

Causal mechanisms and institutionalisation of open government data in Kenya

Paul W. Mungai

University of Cape Town, Cape Town, South Africa

Correspondence

Paul W. Mungai, University of Cape Town, Cape Town, South Africa.

Email: wandopm@gmail.com

Abstract

Open data—including open government data (OGD)—has become a topic of prominence during the last decade. However, most governments have not realised the desired value streams or outcomes from OGD. The Kenya Open Data Initiative (KODI), a Government of Kenya initiative, is no exception with some moments of success but also sustainability struggles. Therefore, the focus for this paper is to understand the causal mechanisms that either enable or constrain institutionalisation of OGD initiatives. Critical realism is ideally suited as a paradigm to identify such mechanisms, but guides to its operationalisation are few. This study uses the operational approach of Bygstad, Munkvold & Volkoff's six-step framework, a hybrid approach that melds concepts from existing critical realism models with the idea of affordances. The findings suggest that data demand and supply mechanisms are critical in institutionalising KODI and that, underpinning basic data-related affordances, are mechanisms engaging with institutional capacity, formal policy, and political support. It is the absence of such elements in the Kenya case which explains why it has experienced significant delays.

KEYWORDS

critical realism, institutionalisation, Kenya Open Data Initiative, open government data

1 | INTRODUCTION

Open data refers to “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 open data and is expected to be accountable and objective in providing this service in the interest of the entire society. Other sources of open data include the private sector, civil society organisations, and non-governmental organisations. The focus of this study is open government data (OGD),¹ a trend stimulated by the Open Government Partnership movement, which requires member states to be open and transparent.²

Open government data comprises a wide array of datasets including business information such as registers, patents, and trademark information. Other government datasets include public tender records, geographic information, legal information, meteorological information, transport information, and statistical data on economics, employment, health, population, and public administration (Ubaldi, 2013). Open data such as OGD has been presented as a resource with the potential of generating social, environmental, and economic value (Jetzek, Avital, & Bjørn-Andersen, 2014). This includes innovation of public services. However, even with the availability of a few datasets, models on how communities can use

¹This study will use the phrase “open data” in cases where the data or discussion is not limited or specific to government data, and will use “OGD” for discussions specific to government open datasets.

²For example, Kenya has been a member of the OGP since 26th August 2011 (Open Government Partnership, 2011).

open data collectively to drive social, environmental, or economic change at the local level are still scarce (Davies & Lithwick, 2010). The question that persists is “how can the use of open data stimulate the generation of sustainable value?” (Jetzek et al., 2014), a question that draws towards the issue of institutionalisation: referred to here as the process of establishing an OGD initiative within government by creating structures that guarantee its sustainability.

Most of the governments that have institutionalised their OGD initiative—though not necessarily being successful in terms of wider impact of such initiatives—are found in Europe, North America, and Australia (Janssen, 2011; Shadbolt, Hara, Berners-Lee, Gibbins, & Glaser, 2007; Ubaldi, 2013; Yu & Robinson, 2012). There are a few recent OGD initiatives in Africa, such as in Kenya and Morocco, that have experienced varying sets of challenges. This could be explained from a stakeholder perspective. For instance, Morocco's initiative was purely government driven, while in Kenya, government buy-in was secured through World Bank intervention during the initial planning and implementation phase.

Since stakeholders are an important part of context, this suggests the need to factor in contextual differences as having an impact on the processes and outcomes of OGD (Heeks, 2002). As another example of contextual impact, 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. It is therefore necessary to avoid generalisation when assessing the underlying causal mechanisms that affect OGD institutionalisation in a particular country and thus necessary to adopt a methodological approach allowing for contextual specificities.

This study is specific to the Kenyan context, and more particularly the Kenya Open Data Initiative (KODI), an OGD initiative, which began in 2011, and only deals with official government data. The research question for this study is as follows: what are the causal mechanisms that enable institutionalisation of the KODI? This question requires an understanding of social structures that affect the institutionalisation process. In this context, social structures comprise individuals, groups, organisations, and a set of rules and practices including technological artefacts. The existence of social structures is dependent on their agents' interpretation of the activities they are engaged in within those structures. They are capable of either constraining or enabling social activities, and are themselves reproduced or transformed by these activities (Wynn & Williams, 2012). It also requires an understanding of the emerging outcomes, which includes their associated context and mechanisms. To achieve this, critical realism was adopted as the research paradigm. The adoption of a critical realist paradigm helps in understanding both the objective reality and the subjective interpretations of the people involved in any process, and also helps in allowing for contextual differences.

The following section provides details of this study's methodology, including a brief description of critical realism and definition of various concepts used in this paper including affordances, causal mechanisms, and structures. The section also includes a description of the approach taken in conducting data analysis. The third section provides an explanation of Bygstad, Munkvold, and Volkoff's (2016) six-step framework, which was the preferred methodological model for this study. The fourth section provides a detailed description of the findings based on the six-step framework. One of the aims of this paper was to demonstrate the practicality of this framework for similar studies, by providing a detailed description of how each step was applied in this study. This was felt necessary because of the relative scarcity of empirical work based on critical realism in the wider discipline of information systems (Henfridsson & Bygstad, 2013), let alone the more specific sub-discipline of ICT4D, in which application of critical realism has been very limited to date (see Heeks & Wall in this special issue). The final section provides a conclusion for the study.

2 | METHODOLOGY

This is a qualitative study whose underlying philosophical assumptions were informed by critical realism. The process of data collection and thematic analysis was guided by the work of Fereday and Muir-Cochrane (2006). Primary data were collected through semi-structured interviews with various stakeholders as outlined in Table 1 below, and supporting documents were obtained during interviews with these stakeholders. Examples of such documents include copies of the KODI data release form, training documents, progress reports, and dataset work plans. The researcher applied purposive sampling in identifying the interview candidates. This was based on the knowledge of the main actors from literature and recommendations by other actors during interviews or email correspondence, especially in cases where someone felt their colleague was better placed to provide certain information about KODI.

TABLE 1 Stakeholder interviews

Stakeholders	Number of Interviews
Researchers	2
Government policymakers	6
County government official	1
KODI team members	5
KODI data fellows	5
World Bank staff members	3
Civil society organisation staff	7
@iLabAfrica (implementing partner) staff	2
Total	31

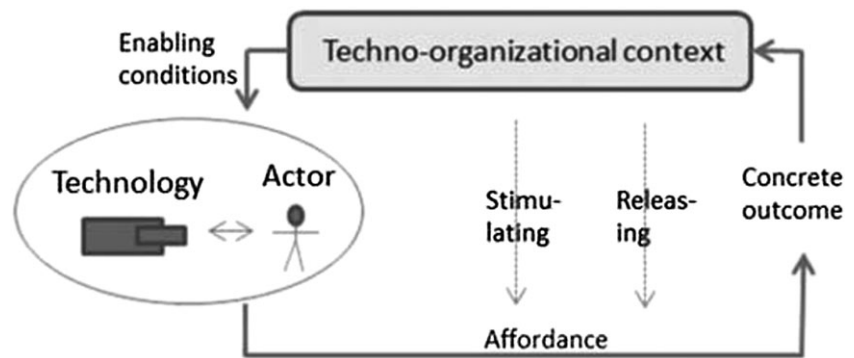


FIGURE 1 Structure of an affordance (Bygstad et al., 2016)

Secondary data comprised media report analysis where relevant newspaper articles and news feeds from television were reviewed.³ The multiple data sources assisted in evidence triangulation and convergence.

2.1 | Philosophical approach

This study applies critical realism as its underlying philosophy. Critical realism is based on the notion that events should be investigated in terms of generative mechanisms that cause those events, not simply in terms of constant conjunction of regular events: the argument being that establishing a constant conjunctive relationship is not sufficient to understand the causal connection between mechanisms and events (Easton, 2010; 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 help reveal causal effects. This differs from empiricism, where events are selected based on perceptibility—the notion that only that which can be perceived can exist (Easton, 2010; Mingers, 2002).

Events and mechanisms can further be explained through stratification: understood from Bhaskar's explanation that reality is both intransitive (existing independent of humans) and stratified. The first form of stratification is between mechanisms, the events that result from those mechanisms, and the sub-set of those events that are experienced. This can be understood across three domains: the real, the actual, and the empirical. The empirical domain contains experiences that may be observed; the actual domain contains events that may or may not occur (the latter being non-events which by definition cannot be observed or experienced (Mingers, 2002)); and the actual domain contains events that may not be observable at all or even when they are, different observers may understand them quite differently owing to the subjectivity of experience (Easton, 2010; Mingers, 2002).

The domains could be likened to an iceberg, where only a part is visible (even of the empirical domain), 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, ie, that no observation is infallible (Easton, 2010; Mingers, 2004). Thus, it is unlikely that observations made in the empirical domain will result in any full understanding of the social situation in question. In addition, there is no definitive criterion to judge the “truth” of a particular explanation. Therefore, there is need for the observer to collect sufficient data that will aid in distinguishing between 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 could be determined by consensus.

Unlike the positivist and interpretivist paradigms, critical realism acknowledges the role of subjective knowledge of social actors but then also allows for the existence of independent structures/mechanisms within a social setting that act as either constraints or enablers (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 (Mingers, 2002; Wynn & Williams, 2012). Once these interpretations are expressed, they cease to be wholly subjective, and become available for investigation, debate, or judgement by others (Mingers, 2002).

Critical realism has three main benefits to information systems research generally and in ICT4D research specifically (Wynn & Williams, 2012; see also Heeks & Wall in this special issue). First, it helps in transcending a number of inconsistencies between stated philosophical assumptions and the actual practice of research under both positivism and interpretivism. Second, it offers a way to address the rigour-relevance gap in research following its approach to causal analysis through multi-method/triangulation and multi-level approaches. 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.

2.2 | Models for an explanatory social science

The aim of social science is to explain social conditions, something which critical realism attempts to achieve. Critical realism has been associated with several explanatory or operational models, including:

³Most of Kenya's media houses publish some television news segments and documentaries via YouTube which makes them easy to analyse.

- The Popper-Hempel explanatory model is based on David Hume's (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. Its central modes of inference are deduction and induction (Danermark, Ekstrom, Jakobsen, & Karlsson, 2002).
- DREI refers to description of law-like behaviour, retroduction, elaboration, and elimination of alternative explanations, and identification of underlying mechanisms. DREI uses transcendental arguments to enable experimentation in social science (Mingers, 2000; Steinmetz, 1998).
- RRRE is defined as resolution of complex events into their components, redescription of components based on theory, retroduction to antecedent causes of the components, and elimination of alternative possible causes of the components (Bhaskar, 1998; Carlson, 2009).
- Margaret Archer's morphogenetic approach aims at understanding emergence and analytical dualism between structure and agency. It consists of a three-stage cycle which involves structure, culture, and agency. Archer's approach helps in explaining 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).
- Pawson and Tilley's realist evaluation model seeks to provide an understanding of how programmes bring change by probing the change apparatus within them (Pawson & Tilley, 2004).
- Danermark's explanatory model incorporates central modes of inference via abduction and retroduction (Danermark et al., 2002).

While these models vary in their particular purpose and role within critical realism, most could provide some methodological basis for operationalisation. However, these were not selected. Instead, Bygstad et al.'s (2016) six-step framework model—description of events and issues, identification of key entities, theoretical re-description (abduction), identification of candidate affordances (retroduction), analysis of the set of affordances and associated mechanisms, assessment of explanatory power—for critical realist data analysis was used. This is a hybrid model that borrows concepts from the models described above. This model was chosen because it identifies and addresses two issues of particular relevance to studying ICT-related cases:

- The identification of mechanisms. Mechanisms are causal structures explaining a phenomenon (Bhaskar, 1998). Bygstad et al. (2016) argue that the challenge for researchers dealing with ICT-based cases is understanding what a mechanism is, finding ways to identify and describe the mechanism, and determining how to evaluate a proposed mechanism.
- The specific role of technology in mechanisms. ICT-based cases involve social and technical entities, which creates the need to identify the elements of socio-technical mechanisms and how such mechanisms interact. Again, Bygstad et al. (2016) identify this as challenging to operationalise.

To address these issues, they argue for use of the concept of affordances as the analytical construct for identifying and analysing mechanisms. Given the broad definition of an affordance as a "possibility for action" (Bygstad et al., 2016: p.84), one can see the similarity to the notion of mechanism as a latent potential to act (eg, Bhaskar, 1998). But the specific relevance of affordances is their development as a conceptual idea linked to ICTs, with Bygstad et al. (2016: p.87) giving the more specific definition as, "the potential for behaviours associated with achieving an immediate concrete outcome and arising from the relation between an object (eg, an IT artefact) and a goal-oriented actor or actors".

Further detail on Bygstad et al.'s (2016) framework is provided in Section 3 below.

2.3 | Thematic analysis

Fereday and Muir-Cochrane's (2006) approach to thematic analysis and code development was applied for abduction. It is a development of the work of Boyatzis (1998) and Crabtree and Miller (1999) who developed a guideline for 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). The stages of the Fereday and Muir-Cochrane approach are developing the code manual from literature, testing the reliability of the codes using sample data, summarising data and identifying preliminary themes, applying the template of codes and additional coding, connecting the codes and identifying themes, and corroborating and legitimating code themes (Fereday & Muir-Cochrane, 2006). This model assisted in identifying, analysing, and reporting on emerging themes or patterns. This model also provided a systematic method for explaining how the identified themes were formulated from the raw data to uncover meanings related to institutionalisation of the OGD initiative. Readers are referred to the original sources for further details.

3 | STEPWISE FRAMEWORK FOR CRITICAL REALIST DATA ANALYSIS

Bygstad et al.'s (2016) framework is based on concepts borrowed from critical realism models and literature including Danermark et al.'s (2002) critical realist explanatory model, Wynn and Williams' (2012) realist evaluation model, and the work of Sayer (1992), Volkoff and Strong (2013),

and De Landa (2006). Table 2 provides a description of this framework. The sequence of steps illustrates its analytical structure and assists in outlining the principles for conducting critical realist data analysis (Bygstad et al., 2016). Section 4 then provides a description of how this framework was applied in the KODI study.

TABLE 2 Critical Realist Data Analysis Model (from Bygstad et al., 2016)

Step	Explanation
1. Description of events and issues	In a critical realist context, events are clusters of observations understood at different levels of granularity dependent on the phenomenon of interest (Bygstad et al., 2016; Sayer, 1992). The researcher, potentially with the help of informants, identifies the events deemed relevant to the study. For example, in an organisational study, the main objective of this step might be to understand organisational changes resulting from the introduction of new artefacts and policies into the existing workflow (Bygstad et al., 2016). (Bygstad et al. (2016) provide no explicit explanation of "issues", but they may be understood as the pre-cursor or surrounding issues to events, eg, the problem or opportunity that was seen as leading into an event.)
2. Identification of key entities	"The key entities are the objects of the case, for example individuals, organizational units, technology and the relationships between them. Together they constitute structures, ie, networks of entities, with causal powers. Entities may be identified from data, in a grounded way or they may be embedded in a theoretical framework." (Bygstad et al., 2016: p.89).
3. Theoretical re-description (abduction)	Bygstad et al. (2016) describe this step as that of abstraction, where different theoretical perspectives and explanations of the identified entities and events are explored as a step towards identifying mechanisms. The researcher needs to choose relevant theories that will assist in the observation, description, interpretation, and explanation of the events and underlying causal mechanisms (Bygstad et al., 2016; Danermark et al., 2002).
4. Retroduction: identification of candidate affordances	This step is a continuation of theoretical re-description. Retroduction is a "mode of inference in which events are explained by postulating (and identifying) mechanisms which are capable of producing them" (Sayer, 1992: p.107 cited in Bygstad et al., 2016). Mechanisms may be seen as an outcome of the interaction between several affordances and thus an affordance is "not the endpoint of mechanism identification and analysis, but a building block of a more complex mechanism (or mechanisms)" (Bygstad et al., 2016: p.88). Bygstad et al. (2016: p.89) use the work of Volkoff and Strong (2013) to provide an outline on how to postulate and explain candidate affordances: <ul style="list-style-type: none"> (1) Identification of immediate concrete outcomes: "something that is directly achieved (or could be achieved) through the use of technology, and is related to the realization of the actor's goals". (2) Analysis of the interplay of human and technical entities. This interplay is complex, mainly because technical objects have "deep technical structures, such as data structures, coded transactions, and authorization tables, and surface structures, such as interfaces". (3) Identification of candidate affordances, which arise "from the relation between human/social and technical entities and reflects the potential for behaviour that may produce the immediate concrete outcome". Like mechanisms, affordances are seen as real, enduring, and potentially never being actualised. (4) Identification of stimulating and releasing conditions. Contextual conditions determine whether, and how, an affordance is actualised, and can either be stimulating or releasing conditions. "Stimulating conditions are typically organizational arrangements that make it easier to act. Releasing conditions are often specific decisions". Figure 1 provides an overview of this.
5. Analysis of the set of affordances and associated mechanisms	After understanding the contextual conditions that stimulate affordances into action, Bygstad et al. (2016: p.89) suggest "First we need to analyse the interaction of affordances, and the dependencies between them. Some have temporal dependencies, while others have structural or complementary dependencies. Next, we should try to understand how affordances are parts of higher-level mechanisms: Macro-micro and micro-macro mechanisms".
6. Assessment of explanatory power	Bygstad et al. (2016: p.89) advise that any proposed mechanisms identified in the previous step "should be treated as a candidate explanation, and the data collection and analysis should be repeated until closure is reached. We search for the mechanisms with the strongest explanatory power related to the empirical evidence, ie, the causal structure that best explains the events observed".

4 | FINDINGS

This section provides a description of how Bygstad et al.'s (2016) six-step framework was applied in conducting data analysis and identifying the causal mechanisms for KODI through affordances. This section is organised according to the steps of the framework.

4.1 | Description of events and issues

A number of events and issues affecting KODI were identified through a series of semi-structured interviews and document reviews as described above. This helps in providing a detailed description of KODI, followed by the identification of critical innovations, and the social and technical structures contributing to the identified events and innovations. Technical structures are necessary in identifying the role of technology in the actualisation of the underlying mechanisms. This information helped in understanding the current workflow and determining the various changes that are emerging from the introduction of the open data phenomenon, which includes technology, laws, and policies. These events and changes are illustrated in a chronological order at the end of this section.

4.1.1 | Introduction of KODI

Kenya Open Data Initiative was formed in 2011, following a successful drive by the then Cabinet Secretary for ICT, Prof. Bitange Ndemo. This initiative required approval from the Head of State, which was eventually obtained after much deliberation. The presidency perceived this to be similar to WikiLeaks, and that its main aim was to cause political instability. This myth was demystified through arguments demonstrating its role in creating jobs for the youth and improving government service delivery.

The initiative started off without the necessary structures, which include policies, job structures and budget, and buy-in from government agencies. In addition, there were no supportive open government directives, laws, and policies that gave the ICT Authority the mandate to collect data and publish government data, which is scattered in 83 government agencies.

This initiative was greatly supported by the World Bank, which saw this as a good measure towards transparency and accountability of their grants and loans. Government on the other hand perceived this as a western initiative, resulting in resistance and low uptake since they thought that the initiative had another agenda which they were not yet aware of. What helped is the buy-in from the permanent secretary in the Ministry of Information and Communications. He helped in securing buy-in from the President and other government agencies including the National Bureau of Statistics.

Unfortunately, the project came to a halt just a year after the initiative began in 2011. Change in government was the greatest contributor since some of the officials including the permanent secretaries were either reshuffled to another post or not reappointed. This was the case for the permanent secretary in the Ministry of Information and Communications. The technical team, which was placed under the ICT Authority, faced capacity issues even before these reshuffles took place. They had only appointed one person on a full-time basis to work on KODI. This person served as a project co-ordinator, and unfortunately during this shakeup, the person terminated their contract to pursue further studies abroad. The ICT Authority did not set up another team until the start of 2014.

Unfortunately, open data was not prioritised as a key deliverable by either the new Minister or permanent secretary at the Ministry of Information and Communications. The CEO of the ICT Authority also failed to keep the project running, despite having sufficient funding from the World Bank. The World Bank had given a sizeable grant that was expected to end in December 2016. It would have been expected that between 2011 and 2016, the Ministry of Information and Communications⁴ would have secured a budget to ensure project continuity post-2016. Unfortunately, none of this was achieved by the end of 2016, and the team which had been established in 2014 had to cease work for lack of budget to run KODI. Although this team had achieved a lot, it was not possible for them to have attained the goals of KODI in a period of just under 2 years (2014-2016).

4.1.2 | What are the enabling technologies?

The most critical technological component is the KODI portal. The portal hosts government data, and in particular demographic, statistical, budget, and expenditure data. Tabula and Pentaho are the two main tools used by KODI and other government agencies to curate data. Tabula, an open source tool, helps in extracting rows of data from PDF documents (Tabula, 2017). Pentaho helps in data integration including analysis, visualisation, reporting, etc. (Pentaho, 2017).

The KODI portal, which can be accessed at www.opendata.go.ke, operates on Socrata, a commercial platform that streamlines data publishing, management, analysis, and reuse (OpenDataMonitor, 2015). This software is owned by a Seattle-based cloud software company that focuses on developing cloud-based solutions that help in making government data more “discoverable, usable, and actionable” (Socrata, 2011). In 2011, the Kenya developer community raised their concerns against procurement of the platform from a foreign company. The argument was that there was enough local talent to build such a platform, and that Kenya had a lot of talent but very few job or business opportunities, especially for the youth. The World Bank was the main financier and made the argument that there was not much time to develop this application. This argument was

⁴Renamed the Ministry of Information, Communication, and Technology in 2013.

supported by the timeline given by the President to launch the initiative. Unfortunately, this becomes another example where aid is given to Africa but ends up benefiting the West more financially during implementation.

Kenya Open Data Initiative maintained a cordial relationship with Socrata, and collaborated in identifying user requirements and building additional features to meet these demands. For instance, the first version of the KODI portal did not have a mechanism where people could request help or additional datasets. The request-a-dataset feature was later introduced, allowing citizens to suggest new datasets. At times, the datasets would exist, only that the user could not find the proper key terms in the search. The team at KODI would then point users to the required datasets.

Some government agencies preferred to publish their data on their own portal—this despite support from KODI. KODI did not discourage this move, since their main aim was making government data open, and not simply populating their own database. An interesting observation was made by some of these agencies. They realised that KODI received a lot of traffic, in comparison to the low traffic on their portal. Some of them opted to push their data to KODI, while others requested KODI to introduce links, which would point users to their portal.

The KODI portal gave government agencies more visibility, as more people were able to engage with their data. Thirty-one out of 83 government agencies had agreed to release some of their data to the public. The portal also provided citizens with an interface that was easier to use, including searching for data, and interpreting emerging patterns through embedded visualisation tools on the Socrata platform. The KODI portal remains the most visited government portal in Kenya.

4.1.3 | What are the events and actors linked to these innovations?

In an effort to promote further openness and use of OGD, KODI supported numerous bootcamps, workshops, data journalism courses, and advocacy campaigns for the formation of open data laws and policies. In some cases, KODI partnered with other stakeholders in the design and implementation of these initiatives. These are listed in Table 3. In addition to these activities, KODI started a data fellows programme, which was meant to provide technical training and support to other government agencies.

Organisations like Google and KEPSA were only involved in the events leading to the launch of KODI in 2011. Other institutions shifted their support, for instance the World Bank stopped providing technical support after the launch but continued providing financial support until December 2016.

There are organisations listed above which were already publishing open data even before the launch of KODI. These include the Ministry of Health which has one of the most up-to-date data portals in Kenya, and KNBS although this was limited as most of their publications are not in open format. KNBS has since invested additional resources and partnered with KODI and @iLabAfrica to make their statistics open, and provide visualisations for ease of interpretation. KODI has published some of the statistics by KNBS on their portal, although KNBS has preferred to publish most of its data on their own portal, which is technically supported by @iLabAfrica and financed through a World Bank grant.

The data fellows programme involved teams of two: a communications expert and a technical support person per government agency. Their collective role, with support from KODI central office was to help these agencies identify datasets that could be of interest to the public and build capacity on how to curate this data for publication.

Despite the success in publishing a few hundred datasets, several organisations interested in the use of open data in their operations cited various challenges. For instance, Code4Kenya cited inconsistency of government datasets with more than one source, which required efforts to verify among the government agencies involved. They also stated that some of the data required significant additional effort prior to use, as some of the datasets were not uploaded in machine readable format. This challenge was also recorded by Mzalendo Watch, who reported the lack of willingness by the National Parliament to issue them with machine readable datasets. This forced them to find alternative, innovative ways to scrape data from the Parliamentary website, which was mostly in PDF format. This approach of obtaining data by Mzalendo Watch came after numerous unsuccessful attempts to access this data through KODI data request channels.

KODI, together with civil society organisations like Civicus, Mzalendo Watch, and Open Institute helped in pushing the agenda to enact the Access to Information Bill between 2011 and 2016. The Bill was enacted into law by the President in September 2016. It helps in resolving the

TABLE 3 List of stakeholders

Group	Stakeholders
Government	Ministry of Information and Communication, Kenya ICT Authority, Ministry of Planning specifically Kenya National Bureau of Statistics (KNBS) and National Treasury, Ministry of Industrialisation, National Transport and Safety Authority, Ministry of Education, Ministry of Health, National Council for Law Reporting, Kenya National Audit Office, Kenya Chief of Staff for National and County Government, Kenya Inter-Governmental Technical Relations Committee, Ministry of Agriculture
Inter-Governmental Organisations	The World Bank, African Centre for Statistics—UN Economic Commission for Africa, Open Government Partnership
Private Sector	Google, Kenya Private Sector Association (KEPSA), @iLabAfrica based at Strathmore University, Kenya tech startups like Data Science Ltd, Socrata
Non-Governmental Organisations (NGOs) and Civil Society	Africa Media Initiative, Civil Society working group, Global Open Data for Agriculture and Nutrition, Open Institute, Code4Kenya, Civicus, Mzalendo Watch—Eye on Kenyan Parliament

challenge of accessing data from government departments (Access to Information Act, 2016). At the time of writing, KODI was working with open data stakeholders to formulate the Kenya Open Data Policy, intended to address several issues, including giving it the official mandate to request data from government agencies on behalf of citizens.

Some of the observations made during interviews relating to poor progress on this policy were lack of clarity about purpose and content, and the lack of inclusive consultation and engagement with the various stakeholders. This would help in ensuring that most of the necessary considerations are factored in. This would help in formalising the mandate of the various stakeholders and building a more coordinated operational framework.

Kenya Open Data Initiative also faces high turnover, which has had a negative impact in its performance. The most recent was in December 2016, when the entire KODI team that was recruited in 2014 had their contracts terminated. Unfortunately, the ICT Authority had not managed to secure funding from government prior to this, despite the knowledge that the World Bank grant would only run until the end of 2016. This resulted in the loss of valuable skills, knowledge, and social capital especially in relation to engagement with other agencies.

The previous KODI team comprised a project manager, project coordinator, statistician charged with data acquisition, curation, and publishing, data analyst charged with analysing, visualising, and storytelling through KODI blog (<http://blog.opendata.go.ke/>), an open data and geographical information specialist, and the data fellows who provide technical support to government agencies. The project manager role was the only role not affected, since they were permanently employed and their scope of work entailed more projects at the ICT Authority besides KODI.

4.1.4 | KODI timeline

Figure 2 provides a detailed outline of KODI. This is based on the information provided in Sections 4.1.1, 4.1.2, and 4.1.3 above. This timeline illustrates the progress made between 2011 and 2016. It includes information about the cumulative number of datasets, proposed or enacted laws, number of government agencies actively publishing their datasets on KODI, and the cumulative number of page views.

Further analysis of these milestones is documented in the following sections.

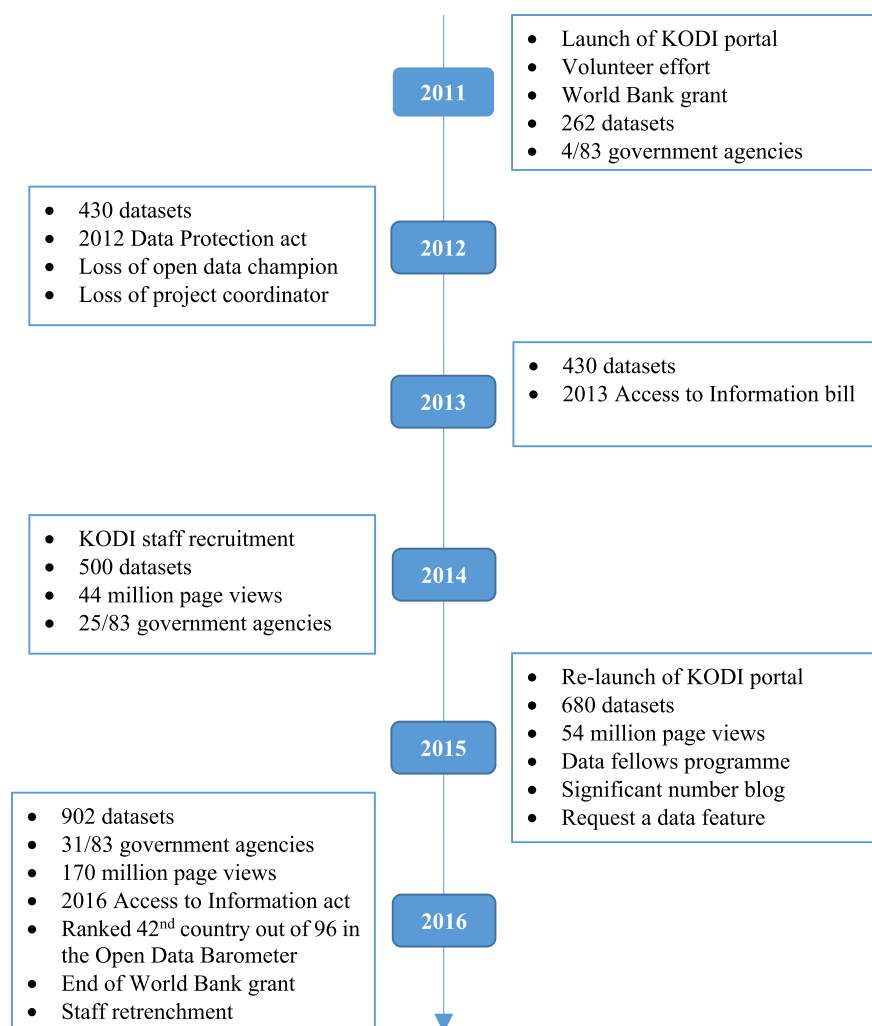


FIGURE 2 KODI timeline

4.2 | Identification of key entities

This step builds on the case description provided above and helps in identifying entities linked to the events. As noted above, entities comprise actors, organisational units, and objects which, as networks, make up structures with causal powers. The core object in this case is the KODI portal, which is dependent on several technical and social structures. The key social structure entities in this initiative go beyond staff at KODI and include some of the stakeholders identified in Table 3 such as partners from civil society, World Bank, NGOs, and information technology start-ups geared towards providing open data services.

Alongside this network of social entities, and linked to it are technical structures, which refers to those processes, features, or tools that enable supply and demand of data. Supply of data is supported by the data release form, data release calendar, and the curation tools. The data release form is an agreement between KODI and a government agency on the release frequency for a given dataset. The data release calendar provides information on the particular dates when datasets listed on the open data portal get released. This helps users to determine when to expect new releases.

Kenya Open Data Initiative provides technical assistance to government agencies on how to use the various curation tools including Tabula, Pentaho, and KODI portal backend, where they upload the datasets for publication. Demand for open data is supported by developer application programming interfaces (APIs), the search feature on the portal, and open data visualisation tools. Data-driven publications such as newspaper articles, blogs and audio/video documentaries from mainstream media, citizen-generated blogs, and infographics also help in increasing demand for data. In addition, the search feature helps ordinary citizens and organisations to search for open data, with the aim of generating new information, or verifying the correctness of certain stories or publications. APIs are also a must have, especially due to the voluminous nature of open data, as they allow developers to write scripts to automate the process of pulling and processing data from the KODI portal. Figure 3 provides a visual illustration of these entities. The network of technical features is not static with, as noted above, a request-a-dataset feature being added following user requests.

Alongside the connected networks of social actors and technical entities, the functionality of the latter is governed by several less-concrete social structures including political goodwill, commitment by government to allocate adequate and sustainable budget and resources, commitment by government agencies to supply data guided by existing laws and policies, and citizen demand and engagement supported by feedback mechanisms on the portal, and by awareness and equality of access. We can therefore summarise KODI key entities as consisting of an overall socio-technical network made up of a combination of formal-social, abstract-social, and formal-technical elements.

4.3 | Theoretical re-description (abduction)

Theoretical re-description helped in identifying several underlying mechanisms that are necessary in institutionalising KODI; that is, in establishing supportive structures that guarantee sustainability. Although guided by the literature on institutions and institutionalisation, we did not attempt here to undertake a deep theorisation; nor did we cycle through different theoretical possibilities in any detail. Instead, we were guided by two key issues that emerge from both conceptual and practical discussions of open data: supply and demand.

Underlying socio-political mechanisms can be understood to affect the use of open data technologies from a demand and a supply perspective. Drawing from both conceptual and practical evidence, we can see that supply requires ownership from government, which is characterised by provision of resources and prioritisation of open data deliverables. There is also a need for budget allocation to ensure sustainability. Resources are not only technology but also human capacity and expertise, which is achieved through training and allocation of dedicated staff for time and skill intensive tasks. These aspects are socio-political because they are mainly dependent on goodwill from government, which also affects law and policy formulation and enforcement.

Demand for open data is also dependent on similar socio-political mechanisms because of the feedback from supply to demand. Even though engagement channels were later introduced, the willingness of government to release open data remains a challenge: if that supply is not forthcoming then demand from citizens and others will dwindle. The KODI support team was necessary in following up with government agencies to

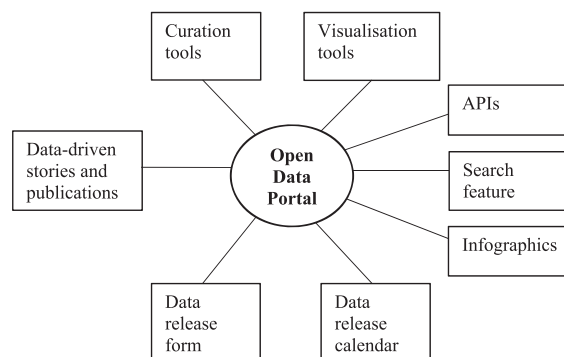


FIGURE 3 Technical entities supporting KODI

release their datasets. Otherwise, citizens would have had to individually follow up with the various agencies, and lack of proper networks and understanding of government operations would have made that unlikely to succeed. Allocation of budget to support the KODI team needs socio-political interventions, driven by an open data champion with the right socio-political connections. And wider issues impacting OGD demand—interest of citizens and other data users in government, belief that data-based action can have an impact, etc. —will also be shaped by the socio-politics of Kenya.

4.4 | Retroduction: identification of candidate affordances

This is a continuation of the theoretical re-description phase and involves further postulation of mechanisms that are capable of producing the observed KODI events and non-events. This is of course a crucial step given the guiding research question: what are the causal mechanisms that enable institutionalisation of the KODI?

Reflecting on the findings of step three above, we observe that the socio-political mechanisms are the main mechanisms affecting institutionalisation of KODI. In order to be more specific in the process of postulating the enabling institutionalisation mechanisms, the research question was therefore revised as follows: what are the socio-political mechanisms that enable institutionalisation of KODI? This question will act as a lens in understanding the underlying affordances, including their associated mechanisms and what stimulates them to be actualised.

4.4.1 | Identification of immediate concrete outcomes

This step helped in identifying immediate concrete technology-enabled outcomes for KODI. The main observation was that although technology could be viewed as a key enabler, it was not the key determinant in the institutionalisation of KODI. Socio-political structures and mechanisms had a larger role to play. This implies that political actors have a key role in determining the success of this initiative. The technical actors may be willing to provide the needed support, but their efforts will only have impact if political leaders and government management provide sustainable structures, including enforcement of enabling laws and policies, which include provision of required financial and technical structures. For instance, the main team at KODI ceased work in December 2016 after the project funds were depleted. This was not an issue of non-performance on their part but an issue of planning, which requires political goodwill. Lack of prioritisation of this initiative could imply that government perceived this as a threat or does not fully understand its value and potential. Hence, the overall relatively limited concrete outcomes of KODI in terms of sustainable release of OGD.

4.4.2 | Analysis of the interplay of human and technical entities

Technical entities comprise deep structures including business rules and interfaces, which guide how human entities interact with technology and each other within a network. These structures determine the kinds of innovations and events that emerge from KODI. Structures are therefore socio-technical and socio-political, which demonstrates the need for human entities, who collectively form the political economy. Political economy is the nexus between government, business, and politics. Analysis of the current complexities, such as lack of budget for KODI from government, suggests that these challenges will remain due to lack of political goodwill. Therefore, introduction of supportive laws and policies like the 2016 Access to Information Act or acquisition of the right technologies such as Socrata does not lead to automatic institutionalisation. The nexus between government, business, and politics in Kenya is thick and hard to penetrate. For example, separation between bureaucracy, politics, and the private sector may be more apparent than real, with the same people involved in all and blocking any initiatives that may threaten the stability of this ecosystem. In most cases, government officials appear to be acting within the law, but they may be actively involved in the private sector and political arena through proxies. As a result, the bureaucratic challenges being experienced by KODI while trying to get other agencies on board, and enforcement of open data laws and policies can only be resolved by securing political goodwill within the wider political economy.

4.4.3 | Identification of candidate affordances

In the case of KODI, two affordances were identified. These were derived from the description of events, the prevailing structures, immediate concrete outcomes, and the interplay between human and technical entities. They are a necessary input in the process of determining the underlying causal mechanisms.

The data release affordance is dependent on three technical entities, namely the data release form, data release calendar, and curation tools. This requires agreement between the KODI team and the government agencies regarding the release of data, and technical support in the process of curating data for publication. This affordance helps in achieving one of the outcomes, which is publication of open data.

The data access affordance is dependent on four technical aspects including API provisioning, interactive search feature on the portal, data visualisation tools, and user training and support. This affordance helps in achieving two more outcomes that are citizen oriented: provision of access to open data, and increased awareness on the existence and benefits of open data.

4.4.4 | Identification of stimulating and releasing conditions

The affordances identified in the previous stage were further analysed with the aim of understanding the contextual conditions necessary for them to be actualised.

The data release affordance is stimulated by formal agreements and goodwill from the management of a government agency in participating in the open data exercise. This also required identification of the right stakeholders within the agency. In the case of KODI, agency executives assumed that communication from the Ministry of ICT was directed to the agency ICT manager. However, this was not necessarily the case since data rests in the departments executing the core mandate of the agency. The releasing condition was identified to be the allocation of dedicated human and technical resources by the agency. Human resources required training and commitment to release data at agreed intervals. This commitment requires management support to allow them to dedicate time to work on open data deliverables.

The data access affordance is stimulated by availability of up-to-date open datasets on the portal, availability of interactive-search features, and API provisioning to allow developers to connect their applications with the open data portal. Ultimately, the releasing condition is dependent on how well KODI reacts to change requests or request for new features. There is also need for continuous awareness campaigns to increase use.

4.5 | Analysis of the set of affordances and associated mechanisms

This step helped in describing the interaction and dependencies between the two affordances that were identified in the previous stage. Higher-level mechanisms were generated from these findings. These mechanisms are treated as a candidate mechanism to allow for further analysis in the next stage.

The data release affordance focused on supply of open data, while the data access affordance focused on the connection between supply and demand for open data. As noted above, demand and supply are interdependent, since demand is what helps in justifying the necessary investment for open data supply. Demand can also only increase when user requests are addressed to their satisfaction, ie, through supply. Conversely, increased supply would potentially attract greater demand for use of increasingly available open data. These affordances were abstracted into a generalised data demand and supply mechanism. This abstraction needs to be conjoined because of the interdependency between demand and supply just noted.

Supply and demand for open data is also dependent on the political economy. The institutionalisation of open data laws and policies, which would facilitate supply and demand of data, requires political goodwill. This includes governance issues such as allocation of required resources and budget, and also enforcement of laws that require agencies and other stakeholders like the private sector to release data that is viewed to be of national interest. Release of such data would push demand, as the country leans more towards facts and less towards hearsay.

We can therefore summarise the overall mechanisms as a set of ever-deeper retroductions; beginning first with the data release and access affordances noted in Section 4.4.3; then moving back into the conditions for activation of those affordances noted in Section 4.4.4; but with the ultimate underpinning relating to the political economy of open data in Kenya as described in Section 4.4.2. The overall is summarised in Figure 4.

4.6 | Assessment of explanatory power

This stage helped in assessing whether the mechanisms identified in the previous stage have solid explanatory power supported by empirical evidence. The two affordances were initially viewed as candidate mechanisms. However, further analysis revealed that one dealt with supply and the other with demand for data. This resulted in the demand mechanism and the supply mechanism. Further analysis revealed that the impact of these mechanisms was inter-linked and interdependent, resulting in a merger of the two mechanisms, leading to the data demand-and-supply

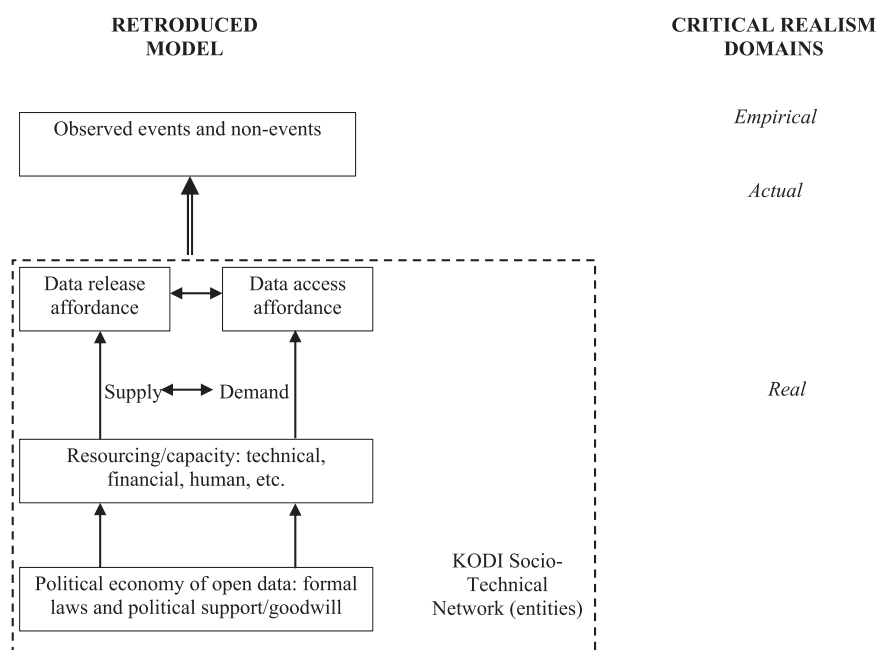


FIGURE 4 Summary of retroduction process

mechanism. This mechanism is responsible for the institutionalisation of KODI. Its implementation requires consideration of the underlying structures and mechanisms that were discussed earlier.

Full assessment of the explanatory power of the proposed mechanism set would require further and deeper investigation of the case, and we acknowledge this. However, given support for the mechanism set from both empirical and conceptual evidence, we can propose that it has at least some level of explanatory power.

For example, we can identify support from other work for the idea of these mechanisms and the layers of intermittent supply of data underpinned by resourcing/capacity challenges underpinned by the political economy of both formal (eg, laws and policies) and informal (eg, political support) institutional factors. Jetzek et al. (2014), for instance, make similar observations, and identify these same challenges as barriers to value generation from open data more generally.

These authors also identify other factors enabling demand for data which may also apply in the Kenyan context. These include user incentives, capabilities, and technical connectivity. We acknowledge that incorporation of stronger demand-related factors such as these and related mechanisms would be of value: our KODI field data and hence analysis were relatively limited in consideration of demand-related mechanisms. Reinforcing the value of this extension, we can see that at the point of KODI launch there were strong incentives for developers and start-up organisations to get involved—a kind of 2011 “data rush.” But these actors did not sustain their involvement nor, hence, their contribution to stimulation of demand once those initial incentives dried up. Nor did the incentives focus on the core of demand: citizen-users and the institutions of civil society.

From this, we therefore note the potential for extensions and refinements to the proposed model, in order to increase its explanatory power.

5 | CONCLUSION

At the heart of critical realism is the search for underlying mechanisms: elements of social structure within the real domain which can help to explain the observations made in the empirical domain. Such a search is of value across a range of research questions, including the question posed here relating to factors enabling and constraining the institutionalisation of OGD in Kenya.

One challenge for those seeking to use critical realism in ICT4D (or other) research is how to operationalise critical realism, particularly the search for mechanisms. In this paper, we propose the value of following the step-by-step approach proposed by Bygstad et al. (2016). We then demonstrated its application through analysis of KODI and uncovered a layered model of mechanism components which had explanatory power. It showed the elements required for institutionalisation of OGD in Kenya, the absence of which has so far prevented institutionalisation. Notwithstanding the value of the model developed, we recognised the way in which further revision to the model could increase its explanatory power.

The particular feature of Bygstad et al.'s (2016) approach is the incorporation of affordances as means to conceptualise the proposed mechanism. We have not delved deep into the notion of affordances in this paper. However, we have shown how it can be incorporated: in this case forming part of the mechanism that was then underpinned by deeper features around resourcing and formal/informal political institutions.

We hope this “proof-of-concept” to operationalising critical realism will encourage others to make use of this particular approach and broaden the application of structured approaches to critical realism within the ICT4D research field.

REFERENCES

- Access to Information Act (2016). *Access to Information Act No. 31 of 2016*. Nairobi: National Council of Law Reporting. <http://www.kenyalaw.org>
- Archer, M. S. (2007). The trajectory of the morphogenetic approach: An account in the first-person. *Sociologia*, 54, 35–47.
- Bhaskar, R. (1998). The critique of the positivist tradition: Explanation, prediction and confirmation. In *The possibility of naturalism: A philosophical critique of the contemporary human sciences* (3rd ed.) (pp. 136–146). Abingdon, UK: Routledge.
- Boyatzis, R. E. (1998). *Thematic analysis and code development: Transforming qualitative information*. London: Sage Publications.
- 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.
- Carlson, S. A. (2009). Critical realism. In Y. K. Dwivedi, B. Lal, M. D. Williams, S. L. Schneberger, & M. Wade (Eds.), *Handbook of research on contemporary theoretical models in information systems* (pp. 57–76). Hershey, PA: IGI Global.
- Crabtree, B. F., & Miller, W. L. (1999). A template approach to text analysis: Developing and using codebooks. In B. F. Crabtree, & W. L. Miller (Eds.), *Doing qualitative research* (3rd ed.) (pp. 93–106). London: Sage Publications.
- Danermark, B., Ekström, M., Jakobsen, L., & Karlsson, J. C. (2002). Generalization, scientific inference and models for an explanatory social science. In *Explaining society: Critical realism in the social sciences* (2nd ed.) (pp. 73–114). Abingdon, UK: Routledge.
- Davies, A., & Lithwick, D. (2010). *Government 2.0 and access to information*. Ottawa, Canada: Library of Parliament, Parliamentary Information and Research Service.
- De Landa, M. (2006). *A new philosophy of society: Assemblage theory and social complexity* (1st ed.). London: Continuum.
- Easton, G. (2010). Critical realism in case study research. *Industrial Marketing Management*, 39(1), 118–128.
- 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.
- Fox, S. (2009). Applying critical realism to information and communication technologies: A case study. *Construction Management and Economics*, 27(5), 465–472.

- Heeks, R. (2002). eGovernment in Africa: Promise and practice (iGovernment Working Paper Series No. 13). Manchester, UK: Centre for Development Informatics, University of Manchester. <https://www.gdi.manchester.ac.uk/research/publications/working-paper-archive/igov/>
- Henfridsson, O., & Bygstad, B. (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.
- Hoxha, J., & Brahaj, A. (2011). Open government data on the web: A semantic approach. 2011 International Conference on Emerging Intelligent Data and Web Technologies, Tirana, Albania, 7–9 Sept, 107–113.
- Hume, D. (1966). *Enquiries concerning the human understanding and concerning the principles of morals* (2nd ed.). Oxford: The Clarendon Press.
- 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.
- Jetzek, T., Avital, M., & Bjørn-Andersen, N. (2014). Generating sustainable value from open data in a sharing society. In *IFIP advances in information and communication technology* (ed., Vol. 429) (pp. 62–82). Berlin, Heidelberg: Springer.
- 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.
- Mingers, J. (2002). Real-izing information systems: Critical realism as an underpinning philosophy for information systems. In *International Conference on Information Systems: ICIS 2002*, Barcelona, Spain, 15–18 Dec, 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.
- Open Government Partnership (2011). *Kenya commitments*. Washington, DC: Open Government Partnership. <https://www.opengovpartnership.org/countries/kenya>
- OpenDataMonitor (2015). *Open data software and APIs*. Brussels: OpenDataMonitor. <http://knowhow.opendatamonitor.eu/odresearch/open-data-software-and-apis/>
- Pawson, R., & Tilley, N. (2004). *Realistic evaluation*. London: Sage Publications.
- Pentaho (2017). *Any analytics, any data, simplified*. Santa Clara, CA: Pentaho. <http://www.pentaho.com/product/product-overview>
- Sayer, A. (1992). *Method in social science: A realist approach* (2nd ed.). Abingdon, UK: Routledge.
- 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.
- Smith, M. L. (2006). Overcoming theory-practice inconsistencies: Critical realism and information systems research. *Information and Organization*, 16(3), 191–211.
- Socrata (2011). *Kenya launches Socrata-powered open data site*. Seattle, WA: Socrata. <http://www.socrata.com/newsroom-article/kenya-launches-socrata-powered-open-data-site/>
- Steinmetz, G. (1998). Critical realism and historical sociology. a review article. *Comparative Studies in Society and History*, 40(1), 170–186.
- Tabula (2017). Tabula. <http://tabula.technology/>
- Ubaldi, B. (2013). Open government data: Towards empirical analysis of open government data initiatives (OECD Working Papers on Public Governance No. 22). Paris: OECD. https://www.oecd-ilibrary.org/governance/open-government-data_5k46bj4f03s7-en
- Volkoff, O., & Strong, D. (2013). Critical realism and affordances: Theorizing IT-associated organizational change processes. *MIS Quarterly*, 37(3), 819–834.
- Wynn, D., & Williams, C. K. (2012). Principles for conducting critical realist case study research in information systems. *MIS Quarterly*, 36(3), 787–810.
- Yu, H., & Robinson, D. G. (2012). The new ambiguity of “open government”. *UCLA Law Review*, 59(178), 178–208.
- Zeuner, L. (1999). Margaret Archer on structural and cultural morphogenesis. *Acta Sociologica*, 42(1), 79–86.

Paul Mungai holds a PhD in information systems from the University of Cape Town. He currently serves as a postgraduate fellow at the University of Cape Town under the supervision of Prof Jean-Paul Van Belle. His research interests include inclusive innovation, open development (stewardship and intermediation models), media for development, and big data for sustainable development (includes open data for transparency and accountability).

How to cite this article: Mungai PW. Causal mechanisms and institutionalisation of open government data in Kenya. *E J Info Sys Dev Countries*. 2018;e12056. <https://doi.org/10.1002/isd2.12056>