monitoring and Evaluation	20	8
Funding, Donors	19	6
Openness	19	8
Domesticating development initiatives and policies	14	7
open data, OGD, & big data	12	12
Jurisdiction	3	2

Table 8.8: Code Groundedness and Representativeness

This table helped in determining whether the grounded score was generated by a high number of quotations in a single document, or whether this was generated from a sizeable number of

documents. This implies that even though a code may have the highest level of groundedness, it may not be the most representative as it emanates from a single document. The findings will help in determining the most grounded and representative codes, which is useful in determining what the documents and interview transcripts suggest as the most important aspects in institutionalizing KODI.

8.5: Connecting the codes and identifying themes

This process involves identifying patterns in data through internalization and matching similar codes resulting in themes that represent these patterns (Crabtree & Miller, 1999). The sub-sections below represent nineteen themes that were identified in this process. The source of these themes was grouped into four categories including; Ministry of ICT and ICT Authority (ICTA), ICTA data fellows, Non-Governmental Organizations (NGOs) and civic tech, other government agencies, and Intra-Governmental Organizations (IGOs). These themes were clustered based on the headings that relate to the respective interview research questions. Identifying themes per data group helped in identifying possible variations between the responses of the various groups (Fereday & Muir-Cochrane, 2006). It also helped in understanding the bigger picture as the interviewees or documents under review were only focused on one aspect. The following sub-sections are designed to first provide an illustration of the emergent themes in tabular format, followed by a summarized description of the emerging themes in pros format.

8.5.1 Definition of open data

Group	Definition of open data
Ministry of ICT and ICT Authority (ICTA)	Data that is available and conveniently accessible to those who need it.
Other government agencies	The idea of open data was aimed at creating an open and inclusive government, and allowing easy access to information.
Non-Governmental Organizations (NGOs) and civic tech	Open data is not limited to government data. It includes data from the media, civil society, private sector and crowdsourced data.

Table 8.9: Definition of open data

The aspect of openness in open data was shared across the three groups of interviewees. In addition to this, they also mentioned that it needs to be conveniently accessible to the public, and that the source is not limited to government but should also include data from the media, civil society, private sector and crowdsourced data. This implies that there are different open data portals, though the data can be mashed up using application programming interfaces (APIs).

8.5.2 Benefits of open data

Group	Benefits of open data
Ministry of ICT and ICT Authority (ICTA)	• Open data assists in determining comparative advantage. For instance, Africa receives more than 40 Billion USD in donor aid, but what would result if Africa increased intra-Africa trade by 5% and countries focused on products their environment supports best.
Other government agencies	• The idea of open data was aimed at creating an open and inclusive government, and allowing easy access to information.
Non-Governmental Organizations (NGOs) and civic tech	• Open data has resulted in creation of exciting jobs for young people. • Open data has also resulted in formation of interesting private companies and not-for-profit civil tech organizations. Examples in Kenya include Code4Africa and Data Science Ltd, Open Institute, and LDRI. • The ordinary citizen is now becoming the center of everything, focusing on better governance and economic environment. • Citizens become more empowered with information, which they can use to keep government accountable.

Table 8.10: Benefits of open data

In the case of open government data, the main objective and benefit if the objective is realized is the formation of an open and inclusive government, which is characterized by a more informed citizenry and engagement channels which allow for public participation in governance. Open data including data from government, has resulted in the creation of exciting jobs for the youth following the introduction of new open-data related products/services within existing organizations or formation of new companies which are either private or not-for-profit organizations.

8.5.3 Stakeholders or partners

Group	Stakeholders or partners
Ministry of ICT and ICT Authority (ICTA)	Stakeholders include: ministry of industrialization, ministry of ICT, ministry of planning specifically KNBS and national Treasury, ministry of education, ministry of health, the World Bank, Google, and the Kenya Law Reports for legal advice and data on judiciary.
Non-Governmental Organizations (NGOs) and civic tech	ICT Authority, The World Bank, Africa Media Initiative, Google, Code4Kenya, Kenya Private Sector Association (KEPSA), Civil Society working group, African Centre for Statistics – UN Economic Commission Africa (UN-ECA), Kenya Chiefs of staff at national and county government, Kenya Inter-governmental Technical Relations Committee, Kenya National Audit Office (KENAO), Ministry of Agriculture, Global Open Data for Agriculture and Nutrition (GODAN).
Intra-Governmental Organizations (IGOs)	Private sector, developer community, government, ministry of ICT, IGOs and NGOs.

Table 8.11: Stakeholders or partners

Some of the organizations listed in table 8.11 above were only involved during the launch of KODI in 2011, such as Google and KEPSA. Others have shifted their contribution. For instance, the World Bank Kenya provided technical support on volunteer basis only until the launch, after which they provided financial support through a grant that ended in December 2016. Other organizations such as the ministry of health were publishing open data even before KODI, and have one of the most updated data portals to-date. Others organizations that publish data but not directly linked to KODI are UN-ECA, GODAN, KNBS, and ministry of treasury in relation to the national budget. The ministry of education used to also provide open data freely, but a few datasets such as primary and secondary school ranking are no longer made public. The ministry has several agencies dealing with primary and secondary schools, but the datasets are also not inter-related and are hard to compare or mashup. The main contribution for this disconnect is the fact that they do not use the same identifier for schools, including a unique identifier and a shared school name across the various portals, mainly as a result of poor data entry.

8.5.4 Roles of ICTA data fellows

Group	Roles of data fellows
Ministry of ICT and ICT Authority (ICTA)	• Strengthen the capacity of agencies on data curation.
ICTA data fellows	• Train agency staff on how to curate data. • Curate and publish OGD. • Communicate and coordinate ICTA process on OGD including signing of data release forms by the agency for each dataset.

Table 8.12: Roles of ICTA data fellows

The open data fellows were charged with the responsibility of strengthening the capacity of agencies on data curation. This included building capacity within government agencies on how to curate and publish data. During the first data fellows programme, ICT Authority found the need to design a data release form. This helped in establishing a contract between ICTA and the agency on the frequency of updating a particular dataset, and the contact person at the agency responsible for that dataset. The aim was to ensure timely update of datasets.

8.5.5 Open data curation tools

Group	Open data curation tools
ICTA data fellows	Tabular and Pentaho curation and publication. Current tools are laborious.
Non-Governmental Organizations (NGOs) and civic tech	Modern technology provides access to phenomenally large samples of data, and analyze them to create news within a short timeframe. Some of these technologies, though sophisticated are easy to use by journalists, and do not require experts as was the case previously.

Table 8.13: Open data curation tools

The main tools used by KODI are Tabular and Pentaho, which have been reported by the data fellows as laborious and hard to learn, especially by agency staff, some of whom are not very tech-savvy, as is the case at the National Transport and Safety Authority (NTSA). There is need to identify less sophisticated tools which can be used by agency staff, and also made available to the public, probably as Software-as-a-Service (SaaS) for free or at an affordable rate. The aim would be to increase open data users, and to build a pool of experts, since most will use the same tool and can teach each other formally or informally.

8.5.6 Challenges faced by KODI data fellows

Group	Challenges faced by KODI data fellows
Ministry of ICT and ICT Authority (ICTA)	• Institutions did not know about open data. • The tendency of institutions to hoard information. • They were perceived as auditors leading to resistance. • Communication from ICTA to agencies landed in the wrong hands at the agency causing delay and confusion. This follows the perception by agency heads that communication from ministry of ICT must be directed to the agency director of ICT. This implies that miscommunication was more within the agency than between the agency and ICTA. • The bureaucratic procurement process made it difficult for ICTA to procure eight laptops for the initial group of data fellows who were contracted for six months. • The data fellows did not have designated workstations and internet points at the agency.
ICTA data fellows	• They were perceived as auditors which created resistance. • Some agencies were not aware of OGD. • Some agencies were also not aware about the data fellows, leaving the introductory work to the fellows. • Some agencies had departments in different locations making it strenuous to collect data. • Not all department in an agency cooperated in the OGD curation process. • Time constraint - The data fellows programme ran for only six months, which was not sufficient for them to collect data from all departments within an agency. • Some data was complex. • Some agencies felt that protocol wasn't followed in the communication leading to resistance (bureaucracy/red tape). • Resource constraints. Specifically, internet at the site and computers. Data fellows used their own computers and sometimes internet via personal modems.

Table 8.14: Challenges faced by KODI data fellows

The main challenge revolved around planning of the KODI data fellows programme. This included proper introduction and awareness of the data fellows programme at the target

institutions, timely acquisition of the required tools and resources, and sufficient time for the programme. The outcome of this shortfall was resistance from the agencies, mainly because the right people at the agency were not aware of the data fellows programme and what it sought to achieve, and lack of resources specifically laptops, internet and working space at the agencies for the data fellows. Unfortunately, though the data fellows raised concerns about these shortfalls, there was insufficient time to go through the required bureaucracy hops and acquire the required resources. This led to a total failure to acquire any datasets in some of the agencies, or key departments within some agencies.

8.5.7 Law and policy

Group	Law and policy
Ministry of ICT and ICT Authority (ICTA)	• Initially, there was no open data policy. Data was collected through established networks by ICTA and not through a structured process. • The following now exist; data protection act 2012, freedom to information act 2016 and the open data policy 2014. The data protection act protects the individual and requires certain information to be anonymized prior to publication.
Non-Governmental Organizations (NGOs) and civic tech	• The freedom of information act was not in place until 2016. This could have led to the reluctance by some agencies to release data. • Africa Data Consensus by African Union. It creates a shared understanding of data revolution, and proposes an action plan for its implementation. • Laws needed for the Open County initiative; County Government Act, inter-governments relation Act, the Public Finance Management Act. This legislation suffices, what remains is its implementation.
Other government agencies	• The main laws are the access to information act, and the data protection act.

Table 8.15: Law and policy

KODI was launched as a proof of concept and an open data awareness campaign targeting all the sectors of the economy including the public sector, private sector and non-governmental organizations and inter-governmental organizations. The main drive for this approach was the fear that if it followed the bureaucratic process of having all the required laws, policies and approvals in place, it could take much longer and probably not materialize in the end. As a result, most of the laws and policies were formed after 2011.

Currently, there are two laws, which include; 2012 Data Protection Act - which protects individuals linked to data either because they are the custodians or are implicated for being part of the project, and 2016 Access to Information Act - which allows citizens to request for data

from government provided it does not compromise national security. There are also two policies, which include; Kenya Open Data Policy which is still under development, and the Africa Data Consensus by the African Union - which helps in creating a shared understanding of open data, data revolution and propose an action plan for implementing open data initiatives in Africa. These laws and policies are at national level, following the devolution of the Kenyan government in 2012, there is need to identify laws which can support this initiative at county government level. Some of these laws include; County government act, Inter-Governments relation act, and the public finance management act.

8.5.8 OGD sourcing, curation and publication workflow

Group	OGD Sourcing, Curation and Publication Workflow
Ministry of ICT and ICT Authority (ICTA)	• Identify interesting datasets then request from responsible government agency. • KODI portal is interactive, allowing the public to make data requests, which defines the demand. • Requests are guided by protocol. Bureaucracy slows down the process. Besides the official communication, KODI staff have to call or physically visit an agency for follow-up, otherwise action is delayed. • The first meeting with the agency is on sensitization. This helps determine what datasets are available, and those that they can release. • Once data is identified, the next step is to check for sensitive information which needs to be anonymized or removed. • Once KODI have finalized with a dataset, they present it to the agency for confirmation of correctness. • Following this confirmation, KODI requests the agency to sign a data release form for that dataset.
ICTA data fellows	• Obtain buy-in from departments within an agency to release data. • In some cases, government officials asked for an email with a questionnaire of the data being requested by the data fellows. • Once the data was curated by the data fellows, it was sent to an ICTA staff for verification. • Once its verified, the data fellows present a data release form to the responsible government agency staff for signing. • Once this commitment and agreement is obtained, the data is released to the public. • Data fellow to send dataset to ICTA staff for verification. • Prior to signing of the data release form, allow the agency to verify correctness of the dataset.

Table 8.16: OGD sourcing, curation and publication workflow

The process of sourcing for open data between ICTA staff and their data fellows had a slight variation. ICTA staff are responsible for making the initial contact with an agency, and that includes following the required bureaucracies, which often slow down the process. However, data fellows realised that following the formal channels mainly inform of letters and meetings is not enough to obtain a dataset. In most cases, they had to follow-up on their request by sending reminders and tracking the progress to identify any bottlenecks and resolve. Once data is

obtained, they then check for sensitive information and sanitize it through consultation with the responsible agency staff.

The data fellows needed to consult with the ICTA staff prior to proposing or making changes to a dataset. This approval was also required before they could publish the data, and even then, it is the ICTA staff who made the final step to push a dataset to the public. This process was supported by the data release form, which is a commitment from the agency to ICT Authority to be publishing data at certain intervals which they define, and identification of a contact person within the agency responsible for that dataset, and through whom the follow-up would be made. The data release form also acted as a contract and agreement by the agency to push the dataset to the public space, taking responsibility from ICTA to the agency.

8.5.9 Measures for ensuring timeliness in data release

Group	Measures for ensuring timeliness in data release
Ministry of ICT and ICT Authority (ICTA)	The agency contact person for a particular dataset as defined in the data release form helps in providing updates of a dataset. KODI team don't have to go through the bureaucratic process again to obtain updates.
ICTA data fellows	The current measure applies to existing datasets. The data release form includes commitment by the agency to release updates at specific times. ICTA staff use this to send reminders.
Other government agencies	Data at treasury is released periodically based on the budget cycle. ICT Authority follows up through calls and emails to get a copy of this data.

Table 8.17: Measures for ensuring timeliness in data release

The main challenge when sourcing for datasets from an agency is the initial contact between the ICTA and the agency. A few bureaucratic steps are required, which involves communication from the Chief Executive Officer for ICTA, or even the Principal Secretary for the Ministry of Information and Communication to an officer of similar ranks in the other agency. Once a go ahead is given at this step, the next step is to identify the departments responsible for the datasets that are of interest to ICTA. Further introduction and agreement to release data is made at this level.

Once these two steps are done, the next step is to identify the responsible staff within the agency for each of the required datasets. This is done with the help of an open data champion within an agency, someone ICTA identifies at the very early stages of the process. ICTA then requests staff within the agency to sign the data release form indicating the frequency the dataset will be updated and submitted to ICTA. This makes it easier for ICTA to determine which datasets to

follow-up on at any given time. It also helps them respond more accurately to public requests on data.

8.5.10 Challenges in obtaining open government data

Group	Challenges in obtaining open government data
Ministry of ICT and ICT Authority (ICTA)	Some agencies do not publish raw data, instead they provide aggregations which are limiting. Sometimes data is shared in pdf format.
Other government agencies	Though treasury publishes the budget on their website, and also sell as books at the government press, there is no clear guideline/policy on how to publish government data to the public. They release to ICT Authority based on trust and also because it is a government agency, otherwise bureaucracy would set in. Treasury can publish more if there is a clear guideline/policy. This policy should allow citizens to assist in monitoring government projects. Legacy data which was stored in print format. When reconstructing data, there are cases where some of the files got misplaced or lost, or contains gaps and inconsistencies.
Intra-Governmental Organizations (IGOs)	Systemic issue. There are many agencies collecting data in different formats. In addition, there are no underlying agreements requiring the agencies to collate the data in one repository or publish to the public.

Table 8.18: Challenges in obtaining open government data

Open data is preferred when it is in its raw format and enriched with metadata, which helps in describing the data. It also needs to be published in machine-readable and open format. Unfortunately, some agencies provide aggregations, which miss important detail that would help in deriving new meanings. Unfortunately, some of the data is released using proprietary formats, mainly pdf, which requires more work in converting it in open data formats which at times compromises on the correctness of the data. Another challenge related to formats is the fact that agencies have their own way of formatting and presenting data, which at times requires more work due to lack of standardization.

8.5.11 Sensitive information and correctness

Group	Sensitive information and correctness
Ministry of ICT and ICT Authority (ICTA)	Sometimes when two agencies publish the same dataset, there are anomalies. For instance, the number of births data published by civil registrations board differs from that of Kenya national bureau of statistics. Some of the data does not contain metadata, which would help in explaining possible anomalies or discrepancies.
ICTA data fellows	Responsibility rests with the agency providing the data. For instance, name field in NTSA drunken driving data. Data release form was used to obtain approval from agencies prior to publication. Main concerns “we are not used to sharing this data, why now? Who is going to use this data? How will this affect us?”
Non-Governmental Organizations	Rule of thumb: Treat all data suspect until verification.

(NGOs) and civic tech	Some innovative ideas are shut down. For instance, an effort to compare donors and their contributions to Kenya was rejected. Open data is not so open. For instance, they publish provisional budget at national and county level, and never the final budget that is approved by parliament. Contentious data is concealed by government. For instance, the name of contractors in controversial projects is removed from the dataset. Checking for correctness is a laborious and often manual process following data inconsistencies. For instance, in a dataset, Nairobi could be recorded in three or more formats; NBI, NBO or Nairobi. Some government agencies do not have an established system of managing data, which makes it hard to analyze. For instance, it's impossible to use KNEC data to compare exam results year on year since schools are not coded and their names may be entered differently each year.
Other government agencies	KNBS perceives that ICTA publishes anything. Treasury has its own internal quality control mechanism, and in the case of budget data, additional verification is done by parliament.
Intra-Governmental Organizations (IGOs)	There was an idea for a quality assurance mechanism/team. The design was that requests for new datasets would be sent to a technical committee who would decide on whether the data can be published, check whether it contains confidential information, anonymize the data, and determine the frequency of data updates.

Table 8.19: Sensitive information and correctness

Correctness is at times difficult to determine especially in situations where two agencies produce varying data on the same phenomenon. Another challenge on correctness is lack of metadata in most of the datasets, this could help clarify discrepancies and also provide further insight into the data. Also, government agencies conceal contentious data, especially data on government expenditure and population. As a result of these challenges, ICTA decided to place the responsibility of the data on the agency technical committee, since they do not have capacity to assess, determine correctness and sensitivity, and take responsibility for each dataset on the portal. Prior to publication, ICTA requests the agency to verify correctness of a dataset, and take ownership through the data release form.

8.5.12 Capacity, training and technical support

Group	Capacity, training and technical support
Ministry of ICT and ICT Authority (ICTA)	There is a period ICTA did not have staff for KODI, but there is now a process to procure and get new staff. It took time for the KODI job structure to be approved. There was need to explain the proposed roles and the linked projects. ICTA funded the civil registry department to digitize all births and deaths, and publish some of it as open data. KODI team now has more capacity, and the current team consists of a project manager, project coordinator, statistician charged with data acquisition, curation and publishing, data analyst charged with analyzing, visualizing and storytelling through KODI blog blog.opendata.go.ke, an open data and geographical information specialist, and the data fellows who provide technical support to government agencies.
ICTA data fellows	Induction training was short but ICTA staff provided additional technical support via email and voice calls.
Non-Governmental Organizations (NGOs) and civic tech	Conduct four days to one-week data bootcamps or data master classes. These are open to the public and attendance includes people from government and private sector. At government agency level, open data is perceived to add an extra layer of work, and at times duplication of effort, in cases where old ways of processing data are retained in parallel to the open data curation process. There is need for more automated systems that eliminate duplication of effort and increase correctness of data. Some government agencies are still using manual systems to record their transactions and data. This implies a challenge in use of technology and resistance to change from the culture of manual to automated systems. For instance, the judiciary.
Other government agencies	There is need for government agencies to receive training on open data. Treasury department benefited from a training on geographical information skills, which they found very useful in monitoring expenditure in the various constituencies. There is a disconnect between ICT department and other departments in some agencies. They do not provide the needed advisory and support in the open data curation process. Most agencies wanted the data fellows to be deployed for a longer period to help in the curation process.
Intra-Governmental Organizations (IGOs)	The world bank Kenya office staff were working on the KODI initiative outside their formal duties and responsibilities. This was motivated by their observation of a real demand for data and the need for immediate response in terms of capacity and funding.

Table 8.20: Capacity, training and technical support

Following the ad-hoc implementation strategy, there were no job structures specific to open data at ICT Authority to support KODI. These job structures took a lengthy period to be formed and approved. However, they are now approved and capacity at KODI has increased. The KODI team at ICTA has been empowering different agencies through technical training workshops. They have also supported several bootcamps, which are aimed at creating awareness and interest among the developer community and the citizens on the affordances of open data. Unfortunately, capacity has not increased at agency level, and open data is perceived as an additional piece of

work, which is not defined in the performance contract. This creates conflict and government staff at times are forced to focus on their core mandate and only deal with open data requests during off-peak. This was the same scenario for World Bank, which has since withdrawn technical support on KODI.

8.5.13 KODI financial support

Group	KODI financial support
Ministry of ICT and ICT Authority (ICTA)	KODI is currently funded by the World Bank. Ideally, it should be funded by government. There is no timeline on when government will start supporting the initiative financially. Reporting to World Bank is based on key performance indicators. The KODI policy, once formulated and approved, could help in securing budget from government.
Non-Governmental Organizations (NGOs) and civic tech	Funding is tied to donor imperatives. Coupled with this are priorities by government and civil society. An alignment of this determines the sustainability of KODI.
Intra-Governmental Organizations (IGOs)	World Bank supported the initiative through the Kenya Transparency Information and Communication Project. An ongoing loan of $5 million was granted to ICT Authority to support open data institutionalization. Money is not the issue to the success of KODI. Cost of data collection, paying for the KODI portal software license (Socrata) and sending data is not expensive. What is critical is leadership, not more money.

Table 8.21: KODI financial support

KODI has been supported by donor funding, almost entirely by the World Bank. ICT Authority is required to report on the project to the World Bank based on the agreed key performance indicators (KPIs). This implies that there could be initiatives that ICTA may want to implement, but cannot since they are not part of the KPIs. This type of funding is not sustainable as it depends on the donor imperatives which may change unexpectedly. Therefore, there is need for ICTA to secure budget from government to ensure sustainability and growth of the initiative. One advantage is that most of the costs related to KODI are known, since the project has been running for a few years, and this may help in negotiating the right budget with government. The structures in place, including job structures and open data policy are also necessary in securing budget.

8.5.14 Culture

Group	Culture
Ministry of ICT and ICT Authority (ICTA)	Government initiatives at times lack ownership. There is need for someone to take responsibility for KODI. Some agencies are used to hoarding data. Almost 60 mobile applications were developed at the start of KODI. However, most of them became useless since data was not updated after agencies withheld/delayed release. There is need for change of mindset on certain processes and information. Data needs to be captured electronically at the first point of entry, for instance when a patient walks into a hospital. This should be linked to an integrated database that contains additional information about the patient. There is need for attention for detail to ensure accuracy/correctness.
ICTA data fellows	Some agencies resisted because they were not used to submit data to any outsiders.
Non-Governmental Organizations (NGOs) and civic tech	Technophobia: Most of those in government offices are from the previous generation, and at times do not understand the potential of open data. Bureaucracy is a big challenge to KODI based on the real-time nature of some datasets, and ever changing user requirements.
Other government agencies	Bureaucracy limits engagement especially where government agencies cannot communicate due to differences. For instance, ICT Authority and KNBS should complement each other and not compete.

Table 8.22: Culture

There are five main aspects that define the culture surrounding KODI including ownership, bureaucracy, secrecy, legacy systems and technophobia. KODI faced ownership challenges but measures such as data release form, approval of job structures at KODI, and the data fellows programme have helped ensure that government agencies, including ICTA commit part of their time and resources to open data. Bureaucracy is still a challenge, especially for ICTA when they are dealing with other government agencies when requesting for data, especially for the first time. This also applies to acquisition of resources, as was the case in the first data fellows programme, where data fellows were forced to use their own laptops since ICTA could not buy the laptops for them. Another challenge is funding, since they are yet to secure funding from government.

Secrecy is still evident within government, especially when it comes to contentious data such as expenditure data at national or constituency level. Related to this is the perception of government staff that data belongs to them, and that releasing it as open data makes them loose control, power and possible revenue stream. A case in point is KNBS which sold some of the statistics in print format, and would not release it in digital formats to reduce the ease of distribution. Legacy systems affect the curation process of open data. Some of the ways used to capture and store data

require additional work to convert in digital format. Most of the data entry work and storage in government is manual. Greater efficiency can be achieved by automating these tasks and linking the various systems and databases together. Government still faces technophobia. For instance during the preparation for the launch of KODI in 2011, some advisors to the President perceived KODI as another wikileaks. Efforts by ICTA to educate various agencies and senior government officials on open data and its capabilities have helped in eliminating this fear and misconception that KODI has come to threaten their jobs or destabilize the current government.

8.5.15 KODI Innovation

Group	KODI Innovation
Ministry of ICT and ICT Authority (ICTA)	Significant number blog posts on blog.opendata.go.ke by KODI team. This involves identifying interesting patterns in a dataset and writing a story about it. For instance, there was an observation from National Transport and Safety Authority (NTSA) that most accidents in Nairobi happen at 6pm, a time when the city is clogged with traffic. In addition, Nairobi contributes 50-60% of accidents nationwide. KODI also produces periodical newsletters which helps in reporting on current undertakings and upcoming initiatives. Application development. KODI has supported several initiatives like the Budget App which demonstrates open data usefulness, and ease of use by people with no statistical background. Newsplex by Nation Media Group. This is a newspaper column with a dedicated team of data journalists. They print inforgraphics of certain datasets on the daily nation newspaper weekly. There is need to find ways to discuss datasets on radio or television.
Non-Governmental Organizations (NGOs) and civic tech	Open Duka by Open Institute. Allows citizens to understand the ownership structure of the various entities in Kenya. Mzalendo.com, a non-partisan project that keeps an eye on parliament and publishes findings on the blog. They sought to obtain data on parliament affairs, but gave up relying on KODI portal as data was never up-to date. They now scrap and clean data from parliament website and constituency websites. kenya.opencounty.org implemented by Open Institute for the Kenya Council of Governors. Provides visualized information of a county's accomplishments.
Other government agencies	Chief Kariuki is developing a system that will capture all the people in his constituency, which helps in tracking peoples movement. The constituency will be divided into clusters, and a leader nominated from the dwellers to help in monitoring their members movement and noting new comers. eProMIS, Electronic Project Monitoring Information Systems by Treasury department. It helps in monitoring projects across government. It makes use of geocodes based on google platform to identify exact location of each project and monitor remotely.
Intra-Governmental Organizations (IGOs)	The KODI platform has the highest traffic in Africa compared to other open data portals. There was steady traffic to the portal even two or three years after the launch when there wasn't much updates on data. Agencies can now have more views on their data by either publishing there or requesting links to their repository to be embedded on KODI platform.

Table 8.23: KODI Innovation

Innovation includes open data blogs, periodical newsletters, open data mobile applications, open data infographics on mainstream newspapers, and startup companies - most of which have been civic technology companies. Examples to these innovations include, significant number blog posts on blog.opendata.go.ke by KODI, periodical newsletters by KODI, budget app, Newsplex weekly newspaper publication by the Nation Media Group (NMG), Open Duka by Open Institute that allows people to understand the ownership of various entities in Kenya, Mzalendo.com a non-partisan publication that assists citizens keep an eye on parliament by publishing parliament affairs, opencounty.org that provides a visualization of accomplishments by county governments, and the KODI platform which is the official portal for publishing government open data. There is also eProMIS, which was a separate initiative to KODI, but has a lot of potential if made public, as it contains data that can help in monitoring and evaluating government projects across the country. Besides these achievements, there is need to ensure that open data is reliable and released timeously, and that requests for new data sets are addressed in time. This will create interest from the developer community, bloggers, media houses and other stakeholders who are interested in creating new meanings and products from open data.

8.5.16 Demand

Group	Demand
Ministry of ICT and ICT Authority (ICTA)	Demand is not high. Not overwhelming. Website optimization is key to data availability. Some agencies lack this on their portals. KODI portal has a data request mechanism. Priority is given to datasets with highest requests. Citizens are updated on their request status. There is no mechanism to determine who is using the data and for what. It is hard to get some datasets, for instance data on primary and secondary exam results. Some users need the data, but do not understand the technical terms that define the datasets. For instance, Mobile Termination Rate (MTR) for client initiated call transactions between mobile phone companies. There is a push from the developer community and the intermediaries who want to create products out of open data. For instance BlackOrwa.com and http://www.nation.co.ke/newsplex are blogposts that create stories based on open data.
ICTA data fellows	Visualizations are the main selling point for KODI portal. Statistics including those by KNBS are complex.
Non-Governmental Organizations (NGOs) and civic tech	One of the challenges is that government at times does not publish raw data, which forces a certain narrative, and limits the user from getting the whole perspective. At times government delays in publishing data as they try to clean it up whereas users prefer raw data. Literacy level, including technological competence affects citizen demand for data, requiring more infomediaries to provide diverse insights.

	Political enablers. There is need to identify ways to make citizens endeared to open data. Developer buy-in during the launch was driven by the perception that data would reform their businesses. National Fibre Optic Backbone (NOFBI) project was a motivation for ICT Authority to embark on open data, to justify the bandwidth need to parliament. Mzalendo.org stopped looking at KODI for data as most datasets are not up-to-date and requests for datasets did not yield much.
Other government agencies	There is a perception that the public does not know about KODI. KODI has helped in visualizing government data in a way the public can consume. Government agencies are able to find new meaning from visualisations, which also help them communicate certain information more effectively. Open data is promoting engagement between government agencies and departments as they find new insights and ideas from the data visualizations, as is the case at National Transport and Safety Authority.
Intra-Governmental Organizations (IGOs)	There was significant demand for data from the ICT community. This was one of the major driving points for the KODI agenda. There was also some push from civil society. If the demand side is fragmented, government will not invest in open data. There is need for a visionary person to lead the initiative. Government has what it needs to implement KODI, but there must be significant demand from the citizens, not donors or foreigners. There is need to use donors strategically. KODI was implementing a feedback mechanism and a feature to request new datasets. There is need to incentivize users to provide feedback on accessed datasets.

Table 8.24: Demand

Demand for open data is high but not overwhelming to require additional capacity. This could be explained by the loss of momentum between 2012 and 2014. The main selling point for the open data portals has been visualizations as they are easier to comprehend compared to raw data. A few challenges exist including lack of website optimization by some government agencies affecting data access/availability, failure of government to produce some much needed datasets such as the primary and secondary school exam results and school rankings (data hoarding), failure of government to publish raw data and opting for processed data which forces a certain narrative following the limitations of the data, open data publication delays, and low literacy levels and access to computers and affordable internet.

8.5.17 Open data champion and buy-in from stakeholders

Group	Open data champion and buy-in from stakeholders
Ministry of ICT and ICT Authority (ICTA)	A champion is necessary to advocate for KODI. Government has many concurrent and competing projects, creating a need for someone to push KODI agenda. The IT cabinet secretary (CS) needs to be a champion to help in influencing other cabinet secretaries, who will then communicate with the agencies and directors in their ministry. There is need for buy-in right at the top. To avoid data hugging, there is need to demonstrate the demand for data.
ICTA data fellows	Capacity building motivated agencies – Give (skills) and Take (data) relationship. Preparedness and proper communication by ICTA to agencies from the onset.
Non-Governmental Organizations (NGOs) and civic tech	Devolution of champions. We need champions at executive, cabinet, ministry and county levels. For instance, David Cameron - the Prime minister of the UK was the open data champion. The true sign of buy-in from government will be when they put substantial and sustainable funding to KODI. At the moment, it's all funded by the World Bank. Ministry of Agriculture and Health data projects are also mainly funded by donors.
Other government agencies	Government needs to issue a directive inform of a circular to all agencies regarding KODI. KODI is perceived to be a donor driven initiative.
Intra-Governmental Organizations (IGOs)	The Permanent Secretary was the main catalyst. He was responsive to the fervent requests for open data from the vibrant ICT community. Political economy is key to buy-in from government. This is the nexus between government, business and politics. It's very thick and hard to penetrate. The same people who run government especially at policy level are the same ones in the private sector and in politics. They allow people who will not unravel those relationships. Government will not spend money on open data because it's not what they want to do. An open data policy and a supportive legislative structure is critical. You need a champion who challenges the political economy and demanding that open data policy be enforced. There is need to find ways to scale initiatives out of KODI to increase buy-in. The World Bank was involved with the aim of strengthening systems that link government to its citizens. This also includes promoting accountability by government to its citizens.

Table 8.25: Open data champion and buy-in from stakeholders

One of the ways to ensure buy-in from government is to issue a directive to all government agencies about KODI. This is a bureaucratic process, but remains a requirement that is yet to be implemented for KODI. This implies that political economy is crititical in securing buy-in from government. As described by the IGOs, political economy is the nexus between government, business and politics, and is described as thick and hard to penetrate since the people running government affairs are the same ones in the private sector and politics. Anything that threatens to unravel that relationship is opposed vehemently. This implies that if open data is perceived as a threat by this class of people, there is need to have an open data champion who is able to

influence them otherwise by creating awareness on the affordances of open data, demystify possible fears, and demonstrate demand for open data.

This implies that an effective open data champion needs a strong political economy. There is also need to have champions at cabinet and county levels, who also possess a strong political economy and understanding of the potential and complexities of open data. Besides having champions with a strong political economy, sustainable buy-in also requires an open data policy, supportive legislative structure, financial support from government and not donors, and open data curation capacity building within government agencies.

8.5.18 Awareness and stakeholder engagement

Group	Awareness and stakeholder engagement
Ministry of ICT and ICT Authority (ICTA)	KODI has been conducting numerous trainings, coding bootcamps, and workshops with government institutions on OGD. KODI has also organized coding competitions with monetary incentives to promote innovative use of open data. There has been an attempt by KODI to work with the private sector, but they don't seem to understand the benefits of OGD. There has been sensitization with the media, focusing more on data based journalism. Sleeping Giant Media, a UK company has been contracted to create awareness and stakeholder engagement. This contract is linked to the data fellows programme. KODI co-sponsors open data events organized by civil society. KODI also planned on having online engagement platform which was motivated by lower costs compared to face to face organized meetings. Some of the founders of KODI have been publishing on national newspapers, blogs and presenting in public spaces with the aim of promoting the open data cause.
Non-Governmental Organizations (NGOs) and civic tech	KODI events are not properly advertised in some instances, making them less open. At times the events are focused on a selected group of people. For instance, KODI re-launch ceremony. After the launch, the code for Kenya fellowship programme was formed. This involved four fellows who were embedded for six months in civil society and media organizations with the aim of creating awareness, transferring skills and demonstrating usefulness of open data. KODI initiative is not operating in silos but in isolation. They've defined their own objectives and are struggling to sell those objectives to government, civil society and private sector. Change of narrative. The narrative should be what data do people need instead of how do people use the data we publish. Open data policy formulation has not been consultative. It has not been subjected to citizen participation and stakeholder engagement. The process has involved the same set of organizations over the years. This reduces the opportunity for additional buy-in. For instance, Parliament through the parliamentary committee has been lacking. Political goodwill is critical. It should not be about the portal but the transformative action it

	brings to improve people's lives. KODI launch was attended by 3000 people including the President, and segments of the launch were broadcast on national television. There was need for a ministerial committee to lobby people from the various ministries in government to embrace open data. There was perceived disconnect at the ministry of ICT with the eGovernment division regarding KODI. This is a critical function and affected efforts to sell the project to other government agencies. The ministry of health has one of the most comprehensive and updated datasets. They started publishing open data before KODI began in 2011. Public participation is hard to achieve since necessary data is either not available or not upto date. For instance data on budget preparation, budget implementation, and government expenditure.
Other government agencies	Chief Kariuki uses short message service broadcasting to communicate with his constituency, and invite them for public barraza's including discussion on constituency funds and what projects to prioritize. This supports citizen engagement in governance and promotes accountability and transparency. People in the grassroots don't know about data or technology but they know about public forums and they attend. The hype should not be the portal but conversion to openness.
Intra-Governmental Organizations (IGOs)	The World Bank held the first meetings with the Permanent Secretary of ICT pushing the open data agenda. They involved various experts at World Bank country office and the international office, with the aim of explaining what was happening with open data globally, and how it could support some of the goals the Permanent Secretary and the ICT board sought to achieve. The ICT permanent secretary organized an open data awareness meeting with key permanent secretaries and directors of agencies that are key holders of data including Kenya National Bureau of Statistics. The World Bank played a supportive role, with attendance from the country director and a few staff members.

Table 8.26: Awareness and stakeholder engagement

In 2011, awareness and stakeholder engagement was mainly driven by the Permanent Secretary of ICT, the World Bank, and the Kenya ICT Board (now ICTA). Challenges include a disconnect between some government agencies as a result of KODI, lack of timely update of some datasets affecting public participation and interest, and hoarding of data which affects innovation and makes it hard to defend the case for open data for lack of interesting products/stories.

Recommendations include, change of narrative, make the open data policy development process more consultative, include new stakeholders into KODI, form a ministerial committee that will assist in lobbying for open data at cabinet level, identify ways to communicate findings from open data to the grassroots - Chief Kariuki's strategy short message broadcasting and calls for public forums could be an option, change focus from the portal to the transformative change that

results from open data, additional sensitization on data-driven journalism, and additional incentive-based bootcamps and coding competitions.

8.5.19 Conceptualization and implementation of KODI

Group	Conceptualization and implementation of KODI
Ministry of ICT and ICT Authority (ICTA)	Idea was conceptualized in 2009 but there was no momentum until 2011. Interest came from the World Bank and several government ministries but spearheaded by the ministry of ICT under the then Principal Secretary Dr. Bitange Ndemo who was a very effective champion for KODI. The initial KODI platform was put together in exactly three weeks in 2011 after the President commissioned the initiative and promised to attend the launch at the end of the three weeks. In 2011, KODI was run by a team of volunteers mainly from civil (tech) society, academia, media and the developer community. The rest of the team came from government and the World Bank. There was resistance from the President's advisers on the risk of exposing government data following the Wikileaks experience in other countries. Data access involves politics. Open Data requires a champion with political capital. Despite being turned down by some agencies, KODI used alternative means to obtain data, including scrapping websites for data. Despite the lack of supportive legistlation, there were successful initiatives on open data including bootcamps, workshops and data journalism courses. There is duplication of work between KNBS and KODI. This makes it difficult for KODI to work with KNBS. For instance, visualization project contracted by KNBS to Strathmore University.
Non-Governmental Organizations (NGOs) and civic tech	Lack of policies, structures and proper awareness explains why KODI came to a halt one year after the launch in 2011. Open data was not pushed as part of the devolution agenda. The 47 County governments were formed without consideration of open data. Besides the World Bank, most other NGOs do not publish the data they collect. This data doesn't need to be published on KODI platform, but needs to be made public digitally.
Other government agencies	KNBS faces the challenge of mandate, capacity and funding in relation to KODI. KNBS suggests that ICTA should spearhead opening of data and not data collection since KNBS is the one mandated to collect, compile and publish. There was poor introduction of the data fellows in the various agencies. This resulted in resistance by agency departments to cooperate.
Intra-Governmental Organizations (IGOs)	Discussions between World Bank and the ICT Permanent Secretary (PS) began in 2010. The PS was motivated by the potential for open data to foster development. There was also a fervent request from the ICT private sector community for open data. The PS was responsive to these requests. The bank provided technical support and initially helped in curating some of the data. Endorsement by the President was critical to the success of the project. Traffic on KNBS portal is low, and the interface is not very user friendly and it's not easy to download some of the datasets. ICT Authority at the beginning placed two consultants at KNBS to assist in the open data initiative with the aim of removing the hesitation on open data. There has been contention between ICT Authority and KNBS on who will get credit for open data. The KODI programme coordinator is now running her own data science company, which results in

	conflict of interest. This needs to be a fulltime role and not part-time. After the 2013 general elections, there was a backlash on open data following the International Criminal Court indictment of the elected Head of State and his deputy. Open data was perceived as a western plot trying to undermine democracy in Kenya. ICT Authority now have the financial resources, if they are keen on KODI, they need to appoint a serious and dedicated team. The project coordinator needs to be someone senior with a strong political capital. The change of ICT Permanent secretary was a big loss to KODI. The initiative needs to focus on comparative metrics between different county governments in Kenya following devolution. Donors have been focusing on push inputs, this should change to pull technology where the best innovations are rewarded, including government agencies with the most comprehensive and updated datasets. There was an idea to have a national steering committee to oversee request and publication of data, and enforce laws and standards like data protection act. No updates or new datasets were uploaded to the KODI platform in 2013. Key government and data producers did not find KODI useful. There is need to create buy-in to KODI, otherwise it will continue to underperform. The number of datasets remained as 439 for a long period. Tracy Lane from World Bank was responsible for budget data, no updates have been made since she left Kenya country office. There is need for line ministries to identify someone who will be responsible to curate and publish data through KODI portal. Open data needs to be included in the Key Performance Contracts of ministry and agency leaders. You cannot do reforms without resources and voluntary work is short lived, which creates the need for dedicated resources.

Table 8.27: Conceptualization and implementation of KODI

Momentum for KODI began in 2011 when Dr. Bitange Ndemo was the Principal Secretary for the Ministry of Information and Communication. He was the main champion for open data, and following consent from the minister for Information and Communication, sought approval from the head of state to launch KODI. Other stakeholders that were keen to see the implementation of KODI were the World Bank, civil society, civic tech society, academia, media and the developer community. The initiative faced some resistance from the presidency, as it was perceived to be similar to WikiLeaks and its main aim was to destabilize the country politically. The KODI team led by the Principal Secretary helped in demystifying this myth on open data, resulting in approval of the initiative by the president. The presidency gave the KODI team three weeks to launch the initiative, which was too tight to setup all the necessary structures, including policies, job structures, funding from government, and buy-in from government agencies.

Some of the challenges that face(d) KODI include; lack of buy-in from some agencies, which resulted in innovative ways to obtain data including scrapping agency websites for data, lack of supportive legislation and policies, open data was not made part of the country devolution agenda following the enactment of a new constitution, most NGOs did not release their data, conflict of mandate between government agencies on open data for instance between KODI and KNBS, lack of a full-time KODI project coordinator with a strong political economy, political backlash on open data which was perceived as a western phenomenon after the head of state and the deputy were indicted by the International Criminal Court for crimes against humanity, and lack of updates on the portal one year after the launch until 2014. Despite these challenges, the following were achieved; successful bootcamps, workshops, data journalism courses, and publication of more than 400 datasets on KODI. There would have been more achievements if the initiative did not stall in in 2012 and 2013.

8.6 Corroborating and legitimating code themes

This is the last stage and involves clustering of themes identified previously from the coded text. This helps in confirming that there is no fabricated evidence that may have been introduced during data interpretation. This implies that the clustered themes should be representative of the initial data analysis and assigned codes. This stage results in an explanatory framework which results from analyzing the codes and themes that emerged from the text. Clustering of themes involves assigning unique phrases that describe the meaning of a particular theme (Fereday & Muir-Cochrane, 2006). This process resulted in the identification of five themes, which include: demand; law and policy; planning, coordination and capacity building; awareness, buy-in and ownership; and advocacy.

8.6.1 Demand

Four clustered themes were identified from the first order theme, which describe the necessary enablers and requirements sustainable open data demand. The table below provides the details of these findings.

First Order Theme	Clustered Theme	Second Order Theme
Open data makes the ordinary citizen to be the center of governance and economic development. The affordances include job creation, and generation of information that helps determine comparative advantage. There is need to create demand (pull) for data in order to realize value. Most of the effort on KODI specifically has been focused on the supply (push) of data. There is little demand from citizens, media community in relation to data driven journalism, and the software development community who should be creating innovative software solutions that generate new insights from open data, and provide an engagement platform where citizens can debate on certain developments. Demand for open data requires government to create engagement channels where citizens can interact with the custodians on certain datasets. Demand for data has been affected by lack of website optimization which improves search results, failure of government in some instances to publish raw data, and failure by government and NGOs in some cases to publish data on time and respond to citizen demands for new datasets.	Theme 1: Open data is effective if it enables the citizen to be the center of governance and economic development. Theme 2: Value of open data requires real demand for open data. Supply without demand is not sufficient. Theme 3: Demand for open data requires citizen engagement channels. Theme 4: Demand for data is dependent on ease of access through search engine optimization (SEO).	Demand When SEO enables citizens to access data with ease and engagement channels allow them to request for additional data or changes on existing datasets.

Table 8.28: Demand

8.6.2 Law and policy

Two clustered themes were identified from the first order theme, which describe the role of law and policy in supporting openness, and the need to eliminate political interference in implementing open data initiatives. Law and policy emerged as the second order theme from this cluster. Table 29 below provides the details of these findings.

First Order Theme	Clustered Theme	Second Order Theme
Supportive laws and policies are critical in the success of open data initiatives. Government needs proper policies and structures to create autonomy of the initiative guarding it from political interference. Open data laws and policies need to be developed during the implementation phase in order to legitimize the initiative, reduce resistance, and create awareness as the laws and policies are adopted by the various agencies within government.	Theme 1: Open data requires laws and policies that support openness. Theme 2: Open data requires autonomy from political interference. This is achieved through law and policy.	Law and policy Success of open data is pegged on the effectiveness of existing laws and policy and the autonomy from political interference.

Table 8.29: Law and policy

8.6.3 Planning, coordination and capacity building

Seven clustered themes were identified from the first order themes. Their focus was on the impact of planning, coordination and capacity building in realizing success in open data initiatives. Two major challenges were identified in government organizations namely bureaucracy and technophobia. As a counter-measure, there is need to plan adequately and

formulate strategies that will assist in overcoming these challenges and increase efficiency in service delivery. Table 30 below provides the details of these findings.

First Order Theme	Clustered Theme	Second Order Theme
Open data initiatives require adequate planning and project management to ensure that possible bottlenecks are addressed beforehand and resources are acquired in time and fully utilized. There is duplication of effort from certain government agencies, for instance ICT Authority and KNBS. There is need for structures and policies that promote coordination between agencies. Technophobia or the fear of technology within government. Some of the officials in government are not tech-savvy and find technology as a threat to their existence and fail to understand the potential benefits of open data. Bureaucracy affects development of KODI. Government agencies require training on open data tools. This should form part of the awareness effort as the agency staff gain an understanding of the capabilities of available open data tools and understand the potential use of their data once it's made public. Data fellows were mistaken for auditors which created resistance. There is need for ICTA to ensure proper communication with the agency prior to deploying data fellows. Quality assurance and correctness of open data is vested on the agency providing the data, with the help of ICTA. All data is treated suspect until verification by the agency prior to publication. Verification is a laborious and often manual process, which is mainly as a result of inconsistencies during data capture and lack of standardization. Agencies may need to form a quality assurance team for data, since the one generating the data may fail to identify some inconsistencies.	Theme 1: Open data initiatives require adequate planning and project management. Theme 2: Open data requires proper coordination to eliminate duplication of effort among government agencies. Theme 3: Technophobia is detrimental to the success of open data initiatives. There is need to devise strategies on how to mitigate this. Theme 4: Bureaucracy affects the progress of open data initiatives and demand in cases where this results in delay to release requested datasets or changes on existing data. There is need to plan on how to address this. KODI for instance devised the data release form to address this. Theme 5: Open data is a give and take process. Government agencies can provide data, but require training and capacity building to achieve this. Theme 6: Capacity building requires proper communication with the target agency to avoid misconception and resistance. Theme 7: Correctness of open data depends on coordination with the agency in charge of the data, and their ability to check for correctness.	Planning, coordination and capacity building Open data challenges include duplication of effort, technophobia, bureaucracy, and skills on how to curate open data. To address this, there is need for proper planning and coordination to ensure that supportive laws and policies are implemented, and that the required skills are provided. In addition, there is need to train, plan and coordinate on quality assurance to ensure correctness and reliability of data.

Table 8.30: Planning, coordination and capacity building

8.6.4 Awareness, buy-in and ownership

Awareness, buy-in and ownership were identified as the second order theme from a collection of nine clustered themes. The clustered themes identified the role of awareness, buy-in and ownership in realizing success of open data initiatives. The key enablers were identified as political goodwill identified by the level of government investment to open data initiatives, and public advertising of open data events to create awareness. The main challenge was political economy which is defined as the nexus between government, business and politics. This results in resistance when the well-being of the drivers of the economy is threatened by open data. There is need to identify open data champions who have strong political economy to challenge or influence the players in this organization. The table below provides the details of these findings.

First Order Theme	Clustered Theme	Second Order Theme
Awareness and buy-in from various stakeholders in government especially the executive branch of national government is critical to the success of open data initiatives. If the initiative is driven by government, there could be greater buy-in compared to when it's driven by donors, which is often challenged politically through resistance from the political class based on the argument that the country's sovereignty is being compromised when asked to report to a foreign body. Political goodwill is critical to the success of open data initiatives. The focus should not be on the portal or technology, but the transformative action it brings to improve people's lives. Public participation on the KODI has been hard to achieve since needed data is either not available or not up-to-date. This is despite the public making requests. This affects buy-in from the public and developer community. There is need to make all open data events public by creating awareness and encouraging participation from the general public. There is need to create awareness with the media, in order to promote more data-driven journalism. The media can benefit from existing tools which help them generate information from huge datasets which would have previously taken a lot of time, or required an expert to generate. Buy-in from private sector is key in ensuring sustainability of the initiative in relation to retaining expertise through revenue generated from commercialization of open data solutions. The true sign of buy-in from government will be when they make a substantial financial investment to KODI by allocating adequate budget and endorsing structures that will guarantee continuity of the initiative. Political economy is key to buy-in from government. This is the nexus between government, business and politics. It's very thick and hard to penetrate. The same people who run government especially at policy level are the same ones in the private sector and in politics. They allow people who will not unravel those relationships. Open data initiatives should be funded by government and not donors since donor imperatives may shift leading to collapse of the initiative. Reforms from open data cannot take place without resources and voluntary work including donor funding is short-lived. There is need for ownership from government and private sector who will dedicate resources to the initiative. The private sector perceive that open data will result in the loss of competitive advantage.	Theme 1: Awareness and buy-in is critical in the success of open data initiatives. Theme 2: Government is the most effective stakeholder in creating awareness and buy-in as it has the greatest political capital. Theme 3: Political goodwill is critical to the success of open data initiatives. Theme 4: Open data events should be advertised to the public to create awareness and ensure inclusivity. Theme 5: Data driven journalism is a major application point of open data, creating demand of open data. Theme 6: Private sector involvement in open data ensures sustainability of open data initiatives, through the revenue generated from products or services based on open data. Theme 7: Government buy-in is measured by the amount of financial and human capital invested on open data initiatives. Theme 8: Open data initiative cannot be sustained by donor funding as this is dependent on donor imperatives. Theme 9: Political economy affects government and private sector buy-in. This is because the nexus between government, business and politics is very thick and hard to penetrate. Some of the actors are associated or are active actors in more than one field either directly or indirectly through proxies. Buy-in will be based on how open data impacts on this relationship.	Awareness, buy-in and ownership The success of open data initiatives is dependent on government buy-in and ownership. Donors may provide initial support, but awareness is required to ensure that government allocates required resources including sustainable budget for sustainability. Government is best placed in creating awareness among its agencies, and creating suitable environment for the private sector to participate and contribute to open data initiatives.

Table 8.31: Awareness, buy-in and ownership

8.6.5 Advocacy

Two major themes were identified from the first order themes in this category. These describe the need for open data champions and the criteria for determining suitable candidates for this this role. There is need for open data champions to have strong political capital in order to influence

the complex organization defining the open data environment, which consists of actors from government, private sector and the political class. Some of these actors appear in two or all of the three sectors, while others depend on each other for survival. The main role of open data champions is advocacy which requires a strong political capital in order to be effective. Table 32 below provides the details of these findings.

First Order Theme	Clustered Theme	Second Order Theme
There is need to identify open data champions at various levels of government including the executive, cabinet, county government, and agency level. This will ensure accountability and reduce bottlenecks since the responsible persons will assist in removing possible bottlenecks in the timely release of open data. Open data champions require political capital to be effective. This helps in resolving possible resistance and circumventing bureaucracy within government. KODI requires open data champions who will challenge the political economy and demand that open data policy be enforced. There is resistance in some innovative ideas, which may threaten a certain establishment either in government or among non-governmental organizations. For instance, an effort to compare donors and their contributions was rejected. This affects buy-in and public participation as the application of open data which makes sense to them is not made available. There is need for an open data champion to advocate for the release of such datasets.	Theme 1: Open data champions are required at various levels of government including the executive, cabinet, county government, and agency level. Theme 2: Political capital determines the effectiveness of an open data champion.	Advocacy There is need for open data champions at various levels of government. These champions need to be nominated based on their political capital, as this determines their effectiveness and influence in formulating and implementing law and policy on open data.

Table 8.32: Advocacy

8.7 Conclusion

This phase started with the development of code manual prior to analysis based on the research questions and theoretical concepts. A definition of what each code represents and a description of the features that can help in identifying the occurrence of a particular code was provided for each. A total of twenty-one codes were identified in the process. This helped in organizing portions of related text and deriving new insights from the patterns that emerged.

The next step involved testing the reliability of the codes defined in the code manual. This was achieved by selecting a sample of five interview transcripts and using them to validate the applicability of the code manual to the data. The first step in testing for applicability was identifying a few phrases from the sample transcripts that represent each code. This process resulted in changes in some of the codes to make them more representative, and removal of others that were not representative of the data. In addition to these modifications, new codes

were formed during analysis in cases where the code manual was not representative of an emerging theme.

After testing for code reliability, the researcher summarized all the interview data based on the research questions, and identified preliminary themes in the process. The aim of this process was to internalize the information contained in the data, with the aim of identifying patterns and new meanings which are described in the form of preliminary themes.

In addition to the summary, the researcher used the raw data from transcripts and project documents to perform additional coding using the code manual. This resulted in further modification of the code manual, and in cases where the modification was fundamentally different from the preliminary code, a new code was formed. Once the code manual was finalized, the level of groundedness and representativeness of each code was determined with the help of Atlas.ti. Groundedness refers to the total number of quotations linked to a particular code, whereas representativeness is the proportion of documents that make reference to a particular code.

This was followed by the process of connecting codes and identifying themes. This process involved matching similar codes together into common themes based on the research questions. These themes were identified per group where it applied, and helped identify the themes that emerged for each group, and whether views from the various groups varied or there was a consensus of opinion in a particular question. These groups consisted of the ministry of ICT and ICT Authority (ICTA), ICTA data fellows, Non-Governmental Organizations (NGOs) and civic tech, other government agencies, and Intra-Governmental Organizations (IGOs). This resulted in further internalization of the data.

The last stage involved corroborating and legitimating code themes, which can be described as the process of clustering themes identified previously from the coded text. The aim is to confirm that there is no fabricated evidence that may have been introduced while interpreting data. The outcome of this process is an explanatory framework of the themes that enable proper institutionalization of KODI. Five themes emerged from this process and consisted of the

following: demand; law and policy; planning, coordination and capacity building; awareness, buy-in and ownership; and advocacy. This implies that the objective of this study has been fulfilled after identifying these themes though further discussion from an institutionalization theory and critical realism perspective will be provided in the next chapter.

9. Phase 4 - Theory Testing

Theory testing aims at refining the understanding of CMO configurations using empirical findings. This is achieved through an assessment and interpretation of the findings arising from phase three on data analysis. In particular, the themes identified in phase three are analysed with the aim of identifying enabling mechanisms. This process starts by analysing these themes using Fereday & Muir-Cochrane (2006) five stage model in search of mechanisms and supportive elements such as causal powers and linked mechanisms. This is followed by hypothesis testing of the candidate mechanisms provided in table 6.6 above. It is not a must for the CMO configurations to match, but is useful in identifying overlaps between the proposed/candidate mechanisms and the observed mechanisms in this phase. This also helps in identifying gaps in the observed mechanisms. Finally, an explanation of the observed mechanisms is provided using a theorical lens, which in this case is the institutionalization theory provided in table 4.1.

This process is repeated several times with the aim of drawing closer to a more accurate explanation of the enabling mechanisms and outcomes of a particular programme. Realist evaluation starts with a theory as illustrated in phase one, and ends in a theory, both of which seeks to describe what works, for whom, in what context and in what respect (Pawson & Tilley, 2004). However, it is important to note that testing of the observed mechanisms/theories is dependent on the researcher's capacity to bring about changes in reality. This is related to the intransitive nature of objects, where they continue to exist and act regardless of our knowledge about them (Bhaskar, 2008, pp. 11–14).

Following an understanding of what this phase entails, the following sections will start by providing a description (table 9.1) of the derived mechanisms in phase three on data analysis. The configuration of these mechanisms constitute the theory which helps in explaining the outcomes of KODI. This will be followed by a comparison with the theory that had been proposed in phase one, to confirm if the proposed mechanisms match those derived from analysis. The last section will discuss the actual mechanisms using the lens of institutionalization as described in chapter four table 4.1 institutionalization thematic table.

9.1 CMOs Affecting KODI Institutionalization

Mechanism and Context	**Real Objects** People/Systems	**Causal powers** Conditions/ Pressures	**Outcome**	**Linked Mechanisms** Dependencies, feedback loop
Demand - Search Engine Optimization (SEO) to increase access and visibility of open data. - Engagement channels to address user needs - meta-data provisioning and use of open data standards to increase accessibility and use.	- Government - Citizens - Civil society - Private sector - Infomediaries who consist of app developers and data wranglers	- Buy-in from government - Openness by government to publish requested datasets	- Use of open data - Up-to-date datasets - Citizen engagement which includes request for new datasets or request for changes on existing datasets - Meta-data provisioning for all datasets - Publication of open data based on open data standards	- Planning, coordination and capacity building
Law and policy - Identify, amend or create suitable laws and policies on data publication and use - Control publication and use of data using copyright laws and disclosure policies. No one should withhold data/release tampered data - Protect the privacy of government agents - Establish and implement a legal framework and policies on right of access to information, confidentiality, exceptions to openness, and intellectual property rights - Establish engagement platforms to support citizen participation in law development or amendment.	- Data Publishers - Data Aggregators - Government - Citizens - Civil societies	- Copyright laws - Disclosure policies - Government commitment - Openness and transparency of government officials	- Publication of data - Confidence among government agents on privacy - Transparency and accountability	- Advocacy
Planning, coordination and capacity building - Develop strategies that increase efficiency - Plan and coordinate the process of developing and implementing required laws and policies.	- Government - Citizens - Civil societies - Infomediaries - Open data champion	- Bureaucracy - Technophobia - Supportive law and policy - Political capital of the project coordinator and open data champions	- Increased efficiency in open data delivery - Capacity to curate and publish open data. - Reduced	- Law and Policy - Awareness, buy-in and ownership - Demand

- Reduce redundancy especially among government agencies. - Understand government processes and plan for delays arising from possible bureaucracies. - Develop a communication strategy to avoid misconceptions when dealing with government. - Identify skills gap and organize for training. This could also be used as a strategy for awareness that may lead to buy-in by the stakeholders. - Facilitate proper staff handover - Set and manage staff roles and responsibilities - Establish competitive channels for staff recruitment - Develop and issue best practice guides - Recruit open data coordinator with a strong political capital since they need to engage with various government agencies and civil societies in the process of implementing the open data initiative		will determine effectiveness.	bottlenecks associated with bureaucracy and technophobia. - Reduced staff turnover. - Efficiency in new staff and client onboarding on how to curate data.	
Awareness, buy-in and ownership - Secure buy-in from key stakeholders including the executive branch of national government. - Identify the potential transformation from open data and use this to build a case for open data. - Develop engagement channels that will foster public participation and ownership. - Promote data-driven journalism - Secure budget from government to support the open data initiative. - Secure buy-in from private sector by eliminating the fear of losing competitive advantage.	- Government - Civil societies - Open data champions - Private sector - Infomediaries - Journalists - Citizens	- Political goodwill characterized by direct investment by government to open data. - Political economy. This is determined by how open data is perceived to affect the nexus between government, private sector and politics.	- Buy-in from government, private sector and the public. - Sustainability following budget approval and resource allocation by government. - Increase demand from private sector including media companies through data-driven journalism - Increased public participation and interest in open data.	- Advocacy. - Demand
Advocacy - Nominate and incentivise open data champions at various levels of	- Civil societies - Open data champions	- Resistance to curate and publish certain datasets	- Political capital to institutionalize the open data initiative.	- Law and policy - Demand

government based on their political capital. This includes champions who are able to influence the executive branch of national government. - Devise strategies to resolve possible resistance on the open data initiative.			- Ownership of the initiative through the open data champions. - Increase in demand of open data	

Table 9.1: CMOs Affecting KODI Institutionalization

9.2 Hypothesis testing

Table 6.6 in chapter 6 on hypothesis provides a description of the five mechanisms were proposed from literature. These include; *Law & policy enforcement* - that involves the creation and amendment of laws and policies dealing with data publication and use, *skills management* - which involves training and strategies for staff retention and handover in cases where a staff was to leave an organization, *efficiency* - which is defined as the timely publication of data coupled with proactive disclosure of data and devolved decision making to reduce possible bureaucracy, *data quality* - which entails meta-data provisioning and use of open standards, and *stakeholder engagement* - which refers to the process of identifying and engaging with stakeholders to identify needs and opportunities.

This section will demonstrate whether these propositions were accurate, by comparing them with the mechanisms that were identified in the analysis phase. This process will help in testing the current mechanisms, and in cases where there are overlaps between the proposed and observed mechanisms, determine whether there are concepts from the proposed mechanism that could improve the observed mechanism.

9.2.1 Demand mechanism

This includes strategies aimed at promoting ease of access including search engine optimization and engagement channels that assist in amplifying the voice of the ordinary citizen and ensuring feedback on each request. From the proposed mechanisms, the efficiency and data quality mechanisms resonate with this mechanism. Though the need for meta-data provisioning and use of open data standards had not been defined in this mechanism, these have now been included as part of the context and outcome of the demand mechanism.

9.2.2 Law and policy mechanism

This mechanism had also been identified as part of the proposed mechanisms though it emphasized on enforcement and was labeled as law and policy enforcement. It was observed that some of the critical laws and policies at the time of data collection had either not been formed or approved by the necessary authorities, which implies that enforcement only applies when the right laws and policies have already been formed and approved. A good example was the access to information act 2015 that was only signed into law by the President in September 2016. This law gives Kenyans the right to freely access information held by government and other public organizations.

9.2.3 Planning, coordination and capacity building mechanism

This mechanism is similar to the proposed skills management mechanism. Planning and coordination of open data activities was found to be critical in the success of the initiative. This implies that even with a skilled workforce, efficiency can only be realized when there is proper planning and coordination. Some of the aspects that were borrowed from the proposed mechanism included the following; facilitate proper staff handover, set and manage staff roles and responsibilities, establish competitive channels for staff recruitment, and develop and issue best practice guides. The envisaged outcomes for these additional aspects are reduced staff turnover since staff feel more motivated when working in a place with sound leadership and clearly defined milestones, including best practice guides, which act as a reference point for their operation. A related outcome to this is efficiency in operations and an increase in demand for data as it is more reliable and accurate.

9.2.4 Awareness, buy-in and ownership mechanism

This mechanism is similar to the proposed stakeholder engagement mechanism. Though citizen engagement and participation as defined in the proposed mechanism is critical, the findings suggested the need for data-driven journalism which is supported by infomediaries that assists journalists in making meaning out of complex data. The Kenyan initiative is purely driven by donor funding. This was perceived as a threat to sustainability and a sign that government is yet to buy-in and own the initiative. Coupled with this is the realization that buy-in, especially within government can best be achieved if the initiative is driven and funded by government and not a donor agency.

These aspects were missing in the proposed mechanism calling for the need to devise a strategy that includes efforts to create awareness and initiate the process of creating required laws and policies, which could be funded by donors at the initial stages, but once buy-in is secured from the national executive, plans to secure budget from government should commence. Once this is achieved, donor funding should be terminated since the initiative could be termed as foreign or western driven, which may adversely affect the political economy following negative perception.

9.2.5 Advocacy mechanism

Like the previous mechanism, the advocacy mechanism is also similar to the proposed stakeholder engagement mechanism though the need for open data champions and the requirement for them to have a strong political capital was more pronounced. These two aspects form the missing components in the proposed mechanism, and the need to define the mechanism as advocacy since this is the main role of the open data champions. This mechanism proposes the need for a strategy on how to nominate and incentivize open data champions, who should have a rich political capital able to influence policy makers at the target agencies. There is need for some of the open data champions to have access and influence the executive branch of national government, in order to ensure continued buy-in and support from the executive, and push for policy and legal changes where necessary.

9.3 Explanation of KODI CMOs based on institutionalization theory

This section is guided by *Table 4.1: Open Data Institutionalization Case Analysis Guide,* that contains institutionalization concepts that were found to be applicable to open data initiatives. This acts as the guide that provides clarity on the assumptions and research questions that are associated with each of the identified concepts.

Six concepts were identified and consist of the following; path dependence - which focuses on the impact of history and external determinants, reciprocal typification or habitualization - which focuses on replicable problem-solving behaviors, objectification - that results from shared meanings and mutual awareness among actors, sedimentation - which investigates whether the main actions acquire a quality of exteriority that is characterized by diffusion and retention of the supporting structures, coercive isomorphism - that refers to the formal and informal pressures an

organization experiences from society and other organizations, and mimetic isomorphism - which studies how organizations imitate others as they learn or in cases of uncertainty on how to implement certain structures.

9.3.1 Demand mechanism

This mechanism is linked to two institutionalization concepts, namely sedimentation and coercive isomorphism. In the case of sedimentation, the need for search engine optimization, the addition of meta-data on all datasets, and monitoring of engagement channels requires action from the responsible agencies. Policies related to these requirements, if adhered to will result in historical continuity, since the actors responsible uphold them. Coercive isomorphism will result once the engagement channels become active, and citizens' requests become more unpredictable, requiring more resources to attend to their requests in-order to maintain buy-in.

These recommendations will help in incorporating the politics of OGD, since more users' interests are dealt with, as they are able to find data with ease, raise concerns, request for changes or additional datasets (Shah, 2013). However, it is important to note that some citizens will still be discriminated as a result of differential privileges in tools and skills access (Gurstein, 2011).

9.3.2 Law and policy mechanism

This mechanism is associated with three institutionalization concepts, namely mimetic isomorphism, objectification and reciprocal typification/habitualization. Mimetic isomorphism is evident because most of the laws and policies developed for the cause of open data have been guided by those by countries which have already implemented open data, or either inter-governmental or non-profit organizations such as the Open Data Institute (ODI). Kenya depended on advisory from the World Bank, who used their experience from other countries, including bringing experts from the United States of America, to help develop similar structures and systems to those in their home countries. It could therefore be argued that this initiative resulted in direct business opportunities for the foreign stakeholders. As a result, it is difficult to completely eliminate capitalism from the transparency and accountability agenda in the case of a developing country such as Kenya, which witnessed a foreign company being awarded a lucrative tender without much competition (Bates, 2012).

Objectification was supported by law and policy which directly results in shared meaning and understanding once people internalize the formulated laws and policies. For instance, the access to information act brought clarity on the data citizens have access to. It also proposes cover on public servants whose fundamental freedom may be threatened if certain datasets are made public. Despite this, the problem of the low-hanging fruit may still arise. Objectification of open data laws and policies does not guarantee that sensitive data will be made available. There is need for management to intervene in ensuring that such data is revised in accordance with the law, and then published without delay (Chignard, 2013).

Reciprocal typification / habitualization is also supported by laws and policies, commonly referred to as structural arrangements that become internalized by the actors of open data, and guide the process of curating and solving open data challenges, and may result in cultural change such as proactive disclosure of open data as a result of these laws and policies. This suggests a spike in OGD release. However, it must be noted that such spikes are often shortlived, mainly because of OGD limitations. Suppliers of OGD in this case may reallocate resources if demand on their datasets does not justify the costs. This is mainly because users find it difficult to determine how OGD can be profitably analysed and used. In addition to providing the necessary laws, there should be policies aimed at facilitating use (Chignard, 2013; Kitchin, 2014).

9.3.3 Planning, coordination and capacity building mechanism

This mechanism is associated with three institutionalization concepts namely, mimetic isomorphism, sedimentation and reciprocal typification/habitualization. Mimetic isomorphism will become manifest in cases where agencies that have just been introduced to open data grapple to understand best practice guides and tools for curating data, and result to imitating those who have succeeded. In the case of KODI, the agency staff learnt through observation as staff from ICT Authority, specifically the data fellows demonstrated on how to use the various tools for curating data. There was enough time to observe and even imitate as the data fellows were deployed at an agency for a period of six months which was assumed to be sufficient for skills transfer. Unfortunately, there was poor planning and communication, and the effectiveness of the data fellows was curtailed. These shortcomings pose a challenge to the sustainability of KODI. They are a result of lack of management intervention, since no arrangements were made

to provide the necessary resources to the data fellows, or followup to ensure that they were properly introduced and hosted at the various agencies (Chignard, 2013; Kitchin, 2014).

Sedimentation was made manifest through policies and best practice guides, including the structures that manage staff recruitment, workflow and the budget. With time, these will be integrated into the current system. However, the level and pace of integration is dependent on the actors involved to uphold them. This will result in historical continuity of these structures when they start being treated as "social givens" by new actors. For instance, they will continue publishing data to the public as part of their daily operations. These efforts should go hand-in-hand with provision of services and tools that facilitate use, since data-demand is what justifies data-supply (Kitchin, 2014).

Another institutionalization concept related to what supports sedimentation in this context is that of habitualization also known as reciprocal typification. With good planning, coordination and capacity building, some of the processes will be documented and their workflow will be defined, making it easier for them to be habitualized by the existing actors and those that may join at a later stage. This was observed in the case of KODI. For instance, ICT Authority developed the data release form which helps in defining the workflow of the curation process, identifying responsible persons within an agency for a particular dataset, determining the frequency that data is released, and securing commitment from the agency, which includes ownership of the dataset since ICT Authority cannot be held responsible for the datasets, as they are not the subject matter experts.

9.3.4 Awareness, buy-in and ownership mechanism

This mechanism was linked to the coercive isomorphism institutionalization concept. Coercive isomorphism will be experienced mainly through the engagement channel once the public understand the need and potential for open-data, and how to make use of the engagement channel. Policies defining the turn-around time between the time it takes for a request to be processed will create pressure in times where the engagement channels are flooded with unique user requests. Additional pressure will be felt in cases where the public demands certain datasets which government finds contentious to release, probably because there was a scandal linked to it. This creates an opportunity for the public to be involved in the running the affairs of government

and increasing trust in government as a result of accountability and transparency. It is also necessary to address the challenge of differential privileges among citizens, by developing programmes aimed at leveraging OGD access and skills among the less privileged. This will assist in promoting greater inclusivity and diversity of users (Kitchin, 2014; Shah, 2013).

9.3.5 Advocacy mechanism

There are two institutionalization concepts that are emerge from advocacy namely, objectification and coercive isomorphism. In the case of objectification, open data champions help in communicating a common understanding of open data, which results in shared meanings and definitions across all stakeholders. Their effort is supported and legitimized by existing laws and policies, such as the access to information act, which requires government agencies to release data to the public. It is during this process that stakeholder interests become manifest. In some instances, political parties and business entities have only shown partial support to non-sensitive agendas. They tend to stay away from right to information movements and whistle blowers, since some of the agendas being raised threaten their very existence (Bates, 2012). They understand that OGD is disruptive, and that some of their business models may be threatened. This creates the need to plan for such resistance, and where possible, provide assurances aimed at securing greater buy-in without compromising transparency and accountability efforts (Kitchin, 2014).

Coercive isomorphism is experienced by government when open data champions with a strong political capital push an agenda for change or implementation of certain laws and policies. The open data champions will also experience functional pressure from government agencies, especially in cases where needed resources and infrastructure are not available. If this is not addressed in time, there could be resistance by government agencies that could be characterized by deployment of resources to other projects. This implies that even when there are laws and policies that help in securing buy-in by government, there is need to create the right environment for agencies to execute their mandate.

9.4. KODI Mechanisms Interconnectedness

Figure 9.1 below provides an illustration of the dependencies and feedback loops that exist between the mechanisms that enable proper institutionalization of KODI. All these mechanisms support each other and are interconnected directly or indirectly. Continuous lines were used to illustrate direct dependencies, while dotted lines were used to illustrate indirect dependencies. For instance, the access to information act is part of law and policy mechanism, and grants citizens the right to request for data held by government agencies. This implies that demand has been legitimized and that law and policy can be said to support demand mechanism.

Demand mechanism is also linked to the awareness, buy-in and ownership mechanism, though this can only be achieved in conjunction with the advocacy and planning, coordination and capacity building mechanisms. Though not illustrated, there is a symbiotic relationship between advocacy and planning, coordination and capacity building mechanisms. Advocacy requires input from planning, coordination and capacity building mechanism, which helps in identifying possible gaps in KODI that could be resolved by amending existing laws and policies or formulating new ones. These laws and policies assist in legitimizing the implementation of the proposed strategies.

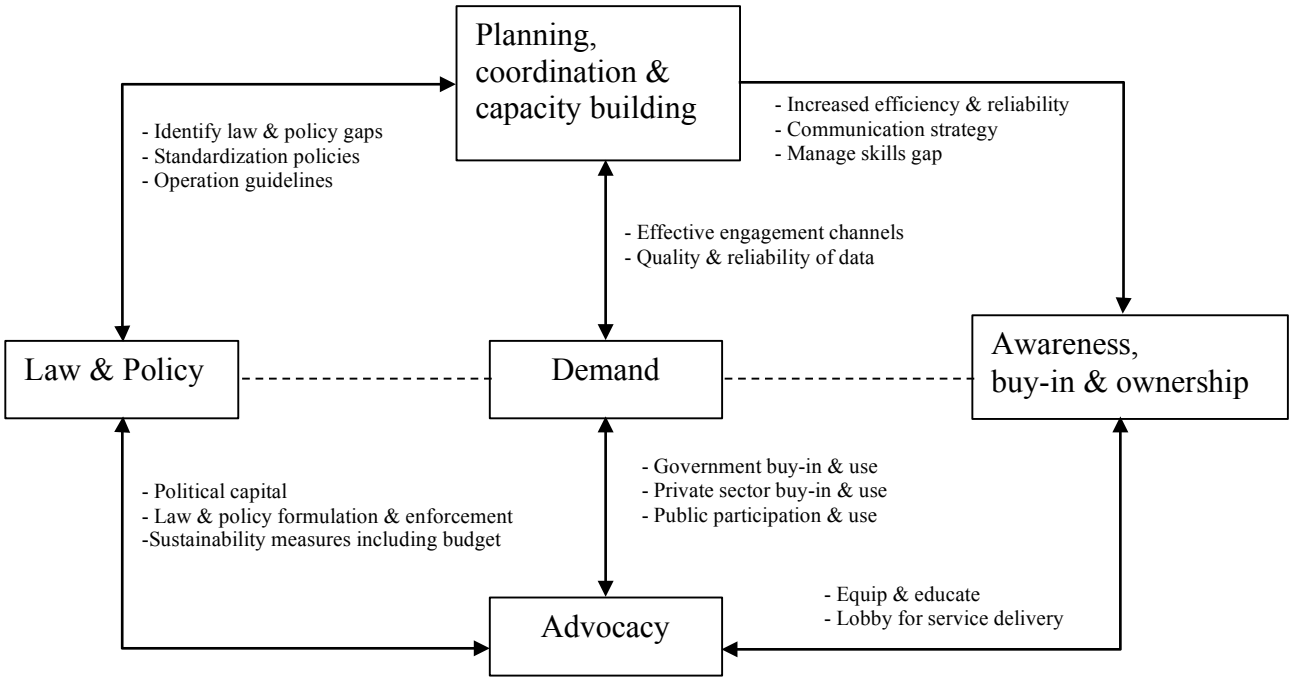

Figure 9.1: KODI Mechanisms Interconnectedness

9.5. Conclusion

This chapter helped in testing the theories that were formulated in the analysis phase by comparing them with the theoretical propositions in the hypothesis phase and conducting further analysis using the theories of institutionalization identified in chapter four. There was an overlap with the proposed theories, the only major difference being that some aspects were more pronounced in the Kenyan context than what literature suggested. This resulted in a different set of CMO configurations that will assist KODI in tapping the full potential of open data once it becomes fully institutionalized.

These will result in objectification, where the various actors will have shared meanings and understanding of what open data refers to and why it is important. In addition, reciprocal typification/habitualization, which is the other institutionalization concept to manifest will be characterized by properly defined structures and systems that support open data, including proper documentation of open data workflow, which results in easy to follow procedures and guidelines.

These will become internalized and incorporated in the organization's culture, resulting in sedimentation, which is the third institutionalization concept in this argument. Sedimentation will be made manifest when the actors in various organization treat open data processes as social givens, and will defend for their continued implementation incase a change is proposed. However, for this to be achieved, there is need for the project coordinator and open data champions to have political capital in order to influence government agencies and the executive in formulating or enforcing laws and policies, and in the process develop shared meanings (objectification) and buy-in which will guarantee sustainability of the initiative. Following this argument, it is correct to state that political capital is paramount in the success of KODI.

10. KODI Case Study

Pawson & Tilley's realist evaluation model assisted in identifying several underlying mechanisms that affect proper institutionalization of KODI. This chapter provides a narrative of KODI and incorporates the findings of the evaluation process, which are summarized in the previous chapter on theory testing. The following section provides an introduction of KODI, followed by a description of the adopted structures, and concludes with a summary of the challenges that KODI has encountered since 2011. Some of these challenges have been resolved adequately, while others are still lurking.

10.1 Introduction

Kenya has been at the forefront of open data since 2011, and according to the 2016 open data barometer report by the World Wide Web Foundation, the country was ranked at position 42 out of 92 countries worldwide. Even though this is the highest rank among countries in Sub-Sahara Africa, the country still faces numerous challenges in the institutionalization of open data. Despite this shortfall, the open government data initiative still needs to be celebrated as it has helped in creating awareness on the value of openness in the country and beyond.

It is spearheaded by the Kenya Open Data Initiative (KODI), which operates under the Kenya ICT Authority, a state corporation under the Ministry of Information Technology. KODI has seen enormous growth in terms of datasets and buy-in from government agencies. Currently, over 902 datasets have been published from 31 out of 62 government agencies. The open data portal has had more than 170 million page views since its launch in 2011 and remains to be the most visited government portal in the country.

The sections below provide a description of the structures that have been put in place to help ensure buy-in and institutionalization by the various government agencies. This includes the challenges that the KODI has faced since it's conception in 2011, and an account of how some of these challenges have been tackled so far.

These strategies and challenges will be described using the lens of institutionalization that was developed in chapter four under *Table 4.1: Open Data Institutionalization Case Analysis Guide*,

and the KODI CMO configurations described in chapter 9. These will help in understanding the institutionalization concepts that are emerging from the case description. This case is derived from the data collected from the various interviews with KODI stakeholders and documents obtained from them during the interview process.

10.2 KODI Strategies and Supporting Structures

There are several useful measures that KODI has undertaken to ensure its sustainability. These include organising capacity building trainings, conducting sensitization and awareness campaigns, and developing and implementing enabling policies and guidelines. A detailed description of these measures is provided in the sections that follow.

10.2.1 Re-launch of the KODI Portal

The KODI portal was re-launched in July 2015. This followed enormous work on the initiative with the aim of ensuring continuous publication of more timely and diverse datasets, and improved user experience. This process was dependent on the planning, coordination and capacity building mechanism and reciprocal typification/habitualization which would help in scaling to other agencies using standardized processes and procedures. The KODI portal had 680 datasets at the time of the re-launch, which was a significant rise from 262 datasets at the 2011 launch. This was as a result of the rise in the number of active government agencies, which shifted from 4 in 2011 to 31 by the time the launch took place.

This is a change that mainly happened in 2014 and the start of 2015, when the team at KODI went on a sensitization and awareness campaign targeting various government agencies. This implies that there was additional buy-in and ownership from the government agencies, which is a sign of awareness, buy-in and ownership mechanism. The rise in datasets saw a significant rise in the number of page views from 44 million views to 54 million views, making it the most visited government portal in the country at the time.

Some of the government agencies, which were previously publishing data on their portals preferred to publish on the KODI portal since their data got more views there compared to their website. This became an additional incentive for agencies to submit their data to KODI. It is projected that this process will eventually reach sedimentation, where staff at the various

government agencies treat the curation process as a 'social given', after the involved processes become infused into the organization culture and upheld by the actors affected by the process.

10.2.2 Data fellows programme

In an effort to build capacity among the various agencies and increase data surfacing, ICTA has been conducting training and deploying sets of data fellows – mainly young graduates to a number of government agencies to assist in build required ICT capacity in these agencies. The contracts run for a duration of six months. Their role entails working with the allocated agency to establish and implement internal procedures and guidelines for data publishing. This resulted in objectification, where agencies developed shared understanding of open data and the related processes, and sedimentation since this effort aimed at diffusing and retaining structures that support open data.

Each agency is allocated two data fellows, comprising of a computer data management fellow and a communications/ public engagement fellow. The former is meant to assist in manipulating and visualizing large data sets in various formats, while the latter should assist in public communication and engagement through social media and strengthening relations between the agency and ICTA in relation to this initiative. ICTA commissioned the first set of data fellows in the second half of 2015. The first set consisted of eight data fellows who were assigned to three government ministries and a county government for a period of six months.

The second set comprises of fifty data fellows aimed at twenty-five government agencies. The planning, coordination and capacity building mechanism is made manifest in this strategy whose main aim was capacity building, with the aim of increasing buy-in and ownership by other government agencies. This would help in addressing the challenges that were experienced in the first phase of the data fellows programme, where data fellows were perceived as auditors resulting in resistance and ineffectiveness of the first phase of the initiative. This was as a result of poor communication between ICTA and the target government agencies.

10.2.3 Significant number blog

This is a KODI team initiative, and involves publication of a blog post based on a significant number identified from one of the datasets. The significant number results from the analysis of a dataset, aimed at identifying an interesting number or pattern from the dataset. For example, one of the blogs pointed out that the highest number of road fatalities in Kenya occur at 6pm. This is weird because this is the time that traffic is at a gridlock. Nairobi, the capital city of Kenya has the highest traffic and contributes the bulk of road accidents in Kenya doing something about 50-60%. So, when we consider that, why are all these people dying and traffic is very slow at 6 PM? (ICTA, 2016).

These are interesting questions to grapple with, which demonstrate how the value of open data can be derived through stories. This is part of data-driven journalism, though it does not restrict non-journalists or non-seasoned bloggers from writing compelling stories. For instance, KODI does not have trained journalists, yet they are able to come up with such compelling stories. This initiative activates the demand mechanism, as citizens and media houses get to understand of the potential of open data, and also derive useful information from the blog posts. Once there is significant demand, KODI will experience coercive isomorphism as society and private companies increase their demand for open data, or stories based on open data. For instance, if a blogpost becomes popular, citizens may request for additional datasets or stories. In this case, this would be termed as social pressure. It could also become political pressure if the opposition party seeks facts explaining certain government operations. However, it is important to emphasise that demand will only increase when tools and services for facilitating use of OGD are made available (Kitchin, 2014).

10.2.4 Request a dataset

KODI online portal has a feature that allows citizens to request for datasets. This helps KODI in identifying the datasets that are of interest to people. This helps them in developing some criteria for determining which government agencies to engage with in obtaining the requested data, and the agencies to prioritize for the data fellows programme. This is an initiative that directly activates the demand mechanism. It is supported by the engagement channel that provides citizens with an interface to interact with government, and demand for data that affects their lives. However, for this to be realized, awareness, buy-in and ownership mechanism is required.

Unfortunately, bureaucracy often slows down the turnaround time to these requests, as the proper communication channel has to be followed. For instance, when dealing with the ministry of education, KODI team have to write to their CEO at ICTA, who then requests the Principal Secretary (PS) at the ministry of Information and Communication to write to the PS at the ministry of education. At times once the communication gets to the other ministry, it is directed to the wrong department, mainly because they assume that all communication from the Ministry of Information and Communication must be directed to the ICT department within the agency.

This requires the KODI team to keep tracking these requests in order to pick up such anomalies. Their experience is that emails don't work between agencies, and that physical letters are the preferred mode of communication. This is tedious and unsustainable. They hope that the Access to Information bill will assist in reducing these bureaucratic hops and that they will have a shorter turnaround time, leading to more relevant datasets. This is a manifestation of coercive isomorphism, since formal functional pressures create inefficiencies which make the portal less reliable, following the uncertainity in the turnaround time needed to obtain some datasets.

10.2.5 Data release form

The data release form was designed by ICTA to help in formalising the contract between them and a government agency per dataset. This provides proof that both organizations consented to the publication and continuous update of a particular dataset. The terms and conditions are as follows;

> **"1)** This schedule confirms that the following listed datasets are to be published to the open data portal, www.opendata.go.ke. **2)** Acknowledges that the data will be licensed using the creative commons 1.0 Universal framework to allow the greatest reuse potential in the long term. **3)** Confirms that the data will not be changed in any way. However, recognizes that the schema may be altered to optimize its overall usability on the Open Data Portal. **4)** It is the responsibility of the data owner to ensure they continuously supply updated records on a regular schedule, or alternatively, **5)** The Data owner agrees to collaborate with the ICT Authority and Kenya Open Data, to phase automatic updates / semi-automatic updates of data

provided. **6)** Further recommends tools or procedures as may be constructed by both parties to support the long term sustainability of the initiative; Including the use of Data Sync, and or other ETL tools."

This is an effort towards habitualization/reciprocal typification since the data release form defines a standardized contractual process between an agency and ICTA. This will eventually result in sedimentation as the actors involved treat the process as a 'social given' and propel their historical continuity. This is a sign of proper planning and assists in coordination between ICTA and agencies, which implies that the planning, coordination and capacity building mechanism assists in habitualization and sedimentation of KODI.

10.2.6 Awareness and capacity building workshops

ICT Authority has organized several workshops targeting government agencies, academia, media, civil society organizations, developers and the general public with the aim of increasing buy-in and ownership of the open data initiative. These workshops have been face-to-face, though there are plans to have webinars, as these do not cost as much, except for the time and manpower in setting up the online meetings. The main outcome of these workshops is objectification, which results from shared meanings and mutual awareness among the actors involved in these workshops. These efforts support the awareness, buy-in and ownership mechanism. However, there is need to ensure greater inclusivity since citizens have differential privileges that often leads to marginalisation and low data demand (Kitchin, 2014).

10.2.7 Internal capacity building

Since ICTA employed more staff in 2014, the capacity issue was resolved temporarily. ICTA used a grant from the World Bank to employ staff working on KODI. Unfortunately, when the grant ended in December 2016, the entire team had to be sent home since ICTA had not secured alternative budget from government. Despite this, the 2014-2016 team made significant strides. They were also assisted by contract staff employed under data-fellows programme during this timeframe. The data-fellows were mainly fresh graduates and were hired for a period of six months, on a non-renewable contract.

The data-fellows reported a few challenges, which affected their productivity. They complained of late payment and lack of resources, in particular personal computers. For the latter, it could have been an issue of bureaucracy at ICTA, though these issues should be addressed prior to recruitment, to guarantee productivity from the onset. This initiative relates to the planning, coordination and capacity building mechanism, and as discussed earlier in section 10.2.2, there is need for better coordination to avoid misconceptions and reduce possible bureaucracies.

10.2.8 Funding

KODI was funded by the World Bank, and since the grant ran-out in December 2016, the initiative came to a stop. ICTA could not retain the staff working on KODI since their salaries were fully dependent on the grant. Delayed action by government could be translated to fear of disruption by OGD. It could also imply that stakeholders within government have other interests that are being threatened by this transparency and accountability agenda, which then explains why some of the sensitive datasets remains unavailable (Bates, 2012; Kitchin, 2014).

To ensure contuinity of the initiative, government needs to allocate budget, and approve a structure that will guarantee project contuinity. For instance, government could incorporate Open Data within the Key Performance Indicators (KPIs) of public officials. This would increase the uptake of open-data and awareness since those tasked with an open data deliverable will strive to understand it, with the aim of attaining their KPIs. Therefore, government has a role in creating awareness and buy-in, which implies that the awareness, buy-in and ownership mechanism is dependent on government support.

10.2.9 Official vs crowdsourced data

KODI only deals with official data from government. This implies that it does not publish data generated by non-governmental agencies or crowdsourced from the general public. However, KODI has been providing advisory and holding joint events with other stakeholders as an effort to exchange knowledge and experiences. Despite this limitation on data sources, the KODI platform needs to be optimized for easy data retrieval and enriched with metadata. This requirement is part of the demand mechanism, as it seeks to address some of the challenges that may affect demand of open data.

10.2.10 Expanded Partnerships

The KODI team has been making deliberate efforts to create and maintain working relationships with the data producers who consist of government agencies, and users who include private citizens and organizations with the aim of ensuring continuous supply and demand of data. They also have a partnership with Socrata, a U.S.A based company. They were appointed following a recommendation by the World Bank in 2010 to the KODI taskforce as the most suitable company to deliver the platform.

This decision was challenged by the Kenyan developer community who perceived this as a lost opportunity to have a home-grown application. The decision was upheld following the argument that there wasn't sufficient time to implement and test, since the launch was scheduled to happen in less than two months. This decision, and the response to the plea by the developer community demonstrates how some business entities gain privilege over others. It also implies that some of the stakeholders had business interests, which were disguised in the transparency agenda. The benefits may not have been monetary, but nationalistic and imperialistic in relation to technology and human expertise (Bates, 2012; Kitchin, 2014).

This initiative also demonstrated characteristics of coercive isomorphism, which arose from negative political pressure from government. This could also explain the slow buy-in of the initiative by government, which has been non-cooperative when it comes to some contentious datasets, such as the school ranking data. This is despite Kenya having laws and policies that support proactive disclosure, including the access to information act which was made into law in 2016. Therefore, there is need for political goodwill and buy-in, which can be achieved through advocacy campaigns by government representatives with strong political capital.

10.3 KODI Challenges

This section outlines the challenges faced by KODI, and the strategies that can be applied to mitigate possible threats and risks arising from these challenges.

10.3.1 Bureaucracy

Compared to countries like South Africa and Tanzania, the open data initiative is managed by their statistical authorities. Kenya has a different approach where the initiative rests under the

ICT Authority (ICTA) and not the Kenya National Bureau of Statistics (KNBS). These two agencies fall under different ministries, that is the ministry of Information and Communication, and the ministry of devolution and planning respectively. This creates a challenge in communication and collaboration as a result of government bureaucracy.

In addition, there are instances where these agencies duplicate effort. An example is the duplication of datasets. These agencies use separate platforms, resulting in version management challenges. KNBS is not mandated to report such changes to KODI. As a result, KODI usually has to devise its own mechanism to keep such KNBS datasets upto date. KODI is justified in maintaining KNBS datasets especially in cases where such datasets have not been made digitally available. Sourcing for these datasets has proven to be difficult in some instances. For instance, the former Principal Secretary for the Ministry of Information and Communications had to scout at KNBS looking for census data prior to the launch in 2011. This demonstrates the need for champions at senior government management level to assist in requesting for policy change or sensitive data from other agencies or ministries. It is not automatic that such a noble idea will be embraced by other agencies as there could be policies or laws such as the official secrets act in Kenya that will block the release of some datasets.

This example characterizes coercive isomorphism from the point of view that KODI through the Principal Secretary pressurized KNBS in releasing the census data (Dacin et al., 2002; DiMaggio & Powell, 1983). This helps in illustrating that there are instances where OGD will be initially made available through coersion.

These challenges require the planning, coordination and capacity building mechanism. There is need to plan and identify overlaps and tension points between the two government agencies. There is also need to coordinate activities at each agency related to open data with the aim of increasing efficiency, reducing bureaucracy and eliminating duplication of effort.

10.3.2 Complexity of data

Government departments operate in silos, follow different standards, leading to incompatible datasets. This has created additional complexity to KODI as they have had to clean and convert such data in the required formats before publishing it to the public. Through the data fellows' initiative, some of the agencies have been trained on how to curate data in the required formats. However, for this to be effective, there is need to engage the planning, coordination and capacity building mechanism to ensure that communication and training reaches the right people, in order to help in resolving some of the data complexities.

Mimetic isomorphism may manifest in cases where an agency, following uncertainties on how to curate OGD imitates another, mainly ICT Authority in the learning process. This has been evident in the data fellows programme, where the agency took a backstage to observe the fellows curate data, before engaging in the process. In some cases, they asked that the data fellows be deployed permanently or have their term extended. This implies that the programme was too short for skills transfer to occur, or that the mode of training needed to be improved to achieve high results within a short timeframe. This also implies that the planning, coordination mechanism is a necessity.

10.3.3 Lack of structures to support KODI

2013 was a difficult year for the initiative. This was aggravated by election campaigns that year and the inadequate staff capacity at KODI at the time. This was a challenge that lasted between 2011 and 2013 where only four government agencies were active, and the number of published datasets stagnated at 430 datasets. However, this changed in 2014 when the ICT Authority increased the headcount at KODI, leading to an increase of 70 additional datasets that year, and an increase in the number of government agencies from 4 to 25.

The new datasets partially explain the rise in number of published datasets. The rest came from datasets which had been submitted by some agencies, but had not been published due to lack of capacity. KODI has since devised ways to sensitize government agencies to publish more data through formal structures such as the data release form and the data fellows programme which ensures skills transfer from ICTA to the various government agencies. The passing of the access to information bill by parliament and it's assent into law by the President in August 2016 has

also assisted legitimizing the call by ICTA to other government agencies to publish their data. This implies that law and policy mechanism is critical in resolving this challenge.

10.3.4 Sensitive data and state interference

Government agencies and county governments have made an effort to release some data. However, the question on granularity remains. An example is county governments budget and expenditure data on the KODI portal, which is currently generalized to sub-county level. For instance, in the 2013-2014 financial period, Kiambu county government committed Kshs. 76 million to develop rural access roads in Ruiru sub-county. This is a massive area in terms of land size and population, and in order for citizens to engage in depth based on facts, it is necessary to provide specifics of which roads in that location have been earmarked for construction and the budget for each. It's also not clear whether these are new roads or repairs, and the type of road surface for each. Another issue is interference from officials in government on what data to publish. For instance, the KODI portal does not have any data on sugar production. This had previously been made available, but after the sugar scandal erupted in 2015, there were orders from above to pull down this data. This followed investigations of corruption and mismanagement of the sugar manufacturing company in Kenya.

There is need for policies which will make KODI more autonomous and independent, and protected by the law for the benefit of the citizen. It is only when such data is released, that journalists and citizens will start to find real value in open data, since conversations can be supported by facts, which can be presented in a court of law if need be. This implies the need for advocacy mechanism, where civil society and open data champions possessing the needed political capital push government to release such data. This call is legitimized by the presence of supportive law and policy. Government in this case will experience coercive isomorphism following political pressure from these actors.

10.3.5 Lack of awareness

The demand side of open data presents an interesting dynamic. Literacy on how to make meaning of data, one would ask, does the ordinary citizen care about open data? This is linked to the question on whether the state is publishing data that either affects their daily lives or helps them keep government and their leaders in check. A case in point is the most recent

announcement by the Ministry of Education in November, 2014 to abolish student and school ranking on exam performance. The reason given was to prevent cut-throat competition among institutions. This new policy also affects school selection in government institutions, favoring learners from public institutions over those from private institutions regardless of their performance.

This has made it difficult for parents to determine where to place their children based on school performance. It also takes away the right of high performers to enjoy the hard-earned and much-deserved recognition. Parents and educators raised these concerns through mainstream media and social media. However, almost two years later, government is yet to revise this decision. This means a lost opportunity for everyone, including innovators who had created mobile/web based solutions based on this data. Therefore, the issue of awareness is not so much whether people know about open data, they are already using it, only that they are not aware that it is called open data, and in some cases, do not know that it is their right. We can only hope that the recently passed access to information bill will help in overturning such policies.

The law and policy mechanism in this case requires amendment or formulation of new laws and policies that promotes disclosure by government. Once the right laws and policies are in place, the aspect of sedimentation arises, where we will observe whether the actors within government are upholding or have rejected these laws and policies. This can be measured by the amount of released datasets based on citizen, private sector, media and civil society requests.

10.3.6 Low innovation

In 2011 and 2012, a significant number of tech innovations emerged based on open data. However, this momentum did not to last for long due to the following reasons. First, the model was based on incentives from the donor community – mainly the World Bank. Developers saw this as an opportunity to make money, and the moment these funds depleted, they moved on to the other financially rewarding initiatives. Second, the open data portal remained dormant for most of 2012 and 2013, with no updates or release of new datasets. This implied that applications that had been built previously lost relevance and traction, which was a great discouragement to the innovators.

This means that despite the pullout of monetary incentives, these developers could not monetize their solutions as they became irrelevant following the lack of data updates. This was also at a time where there was a reshuffle in the Ministry of Information and Communications, which saw Kenya's main open data champion – Dr. Bitange Ndemo replaced as the Principal Secretary. He is the one who had promised this community of change, and unfortunately he did not stay long enough to see his vision come to fruition. This suggests that perhaps additional policies would have been laid during his tenure, which would have helped maintain the momentum of this initiative, even after his departure.

This indicates lack of sedimentation, especially because a few laws and policies promoting openness of government operations had been approved, but the actors within government were yet to uphold them by publishing government open data. The awareness, buy-in and ownership mechanism may be required to sensitize government officials on the potential value of open data, and demystify the myths that may cause them not to publish data.

10.4 Conclusion

This study demonstrates that institutionalizing a country open data initiative requires more than lip service from the various stakeholders - government included. There is need for political goodwill which should be supported by enabling laws and policies, and a dedicated resource team whose mandate includes capacity building, awareness and sensitization. To ensure sustainability, this initiative should be funded by government. Donor funding could help at the initial stages, but no government should peg such an initiative on donor resources. This follows the fact that this funding is premised on donor imperative.

On the issue of laws and policies, it was observed that national elections and change of government in Kenya affected the implementation process. While it is expected that government priorities will shift based on who is in power, such an initiative should be shielded by law and policy. With proper structures, the life of valuable government initiatives is not sapped by change of guard or other political initiatives. An open data initiative for instance should be in higher demand during such times, as people dig into history to create a case for themselves, or challenge others into stepping aside for lack of results.

On the same note, the open data initiative should no longer be viewed as a tool by the west against African governments. This initiative needs to be localized and driven more by locals than foreigners. This implies that government and the private sector will need to join hands in providing more support on open data initiatives including conferences and training workshops. This calls for innovation and monetization. Government and the private sector will definitely be involved on any initiative that has social or economic value. Perhaps the initial phases of open data institutionalization should focus on this aspect as they seek to answer the question, where is the value or social impact of the open data initiative?

Related to innovation is the need for sustainable supportive structures. In the case of Kenya, there were numerous incentive mechanisms which saw many developers interested in creating valuable open data-driven solutions. Unfortunately, these structures were only short-lived, and despite the withdrawal of monetary incentives, data was not updated timeously which significantly reduced the value and demand of the apps. Following this, developers moved on to other initiatives, and ever since, it became an upward task to redeem their trust on the initiative. New successful initiatives have emerged, but it would have been great to have initiatives that began when the initiative was conceptualized in 2011.

It is therefore imperative following the heterogeneous nature of open data, to create structures that will guarantee timeous release of data, which follows adequate planning that factors the complexities that may arise in curating and publishing a dataset. There is also need for flexibility, as has been observed in Kenya. For instance, government agencies should be allowed to publish their own datasets on their portal, though many may change their approach with time, as they realize that the national open data portal has more traffic than the individual agency portals.

In conclusion, there is need for goodwill from government, formulation and enforcement of open data laws and policies including a revision of laws that may contradict those supporting openness, establishment of a government agency to manage the institutionalization of the open data initiative - preferably a separate agency from the statistical authority, ownership by government with less dependency on donor support, timeous release of data, and an

understanding of the social and economic value of open data which will help in driving demand and supply of open data, and consequently guarantee the sustainability of the open data initiative.

11. Conclusion and Recommendation

The aim of this study was to identify the causal mechanisms that lead to effective institutionalization and sustainability of OGD initiatives in Kenya. In particular, it sought to identify the social, economic, cultural, political structures and mechanisms that impact on the institutionalization of the Kenya Open Data Initiative (KODI). These mechanisms and structures could either result in positive or negative impact, where they curtail the furtherance of the institutionalization process. To achieve this goal, critical realism was identified as the underlying philosophy, which was implemented using Pawson and Tilley's (2004) realist evaluation model. Institutionalization theory acted as a lens to help in identifying and describing institutionalization structures and mechanisms. The following sections provide the summary of findings, validity concerns, limitations, contributions and recommendations for further research.

11.1 Summary of Findings

Table 9.1 in chapter nine on theory testing provides the Context-Mechanism-Outcome (CMO) configuration of the KODI institutionalization process. Figure 9.1 illustrates the associations between the emergent mechanisms and Table 9.1 helps in explaining the CMO configuration. The CMO configuration helps in formulating a detailed description of the observable, by referencing the underlying structures and mechanisms. Therefore, it contributes to the details of this study, which is compiled in form of a case study in chapter ten. This case study is based on the outcome of the theory testing phase in chapter nine, which was guided by the institutionalization theory described in chapter four.

Chapter four provides a detailed literature review on institutions and institutionalization with the aim of identifying institutionalization concepts and assumptions that would help in analyzing and theorizing open data initiatives. The findings are presented in Table 4.1 Open Data Institutionalization Case Analysis Guide, and comprises of a set of concepts that characterise institutionalization. These concepts include path dependence, habitualization or reciprocal typification, objectification, sedimentation, coercive isomorphism, and mimetic isomorphism.

Analysis of the programme theories and empirical findings revealed that for KODI to become sustainable, five mechanisms are necessary. These include; demand mechanism, law and policy

mechanism, planning, coordination and capacity building mechanism, awareness mechanism, and advocacy mechanism. A detailed description of these mechanisms is provided in chapter nine, though a brief introduction is provided in the following paragraph.

Demand mechanism requires the formation and implementation of strategies that promote accessibility and reliability. This is achieved through open data standards, timely publication of data, and meta-data provisioning. The law and policy mechanism emphasizes the need for supportive laws and policies, such as the 2015 access to information act which gives Kenyans the right to request for information held by government. The planning, coordination and capacity building mechanism suggests that even with a skilled workforce, efficiency can only be realized through proper planning and coordination. There is also need to plan and coordinate capacity building efforts, in order to ensure that such efforts address the most pressing needs.

The awareness, buy-in and ownership mechanism focuses more on policy makers, and suggests that buy-in and ownership is best characterized by government financial commitment on KODI. This is what guarantees sustainability of KODI, and is a sure sign of ownership by government.
The advocacy mechanism emphasizes the need to have open data champions for advocacy on open data issues. These champions need strong political capital for them to be effective, as they will need to push for the formation and implementation of open data laws and policies.

This set of mechanisms have helped in addressing the aim of this research, as they highlight the strategies that will help in ensuring the sustainability and institutionalization of KODI. This will be characterized by a more informed citizenry, data-driven decision making by government and citizens, increased democracy, emergence of new services built on OGD, increased efficiency in government services, and a more transparent and trusted government.

11.2 Quality Concerns

This study applied various quality checks with the aim of protecting its objectivity, reliability, internal validity, external validity and applicability/utilization. These checks were borrowed from Miles, Huberman & Saldana (2014, pp. 331–335). These checks came with some limitations. To ensure objectivity, the researcher remained impartial by operating in an unbiased and value-free way as was possible. The aim was to ensure fairness while collecting data, which was mainly

through interviews, and while analysing and theorising the findings. This process of data collection and analysis was guided by a five stage model developed by Fereday & Muir-Cochrane (2006). A description of this process is provided in chapter 8 of this study.

On objectivity, the research methods applied in this study helped in ensuring consistency. However, since this was a doctoral study with limited resources, identification of codes and themes from the data was carried out by one person, including the analysis and theory formation. It would have helped to have multiple perspectives, especially during the process of identifying codes and themes. In addition, critical realism is a fallible philosophy (Sayer, 1992, p. 227), implying that the identified structures and generative mechanisms may not fully explain KODI.

To ensure reliability, this study followed two main models, namely realist evaluation model, and Fereday & Muir-Cochrane's thematic analysis model. Pawson & Tilley's realist evaluation model helped in implementing critical realism, which is the underlying philosophy for this study. Fereday & Muir-Cochrane's thematic analysis model assisted in analysing and integrating data during the deductive phase, and theory formation during the inductive phase. Institutionalization theories also guided in the development of data collection instruments, specifically the semi-structured interview questions, code and theme formation during the analysis phase, and theory formation in the fourth phase of Pawson & Tilley's theory testing phase.

On internal validity, the research ensured credibility and authenticity by following accepted research methods, which assisted in forming suitable research instruments for data collection and analysis, and peer review through publication. The researcher wrote and presented several conference papers including the 2016 Kenya data report, which formed part of the Africa Data Revolution Report (ADRR). The Kenya data report helped in soliciting reviews from various actors in the Africa open data space. This report was expanded based on institutionalization concepts to form chapter ten of this study.

On external validity, the findings can be generalized to other settings, though the CMO configurations may differ as some mechanisms may be more pronounced in that context. This implies that some of the mechanisms that were not observed and recorded in this study may be

more pronounced in another context. This therefore creates a limitation on generalizability since mechanisms manifest differently from one context to another following the variation in structure and causal pressures.

On applicability/utilization, this study helped in identifying the causal mechanisms that help in institutionalizing KODI. The case study in chapter ten and the Kenya data report can help policy makers in Kenya to understand the current context and how to improve the initiative. It can also help other countries similar to Kenya, understand what to expect when they start implementing open government data, and what they need in terms of structures and mechanisms.

11.3 Contribution

This section helps in describing the various theoretical, methodological and practical contributions arising from this study. This will be followed by the recommendations of this study.

11.3.1 Theoretical contribution

The first contribution is on the use of critical realism. This study helps in addressing the scarcity of empirical work based on critical realism in the discipline of Information Systems (Henfridsson & Bendik Bygstad, 2013). Experience from this study suggests that in order to effectively conduct a critical realism study, one needs to start by identifying a suitable explanatory model.

Section 3.5 provides a detailed description of seven explanatory models to choose from. This study applied the realist evaluation model, which consists of four stages namely; hypothesis, data collection, analysis and theory testing. This is followed by identification of a suitable theory for testing mechanisms and hypotheses. An example of such a theory is provided in chapter four of this study. An illustration of how this theory was applied in developing and testing the mechanisms is provided in chapters six to ten.

Case study approach compliments critical realism. An explanation of how it can be applied is provided in section 3.6 and 3.7. Chapter 11 provides an example of the resulting case study from this research. Data analysis is a critical stage in either of the seven models. It is advisable for one to identify a suitable model for carrying out this process. This ensures that all the necessary steps

of data analysis are addressed systematically. This study applied Fereday & Muir-Cochrane's (2006) five stage model as its model for data analysis, and found it to be easy to follow and comprehensive.

The second contribution is the open data institutionalization analysis guide, which was presented in table 4.1 in chapter four. It contains a set of institutionalization concepts that could help in analyzing open government data initiatives. These concepts include path dependence, reciprocal typification/ habitualization, objectification, sedimentation, coercive isomorphism, and mimetic isomorphism.

The third contribution is the identification of mechanisms supporting institutionalization of the KODI. Five mechanisms were identified including 1) law and policy, 2) demand, 3) awareness, buy-in and ownership, 4) planning, coordination and capacity building, and 5) advocacy. Figure 9.1 helps in illustrating how these mechanisms are interlinked. These mechanisms could act as a base for analyzing OGD initiatives in a developing country context. A comparative study would help in creating generalizations to some of these mechanisms, especially in cases where they manifest strongly in more than one context.

11.3.2 Practical contribution

This study helped in identifying the mechanisms that affect institutionalization of open government data initiatives, with a particular focus on KODI. The findings will assist policy makers in creating and enforcing the necessary structures necessary for proper institutionalization, which will help in realizing the goals of KODI and similar initiatives. This implies that the outcome of this study contributes to the knowledge base on open government data initiatives and institutionalization.

More specifically, this study presents three practical benefits to policy makers. Firstly, by understanding the underlying causal mechanisms that enable effective institutionalization of OGD, government may be able to form or strengthen existing policies that create an enabling environment. Secondly, knowledge of the underlying enabling causal mechanisms could assist government in determining what to assess when conducting periodic evaluations, and use the findings to identify gaps and opportunities within the OGD phenomenon. Thirdly, mainly from

the social perspective, knowledge of the underlying mechanisms will be valuable to non-governmental organizations and the general public, as they will know what causal mechanisms to assess. This will help in keeping government in-check on the OGD deliverable, by finding opportunities and gaps, and possible ways of addressing them, which are discussed with the government for possible action.

Practitioners involved in information systems evaluation studies can also evaluate the merits of adopting a critical realist metatheory over conventional approaches. This study helps in demonstrating the importance of iteration versus a linear approach founded on assumptions of universal regularities and repetition to information systems study.

11.4 Recommendations for Future Research

There are a few gaps on OGD institutionalization that follow from our findings, which would benefit from further research;

1. A comparative study between KODI and another country that has successfully institutionalized OGD to help in identifying gaps, especially from a structure perspective. This would require the researcher to apply the same data collection and analysis instruments. The findings will help in providing further validity of the findings of this study.
2. There is need to assess evaluation as a supportive structure for KODI. This could help in monitoring the progress of the initiative, and identifying bottlenecks before a crisis emerges. It would also promote ownership as it provides a chance for stakeholders to learn more about the existing structures, which are mainly policies and laws.
3. It would also be helpful to study the individual ministries such as health, agriculture and treasury in relation to how they are institutionalizing open data. This study looked at the national context, and though representatives from these agencies were interviewed, it would help to delve deeper into understanding the underlying dynamics. Devolution of government had an implication in some of these ministries, where their operations were decentralized to the forty-seven sub-national governments. This adds complexity since operations among the various sub-national governments is yet to be standardised. Even though one may be successful in formulating and implementing supportive policies in one sub-national government, those policies and implementation strategies may not work

in the rest as they could be other constraining structures and mechanisms that were not factored in the initial strategies. This implies that it may not be feasible to design a CMO configuration with the expectation that it will apply across different contexts.

12. References

Al-kubaisi, A. S. (2014). *Openness and efficiency of e-Government in Qatar in the era of open government data, and beyond*. Queensland University of Technology. Retrieved from http://eprints.qut.edu.au/75540/2/Ali_Al-Kubaisi_Thesis.pdf

Albano, C. S., & Reinhard, N. (2014). Open government data: Facilitating and motivating factors for coping with potential barriers in the Brazilian context. In *International Conference on Electronic Government* (pp. 181–193). Springer Berlin / Heidelberg. https://doi.org/10.1007/978-3-662-44426-9_15

Archer, M. S. (1995). *Realist social theory: The morphogenetic approach.* Cambridge University Press.

Archer, M. S. (2007). The trajectory of the morphogenetic approach: An account in the first-person. *Sociologia*, *54*, 35–47.

Archer, M. S. (2010). Morphogenesis versus structuration: On combining structure and action. *British Journal of Sociology*, *61*(SUPPL. 1), 225–252. https://doi.org/10.1111/j.1468-4446.2009.01245.x

Bates, J. (2012). "This is what modern deregulation looks like" : co-optation and contestation in the shaping of the UK's Open Government Data Initiative. *The Journal of Community Informatics*, *8*(2), 1–13. Retrieved from http://ci-journal.net/index.php/ciej/article/view/845

Bates, J. (2014). The strategic importance of information policy for the contemporary neoliberal state : The case of Open Government Data in the United Kingdom. *Government Information Quarterly*, *31*(3), 388–395. https://doi.org/10.1016/j.giq.2014.02.009

Benbasat, I., & Zmud, R. W. (1999). Empirical research in information systems : the practice of relevance. *MIS Quarterly*, *23*(1), 3–16.

Berger, P. L., & Luckmann, T. (1967). *The social construction of reality: A Treatise in the Sociology of Knowledge* (10th ed.). Penguin UK. https://doi.org/10.2307/323448

Bertot, J. C., Jaeger, P. T., Grimes, J. M., & Hansen, D. (2010). Using ICTs to create a culture of transparency: E-government and social media as openness and anti-corruption tools for societies. *Government Information Quarterly*, *27*(3), 264–271. https://doi.org/10.1016/j.giq.2010.03.001

Bhaskar, R. (1998). *The Possibility of Naturalism: A Philosophical Critique of the Contemporary Human Sciences (Critical Realism--Interventions)*. Routledge.

Bhaskar, R. (2008). *A Realist Theory of Science*. Routledge. https://doi.org/10.2307/2184170

Böhm, C., Freitag, M., Heise, A., Lehmann, C., Mascher, A., Naumann, F., … Schmidt, M. (2012). GovWILD : Integrating Open Government Data for Transparency. In *21st international conference companion on World Wide Web. ACM* (pp. 3–6).

Boyatzis, R. E. (1998). *Thematic analysis and code development: Transforming qualitative information*. London and New Delhi: Sage Publications.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Breitman, K., Salas, P., Saraiva, D., Gama, V., & Casanova, M. A. (2012). Open Government Data in Brazil. *IEEE Intelligent Systems*, *27*(3), 45–49.

Bygstad, B. (2010). Generative mechanisms for innovation in information infrastructures. *Information and Organization*, *20*(3–4), 156–168. https://doi.org/10.1016/j.infoandorg.2010.07.001

Bygstad, B., Munkvold, B. E., & Volkoff, O. (2016). Identifying generative mechanisms through affordances: A framework for critical realist data analysis. *Journal of Information Technology*, *31*(1), 83–96. https://doi.org/10.1057/jit.2015.13

Carlson, S. A. (2009). Critical Realism. In M. Dwivedi, Y. K., Lal, B., Williams, M. D., Schneberger, S. L. & Wade (Ed.), *Handbook of Research on Contemporary Theoretical Models in Information Systems* (pp. 57–76). IGI Global.

Chattapadhyay, S. (2013). Towards an expanded and integrated open government data agenda for India. In *Proceedings of the 7th International Conference on Theory and Practice of Electronic Governance - ACM* (pp. 202–205). New York, New York, USA: ACM Press. https://doi.org/10.1145/2591888.2591923

Chignard, S. (2013). A brief history of Open Data. Retrieved October 14, 2017, from http://parisinnovationreview.com/2013/03/29/brief-history-open-data/?media=print

Collier, A. (1994). *Critical realism: An introduction to Roy Bhaskar's philosophy*. Verso. Retrieved from https://philpapers.org/rec/COLCRA

Crabtree, B. F., & Miller, W. L. (1999). *Doing qualitative research*. SAGE Publications.

Dacin, M. T., Goodstein, J., & Scott, W. R. (2002). Institutional theory and institutional change: Introduction to the special research forum. *Academy of Management Journal*, *45*(1), 43–56. Retrieved from http://www.jstor.org/stable/3069284

Danermark, B., Ekstrom, M., Jakobsen, L., & Karlsson, J. C. (2002). *Explaining Society: Critical realism in the social sciences*. Routledge.

Dawes, S. S. (2005). A Realistic Look at Open Data. *Center for Technology in Government, University at Albany/SUNY*. Retrieved from http://www.w3.org/2012/06/pmod/pmod2012_submission_38.pdf

Dawes, S. S., & Helbig, N. (2010). Information Strategies for Open Government : Challenges and Prospects for Deriving Public Value from Government Transparency. *Electronic Government*, 50–60.

DiMaggio, P. J., & Powell, W. J. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*(2), 147–160.

Dube, L., & Pare, G. (2003). Rigor in Information Systems Positivist Case Research: Current Practices, Trends, and Recommendations. *MIS Quarterly*, *27*(4), 597–636.

Easton, G. (2010). Critical realism in case study research. *Industrial Marketing Management*, *39*(1), 118–128. https://doi.org/10.1016/j.indmarman.2008.06.004

Eisenhardt, M. K. (1989). Building Theories from Case Study Research. *The Academy of Management Review*, *14*(4), 532–550. Retrieved from http://www.jstor.org/stable/258557

Eriksson, O., & Goldkuhl, G. (2013). Preconditions for public sector e-infrastructure development. *Information and Organization*, *23*(3), 149–176. https://doi.org/DOI 10.1016/j.infoandorg.2013.04.001

Fereday, J., & Muir-Cochrane, E. (2006). Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development. *International Journal of Qualitative Methods*, *5*(1), 80–92. https://doi.org/10.1063/1.2011295

Flyvbjerg, B. (2006). Five Misunderstandings About Case-Study Research. *Qualitative Inquiry*, *12*(2), 219–245. https://doi.org/10.1177/1077800405284363

Fox, S. (2009). Applying critical realism to information and communication technologies: a case study. *Construction Management and Economics*, *27*(5), 465–472. https://doi.org/10.1080/01446190902842292

Freeman, M., deMarrais, K., Preissle, J., Roulston, K., & St. Pierre, E. a. (2007). Standards of Evidence in Qualitative Research: An Incitement to Discourse. *Educational Researcher*,

36(1), 25–32. https://doi.org/10.3102/0013189X06298009

Giddens, A. (1979). *Central problems in social theory: Action, structure, and contradiction in social analysis* (241st ed.). Univ of California Press.

Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday*, *16*(2), 9–11. Retrieved from http://firstmonday.org/ojs/index.php/fm/article/viewArticle/3316/2764

Heeks, R. (2002). *eGovernment in Africa: Practice and Practice* (No. 13). Manchester, UK. Retrieved from http://idpm.man.ac.uk/wp/igov/index.htm

Heeks, R. (2005). e-Government as a Carrier of Context. *Journal of Public Policy*, *25*(1), 51–74. https://doi.org/10.1017/S0143814X05000206

Henfridsson, O., & Bendik Bygstad. (2013). The generative mechanisms of digital infrastructure evolution. *MIS Quarterly*, *37*(3), 907–931.

Horrocks, I. (2009). Applying the Morphogenetic Approach. *Journal of Critical Realism*, *8*(1), 35–62. https://doi.org/10.1558/jocr.v8i1.35

Hoxha, J., & Brahaj, A. (2011). Open Government Data on the Web: A Semantic Approach. *2011 International Conference on Emerging Intelligent Data and Web Technologies*, 107–113. https://doi.org/10.1109/EIDWT.2011.24

Huijboom, N., & Broek, T. Van Den. (2011). Open data : an international comparison of strategies. *European Journal of ePractice*, *12*(March/April), 1–13. Retrieved from www.epracticejournal.eu

ICTA. (2016). Kenya ICT Authority Blog. Retrieved November 13, 2016, from http://www.icta.go.ke/category/blog/

Ilkka, N. (1991). Realism, Relativism, and Constructivism. *Scientific Knowledge and Scientific Practice*, *89*(1), 135–162.

Janssen, K. (2011). The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly*, *28*(4), 446–456. https://doi.org/10.1016/j.giq.2011.01.004

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. https://doi.org/10.1080/10580530.2012.716740

Jepperson, R. L. (1991). Institutions, institutional effects, and institutionalism. In *The new*

institutionalism in organizational analysis (pp. 143–163). Retrieved from https://books.google.co.za/books?id=jbTbAgAAQBAJ&printsec=frontcover&source=gbs_atb#v=onepage&q&f=false

Jetzek, T., Avital, M., & Bjørn-Andersen, N. (2013). The generative mechanisms of open government data. In *21st European Conference on Information Systems (ECIS 2013)*. Utrecht, Netherlands.

Kags, A. (2014). A look back at Kenya's open data jouney. Retrieved October 12, 2017, from http://www.openinstitute.com/a-look-back-at-kenyas-open-data-journey/

Kitchin, R. (2014). Open and Linked Data. In *Data Revolution* (pp. 48–66). Sage Publications. Retrieved from https://uk.sagepub.com/en-gb/afr/the-data-revolution/book242780#preview

Kraft, M. E., & Scott R. Furlong. (2015). *Public policy: Politics, analysis, and alternatives*. *Sage*. Sage Publications.

Kvale, S. (1996). The 1000-Page Question. *Qualitative Inquiry*, *2*(3), 275–284.

Lee, A. S. (1989). A Scientific Methodology for MIS Case Studies. *MIS Quarterly*, *13*(1), 33–50.

Meijer, A. (2012). Understanding the Complex Dynamics of Transparency Albert. *Public Administration Review*, *73*(3), 429–439.

Meyer, J. W. (2006). Reflections on Institutional Theories of Organizations. *The Sage Handbook of Organizational Institutionalism*, 788–809. https://doi.org/10.4135/9781849200387.n35

Miles, M., Huberman, M., & Saldana, J. (2014). *Qualitative data analysis: a methods sourcebook* (3rd ed.). Sage Publications. https://doi.org/10.7903/cmr.14043

Mingers, J. (2000). The contribution of critical realism as an underpinning philosophy for OR=MS and systems. *Journal of the Operational Research Society*, *51*(11), 1256–1270. https://doi.org/10.1016/j.infoandorg.2003.06.001

Mingers, J. (2002). Real-izing Information Systems: Critical Realism as an Underpinning Philosophy for Information Systems. In *International Conference on Information Systems* (pp. 294–303).

Mingers, J. (2004). Real-izing information systems: critical realism as an underpinning philosophy for information systems. *Information and Organization*, *14*(2), 87–103. https://doi.org/10.1016/j.infoandorg.2003.06.001

Morton, P. (2006). Using Critical Realism To Explain Strategic Information Systems Planning. *Journal of Information Technology Theory and Application (JITTA)*, *8*(1), 1–20.

Myers, M. D., & Klein, H. K. (2011). A set of principles for conducting critical research in information systems. *MIS Quarterly*, *35*(1), 17–36.

Nelson, R. R., & Winter, S. G. (1982). The Schumpeterian Tradeoff Revisited. *American Economic Review*, *72*(1), 114–132. https://doi.org/10.2307/1808579

Open Knowledge Foundation. (2012). The Open Data Handbook. Retrieved January 5, 2015, from http://opendatahandbook.org/guide/en/

Orlikowski, W. J., & Baroudi, J. J. (1991). Studying Information Technology in Organizations : Research Approaches and Assumptions. *Information Systems Research*, *2*(1), 1–28. Retrieved from http://www.jstor.org/stable/23010611

Ostrom, E. (2007). Institutional rational choice: An assessment of the institutional analysis and development framework. In *Theories of the Policy Process* (pp. 267–306). Westview Press.

Parycek, P., Höchtl, J., & Ginner, M. (2014). Open Government Data Implementation Evaluation. *Journal of Theoretical and Applied Electronic Commerce Research*, *9*(2), 13–14. https://doi.org/10.4067/S0718-18762014000200007

Pawson, R., & Tilley, N. (2004). Realist evaluation. *Changes*, 1–36.

Popper, K. (2014). Corroboration, or how a theory stands up to tests. In *The logic of scientific discovery* (pp. 248–283). Routledge.

Porter, S. (2015). Realist evaluation: An immanent critique. *Nursing Philosophy*, *16*(4), 239–251. https://doi.org/10.1111/nup.12100

Powell, W. W., & Colyvas, J. a. (2008). Microfoundations of institutional theory. In *The SAGE Handbook of Organizational Institutionalism* (pp. 276–298). https://doi.org/10.4135/9781849200387

Provost, C. (2011). Kenya opens its books in revolutionary transparency drive. Retrieved October 12, 2017, from https://www.theguardian.com/global-development/poverty-matters/2011/jul/13/kenya-open-data-initiative

Ranmuthugala, G., Cunningham, F. C., Plumb, J. J., Long, J., Georgiou, A., Westbrook, J. I., & Braithwaite, J. (2011). A realist evaluation of the role of communities of practice in changing healthcare practice. *Implementation Science*, *6*, 49. https://doi.org/10.1186/1748-5908-6-49

Robinson, D., Harlan, Y., Zeller, W. P., & Felten, E. W. (2009). Government data and the invisible hand. *Yale JL & Tech*, *11*(1).

Sayer, A. (1992). *Method in Social Science: a realist approach* (Second Edi). Routledge. https://doi.org/10.4324/9780203310762

Schutz, A. (1967). *The Phenomenology of the Social World. Northwestern University Press*. https://doi.org/10.2307/2092457

Schutz, A. (1971). *The problem of social reality*. Nijhoff.

Scott, W. R. (2004). Institutional theory. *Encyclopedia of Social Theory*, *1*(11), 408–414.

Scott, W. R. (2005). Institutional theory: Contributing to a theoretical research program. In *Great Minds of Management: The Process of Theory Development* (pp. 460–485).

Scott, W. R. (2008a). Approaching adulthood: The maturing of institutional theory. *Theory and Society*, *37*(5), 427–442. https://doi.org/10.1007/s11186-008-9067-z

Scott, W. R. (2008b). Institutions and Organizations Ideas, Interests, and Identities. SAGE Publications.

Selznick, P. (2011). Leadership in administration: A sociological interpretation. In *Quid Pro Books*.

Sewell, W. H. J. (1992). A Theory of Structure: Duality, Agency, and Transformation. *The American Journal of Sociology*, *98*(1), 1–29. https://doi.org/10.1086/229967

Shadbolt, N., Hara, K. O., Berners-lee, T., Gibbins, N., & Glaser, H. (2007). Linked Open Government Data: Lessons from Data.gov.uk. *IEEE Intelligent Systems*, *27*(3), 2–10.

Shah, N. (2013). Big Data, People's Lives, and the Importance of Openness. Retrieved October 14, 2017, from https://dmlcentral.net/big-data-people-s-lives-and-the-importance-of-openness/

Smith, M. L. (2006). Overcoming theory-practice inconsistencies: Critical realism and information systems research. *Information and Organization*, *16*(3), 191–211. https://doi.org/10.1016/j.infoandorg.2005.10.003

Steinmetz, G. (1998). Critical realism and historical sociology. A review article. *Comparative Studies in Society and History*, *40*(1), 170–186.

Tinati, R., Carr, L., Halford, S., & Pope, C. (2012). Exploring the Impact of Adopting Open Data in the UK Government. *Digital Futures*, *3*.

Tolbert, P. S. (1988). Institutional Sources of Organizational Culture in Major Law Firms. *Institutional Patterns and Organizations : Culture and Environment*, 101–113.

Tolbert, P. S., & Zucker, L. G. (1996). The Institutionalization of Institutional Theory. In

Handbook of organization studies (pp. 175–190).

Ubaldi, B. (2013). *Open government data: Towards empirical analysis of open government data initiatives* (OECD Working Papers on Public Governance No. 22). Retrieved from http://dx.doi.org/10.1787/5k46bj4f03s7-en

Walsham, G. (2006). Doing interpretive research. *European Journal of Information Systems*, 320–330. https://doi.org/10.1057/palgrave.ejis.3000589

Wiseman, A. W., & Baker, D. P. (2006). The Symbiotic Relationship between Empirical Comparative Research on Education and Neo-Institutional Theory. *International Perspectives on Education and Society*, *7*(January 2016), 1–26. https://doi.org/10.1016/S1479-3679(06)07001-0

Wynn, D., & Williams, C. K. (2012). Principles for conducting critical realist case study research in information systems. *MIS Quarterly*, *36*(3), 787–810.

Yin, R. K. (2002). *Case Study Research: Design and Methods* (Third). SAGE.

Yu, H., & Robinson, D. G. (2012). *The New Ambiguity of "Open Government"* (Vol. 178). Retrieved from http://ssrn.com/abstract=2012489

Zeuner, L. (1999). Margaret Archer on Structural and Cultural Morphogenesis. *Acta Sociologica*, *42*(1), 79–86.

Zucker, L. G. (1977). The Role of Institutionalization in Cultural Persistence. *American Sociological Review*, *42*(5), 726–743.

Zucker, L. G. (1987). Institutional theories of organization. *Annual Review of Sociology*, *13*(1987), 443–464.

13. Appendix 1: Semi-Structured Interview Guides

A. Data Fellows Interview Questions

Section 1

1. What was your role as a data fellow?
2. Have these roles been included as part of your annual Key Performance Indicators (KPIs)?
3. What data sets do you help in curating?
4. How does your team identify the open data needs of (external) users?

Section 2

1. Describe the workflow for curating and publishing data?
2. What documents support the workflow (request for data, activity signoff documents)?
3. What quality assurance measures are in place?
4. What measures are in place to ensure timely release of data?
5. What challenges do you encounter as a data fellow?
6. What other challenges have you/your team faced in the past?
    a. How were they addressed?

Section 3

1. What policies have been adapted within the institution you worked in to assist in implementing the initiative?
2. What amendments/improvements would you suggest on existing policies and workflow?
3. In what ways has open data disrupted operations in your agency?
4. What collaborations on OGD exists with other departments/ institutions?
5. Who else can you recommend for this interview?

B. Government Agency Staff Interview Questions (Management)

Section 1

1. What was the origin of the Kenya Open Government Data initiative (KODI)?
2. What/who were the driving forces that led to the conceptualization, development and implementation of the KODI portal?
3. KODI Stakeholders prior 2015
    a. In your view, who are the key stakeholders?
    b. Why do you find them to be so critical?
    c. How did you or your team engage with them?
    d. What were their roles?

Section 2

1. What is/was the focus of your work around KODI?
2. What processes were affected by the KODI?
3. Were KODI related tasks been included in staff performance contracts?
4. What were the employees' perceptions about changes that resulted from KODI?

Section 3

1. What were the key achievements from the KODI?
2. What were the most prevalent disabling forces/challenges in the implementation process prior 2015?
3. What laws and policies were formulated or adapted to support OGD implementation?
4. What were the adopted sensitization and engagement channels with citizens and NGOs?

Section 4

1. How did your team identify the OGD needs of external users?
2. What was the objective behind publication of OGD?
    a. Has this been realized?
    b. How is this measured?
    c. What challenges were encountered?
3. What was the source of funding/resources and what motivated the allocation?
4. What collaborations on OGD exists with other departments/ institutions?

Section 5

7. When and how was data sourced and converted to machine-readable format?
    a. What plans had been placed?
    b. What were the challenges?
8. How can one access logs on the usage of KODI data by the public since 2009?

Section 6

6. In what ways has KODI evolved?
    a. What drove this evolution?
7. Have your interpretations of what happened prior 2014 changed?
8. Have your expectations about KODI changed post 2014?
9. Is there any documentation on this project that you can share for the purpose of research?
10. Who else can you recommend for this interview?

C. Government Agency Staff Interview Questions (Non-Management)

Section 1

1. What is the focus and objective of your work in relation to open data?
2. What is the structure of the team that deals with open data?

Section 2

5. What data is your institution mandated to publish to the public?
6. What external policies support this mandate?
7. What internal policies have been formulated to support this mandate?

Section 3

9. Describe the workflow for curating and publishing data?
10. What documents support the workflow (request for data, activity signoff documents)?
11. What quality assurance measures are in place?
12. What measures are in place to ensure timely release of data?
13. What are the existing challenges in the open data initiative?
14. What other challenges have you/your team faced in the past?
    a. How were they addressed?

Section 4

11. What policies have been adapted within your institution to assist in implementing the initiative?
12. What amendments/improvements would you suggest on existing policies and workflow?
13. In what ways has open data disrupted operations in your agency?
14. What collaborations on OGD exists with other departments/ institutions?
15. What is the source of funding for the open data initiative within your institution?
16. Who else can you recommend for this interview?

D. CSO and Private Sector Interview Questions (Users)

The following questions were targeted at civil society organization (CSO) actors and members of the private sector who are engaging with open data.

Section 1

4. In your view, what was the origin of the Kenya Open Government Data initiative (KODI)?
5. What/who were the driving forces that led its implementation?

Section 2

5. What is/was the focus and objective of your work around KODI?
6. How do you identify the needs of citizens/your users?
7. What are your Open Data sources?
    a. What challenges do you face in sourcing for this data?
    b. How much work is required for you to clean this data before deriving meaning from it?
8. What is the source of funding for your initiative(s)?

Section 3

5. In your view, what were the key achievements of the KODI?
6. What are the most prevalent disabling forces/challenges of this initiative?
7. What laws and policies need to be developed/adapted to support this initiative?

Section 6

17. In what ways has KODI evolved?
    a. What has supported this development?
18. Have your interpretations of what happened prior 2014 changed?
19. What are your expectations about KODI?
    a. Have they changed since the initiative was launched in 2011?

E. Ethical Clearance

UNIVERSITY OF CAPE TOWN

Faculty of Commerce
Ethics in Research Committee
University of Cape Town Private Bag Rondebosch 7701
Email: kincaidharold592@gmail.com
Telephone: 071 823 7573

March 9, 2015

Paul Mungai Wando
Information Systems

Project title: CAUSAL MECHANISMS THAT ENABLE INSTITUTIONALISATION OF OPEN GOVERNMENT DATA IN KENYA

Proposal no. 32-2015

Dear Researcher,

This letter serves to confirm that this project as described in your submitted protocol has been approved.

Please note that if you make any substantial change in your research procedure that could affect the experiences of the participants, you must submit a revised protocol to the Committee for approval.

Regards,

Professor Harold Kincaid

H. Kincaid

Commerce Faculty Ethics in Research Committee

F. Consent Form

Department of Information Systems
Leslie Commerce Building
Engineering Mall, Upper Campus
OR
Private Bag X3 - Rondebosch - 7701
Tel: +27 (0) 21 650 2261 Fax: +27 (0) 21650 2280
Internet: http://www.commerce.uct.ac.za/informationsystems/

26th November 2014

Dear <Participant Name>

I would like to invite you to participate in an academic research case study on identifying the context and causal mechanisms that enable proper institutionalization of OGD initiatives in Kenya. This research has been approved by the University of Cape Town (UCT)'s Commerce Faculty Ethics in Research Committee, and is endorsed by <Research authorisation body in Kenya> and <Head of Department in Kenya>.

The aim of this study is to gain understanding and insight into the continued usage and diffusion of business intelligence and analytics solutions within <Organisation Name>, and to identify possible areas (or "gaps") for future consideration, by conducting one-on-one interviews among selected participants across the company over the next two to three months.

Your participation in this research is voluntary. All information will be treated in a confidential manner and used exclusively for the purpose of this study. No individual names will be recorded or published. You will not be requested to supply any identifiable information, ensuring anonymity of your responses. You can choose to withdraw from the research at any time for whatever reason, in accordance with ethical research requirements.

The one-on-one interviews will take approximately 45 minutes at <Organisation Name>s premises in <Location, City, Kenya>. If you are willing to participate in this study, kindly sign the attached form and return to me at your earliest convenience.

Should you have any questions regarding this research, please feel free to email me on wandopm@gmail.com.

Thank you for your time and participation.

Sincerely,

Signed

Wando, Paul Mungai
PhD Student
Department of Information Systems
University of Cape Town
Email: wandopm@gmail.com

Professor Jean-Paul Van Belle
Research Supervisor
Department of Information Systems
University of Cape Town
Email: jean-paul.vanbelle@uct.ac.za

Research Participant Consent Form

I, _____, consent to participate in the research on identifying the context and causal mechanisms that enable proper institutionalization of OGD initiatives at <Organisation Name> in Kenya.

I am aware that participation is voluntary and that I may choose to withdraw from this study at any time, should I choose to do so.

_____ _____

Signature Date

G. Sample Interview Transcript

Interviewer: Maybe you can repeat again for me your role and how you…what's your relationship…your mediating role between Treasury and ICT Authority?

Interviewee: … I am the assistant director and by the virtue of the Kenya Gazette, ICT function falls under the ICT Authority and I am deployed to Treasury resource mobilization department where I am the database development manager. We source money from across all the donors, all the development partners where the government signs a financing agreement then we incorporate that budget with the national budget.

Interviewer: Most specifically the Open Data project what has been your key function?

Interviewee: With the Open Data Initiative, we have partnered for some time, they started off through technical assistance project from the World Bank where I happen to have worked with a team to develop a certain application called eProMIS, Electronic Project Monitoring Information Systems that's being used to across the whole government for monitoring purpose of projects. So, we realized Geo coding is very important particularly it's a good functionality when it comes to monitoring, Geo codes for spartial information...

We can Geo code every project [inaudible] a government, it will agree a lot with monitoring because we'll be able to give using Google Maps or Google Earth. The project on the ground while you're still in the office and you could comfortably compare what we're supposed to approve with what is existing on the ground. So, we felt we needed to improve on the data that we had on that application that's why we approached the World Bank, they assisted us with a certain object to Geo code.

At that point, the ICT Authority was a project, Kenya Open Data Initiative which is still World Bank, so we were linked up. After that we have worked quite closely with the Kenya Open Data Initiative where we supply some data, we publish our data on development budgets in that directorate. We avail them with data whenever they request in partnership, that has been our relationship. They have also assisted us in that application to train through the same project in most of the functions.

Interviewer: When it comes to the data, what's usually the flow is there a schedule for releasing data to ICT Authority that you've adopted?

Interviewee: We haven't agreed on the schedule but it's automatically known, our budget cycle is periodical. When we're done with a certain phase they just call, and avail them the data, particularly with a budget. We haven't published a lot, we started with a little amount of data like the development budget allocations. So we're starting with the development budget which we do after every cycle. When we're done, he is very aggressive, he calls and we email it to him, he goes format it at [0:04:23]...

Interviewer: They require to format?

Interviewee: Yeah. So, it's periodical and we hope to improve that in future, we will be able to avail more data like now expenditure which can be done at whatever period.

Interviewer: Do you feel that there's some extra work that you need to do for you to be able to give them the data as per their need or you just submit the data as it is?

Interviewee: Fortunately when you have a database you just shoot a query and you generate a data as required so, to me it's not a burden. In fact, we really appreciate them, because that that website is an outlet to present our data from our application to the people, people who are not in government, to us that was a big boost for us to market our government data which has already been lying idle, you see all these books here they have a lot of data, and nobody knows where to access the data.

When we publish it in the KODI website we've come to learn that there are a lot of hits, people access that data [inaudible]...

Interviewer: And cheaper because they do not print?

Interviewee: Yeah, it is responsive, and in that case we reach more people that we normally interact with there.

Interviewer: Is there any, now when it comes to the politics that govern this particular process, are there any that mandate Treasury to publish data to the public or what guides you to do it?

Interviewee: What a very good question, in actual fact that was one of the challenges that we had when we started partnering with the KODI team, we had to be very cautious. You know the red-tape, we haven't had any clear guidelines on how we're supposed to release the government data to the public. Why? The other reason why we were comfortably because we're dealing with a government of the government, the Ministry of Information is actually a government agency.

So when handed them over that data us we're handing over to a government institution, so them they have their own guideline they publish but I agree with you, we need the [inaudible] way. There's a lot of data here in the government office which can be published as long as there's a clear guideline.