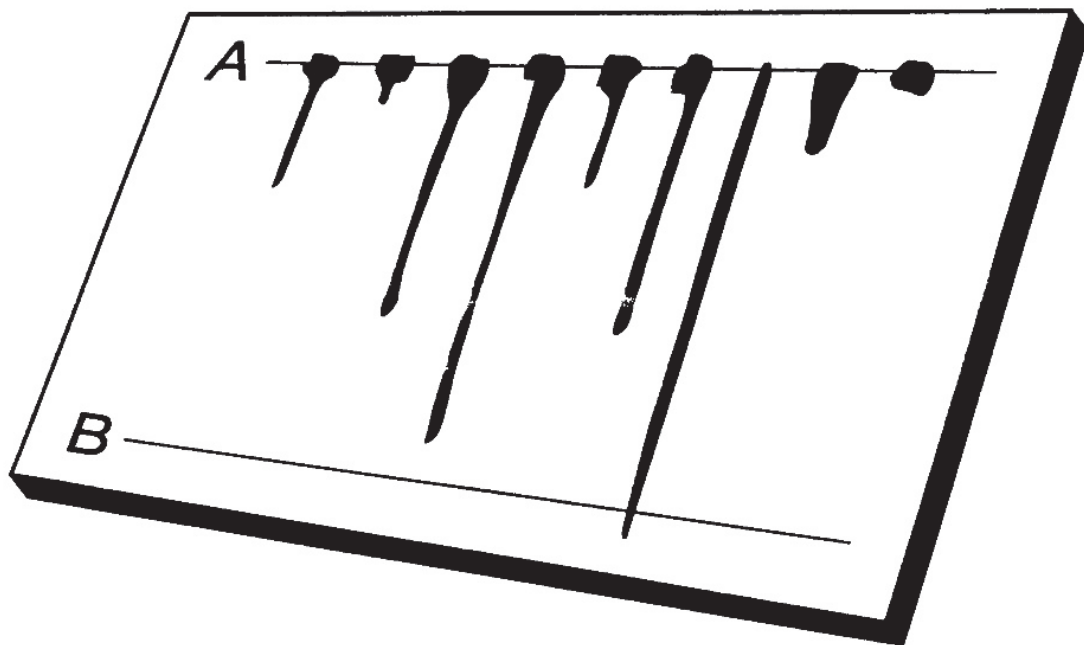


# Viscous open data: The roles of intermediaries in an open data ecosystem

François van Schalkwyk, Michelle Willmers and Maurice McNaughton



*This is an Accepted Manuscript of an article published by Taylor & Francis Group  
in the Journal of Journal of Information Technology for Development*

First author: François van Schalkwyk\*  
Affiliation: University of Cape Town, South Africa  
Postal address: University of Cape Town, Private Bag X3, Rondebosch 7701,  
South Africa  
Telephone number: +27(21)852-7093  
Email address: francois@compressdsl.com

Second author: Michelle Willmers  
Affiliation: University of Cape Town, South Africa  
Postal address: University of Cape Town, Private Bag X3, Rondebosch 7701,  
South Africa  
Telephone number: +27(21)650-5061  
Email address: michelle.willmers@uct.ac.za

Third author: Maurice L. McNaughton  
Affiliation: Mona School of Business and Management,  
University of the West Indies, Jamaica  
Postal address: The University of the West Indies, Mona Campus, Kingston 7,  
Jamaica, W.I.  
Telephone number: +1(876)977-6035  
Email address: maurice.mcnaughton@uwimona.edu.jm

\* Corresponding author

The funding for this work has been provided through the World Wide Web Foundation  
'Exploring the Emerging Impacts of Open Data in Developing Countries' research project,  
supported by grant 107075 from Canada's International Development Research Centre  
(web.idrc.ca). Find out more at [www.opendataresearch.org/emergingimpacts](http://www.opendataresearch.org/emergingimpacts)

## **Abstract**

Open data has the potential to improve the governance of universities as public institutions. In addition, open data is likely to increase the quality, efficacy and efficiency of the research and analysis of higher education systems by providing a shared empirical base for critical interrogation and reinterpretation. Drawing on research conducted by the Emerging Impacts of Open Data in Developing Countries project, and using an ecosystems approach, this research paper considers the supply, demand and use of open data as well as the roles of intermediaries in the governance of South African public higher education. It shows that government's higher education database is a closed and isolated data source in the data ecosystem; and that the open data that is made available by government is inaccessible and rarely used. In contrast, government data made available by data intermediaries in the ecosystem are being used by key stakeholders. Intermediaries are found to play several important roles in the ecosystem: (i) they increase the accessibility and utility of data; (ii) they may assume the role of a 'keystone species' in a data ecosystem; and (iii) they have the potential to democratise the impacts and use of open data. The article concludes that despite poor data provision by government, the public university governance open data ecosystem has evolved because intermediaries in the ecosystem have reduced the viscosity of government data. Further increasing the fluidity of government open data will improve access and ensure the sustainability of open data supply in the ecosystem.

**Keywords:** open data; ecosystem; intermediary; governance; university; higher education; South Africa

## 1. Introduction

Higher education has a critical role to play in development. The creation of new knowledge, innovation, the training of professionals and instilling democratic values are in the hands of the contemporary university (Annan, 2015; Castells, 2001; Cloete et al. 2015a). However, to be effective, efficient and efficacious in the execution of these obligations universities require access to basic but critical data. In particular, *open* data may play a role in improving the governance of institutions, including universities (Cloete et al. 2015b), by increasing the transparency of decision-making as well as the accountability of those tasked with implementing processes that serve the interest of society.<sup>1</sup>

However, governance that is not premised on informed decision-making has the potential to foster weak and fragmented institutions prone to corruption and/or the inappropriate allocation of resources. This potentially destructive combination is among the reasons for 5 of the 23 public universities in South Africa being under administration at the time of writing.<sup>2</sup> University councils need accurate and informative data on the state of their institutions in order to shift the debate from one that is driven by ideology and self-interest to one that is empirically based and in the interest of the performance of the institution.<sup>3</sup>

---

1 In its simplest form, governance is understood to be concerned with the processes of decision-making and implementation (UN ESCAP, n.d.).

2 See *Sunday Independent*, 29 July 2012, 'Poor leadership cripples tertiary institutions'.

3 This need is confirmed by the number of requests received by the Centre for Higher Education Transformation in 2013 from four South African public universities (and three African universities) to present to council (and management) a set of institutional-level indicators. N. Cloete (personal communication, 20 January 2014).

The importance of higher education in development and consequent requirement for well-governed universities, is situated in a context where there has been a phenomenal growth in the supply of open data. Davies et al. (2013) estimate that governments alone have posted in excess of 1 million datasets online.

This paper shares the findings from research that sought to examine the supply, (re)use and possible impact of open data on the governance of South African public universities. Three basic questions were considered: (1) How can the flow of data to and from the government database on higher education be described? (2) Is the provision of data by the government contributing to the evolution of what can be described as the South African public university open data ecosystem? (3) How does the presence of those intermediating between the provision and use of data contribute to the functioning of this ecosystem?

## **2. Conceptual framework**

Yu and Robinson (2012) describe open data as being either “adaptable” or “inert”. Manyika and colleagues, in their recent work on quantifying the economic value of opening up data, alight on the notion of open data being “liquid” (Manyika et al. 2013). That is, open data unlocks value as it flows from governments, between firms, researchers and entrepreneurs, and to citizens, and is adapted in the process. To extend the analogy, the flow of data could result in a virtuous cycle, becoming a stable but dynamic part of an ecosystem. But equally possible, data could, despite being open, become inert and flow too slowly or not at all; it could be too viscous to contribute to the evolution of the ecosystem.

The conceptual framework for this analysis borrows from Helbig et al.’s (2012) “information polity” heuristic. However, we extend the information polity heuristic partly because we believe that the concept of an ecosystem enables the more accurate

reflection of the resources, sources, providers and users in a context broader than when government alone acts as the primary collector and provider of data, and partly because we believe that the concept of the ecosystem will resonate to a greater degree with both practitioners and scholars.

The concept of the ecosystem has already gained a degree of traction in the analysis of how ICTs are driving change, be this in discussions on open government or open data. Harrison et al. (2012), in a review of the ecosystem metaphor in the open government literature, identify several key features of ecosystems. Ecosystems are seen as consisting of mutually interacting organisms; complex in their arrangement; characterized by the interdependency of and between organisms and resources; dynamic rather than static – seeking equilibrium through motion rather than stasis; populated by keystone species that play a critical role in facilitating exchange in the ecosystem thereby ensuring dynamism and constant movement; movement tends to be cyclical and reinforcing making the system resilient (adaptable and restorative); but it is also vulnerable to exogenous forces which may disrupt or destroy the ecosystem.

Martin Fransman, in his book *The New ICT Ecosystem*, draws on the work of evolutionary economist Joseph Schumpeter to describe the components of socio-economic ecosystems and to recast these components in the context of ICT, which he argues constitutes one of many sectorial ecosystems within the larger socio-economic ecosystem. He identifies the dynamically interacting organisms in the ICT ecosystem (firms, non-firms, intermediaries and consumers) bound by exchange as well as by the institutions (the repositories of rules, values and norms) in which they are embedded. Key to his exposition of the ICT ecosystem is that the ICT ecosystem is driven by innovation (i.e. the injection of new knowledge into the ecosystem). Firms compete and co-operate symbiotically, and the interaction between firms and consumers (that is,

between knowledge creators and knowledge consumers) generates new knowledge which leads to innovation in the ecosystem. It is the pursuit of innovation that keeps the ICT ecosystem in motion.

For the purposes of our analysis of a particular data ecosystem: If knowledge creation as a simplified process moves from observation to recording those observations to analysis to testing to validation, and data is the codified retrievable recording of observations in this process of knowledge creation, then it seems reasonable to assume that the open data ecosystem is a key component in the broader ICT ecosystem, particularly if it is premised on innovation as a key driver. What is less clear is whether innovation per se is a driver in an open data ecosystems or, if it is a driver in the open data ecosystem, what conditions need to be in place to ensure the sustainability of such an innovation-driven ecosystem.

An ecosystem consists of at least three contextual conditions under which actors in the ecosystem function and which motivate, direct and/or constrain their actions as data providers, intermediaries or consumers.

The first of these is the regulatory condition – laws, policies, standards and agreements which have a bearing on how the components of the ecosystem are structured and how they interrelate.

The second condition is that of the institutional context in which the actors operate. Each institutional context provides the taken-for-granted values, rules and norms shared by actors who operate within that particular institutional context (Scott 2014). These values, rules and norms inevitably propel and restrain the behaviours of actors in the ecosystem (Janssen et al., 2012).

The third condition is that of current information and communications technologies, that is, the network elements, the network operators and the

communications protocols that connect and interconnect the networked elements, operators and users. Principle among these is the internet as a key enabler that introduces new actors to the ecosystem by connecting them to legacy components.

### 3. Method

In order to answer the research questions posed by this study, the Centre for Higher Education Trust's (CHET) open data platform was used as a case to examine the dynamics of data supply, (re)use and the role of intermediaries in the open data ecosystem.

In 1999, CHET, a non-governmental organization, initiated a project on performance indicators in South African higher education. The project arose from the question: "Is the South African higher education system transforming?" By 2000, the concept of "transformation" had become so ideologized, CHET argued, that the concept no longer had any empirical use. It maintained that a combination of empirical indicators and theoretical reflection was the only way to initiate constructive dialogue between stakeholders on the transformation of the South African higher education system.

In 2009, based on feedback from universities and the refinement of the indicators proposed in 2004 (Bunting & Cloete, 2004), CHET published *Performance Indicators: The South African Higher Education System 2000–2008* (Bunting et al., 2010). For the first time, the data from the publication were also made available on the CHET website.<sup>4</sup> The intention was for the published university performance profiles, in

---

<sup>4</sup> <http://www.chet.org.za/data/sahe-open-data>



conjunction with the online open data, to assist university planners and councils to make assessments that would contribute to evidenced-based management and governance.

#### *Actors in the ecosystem*

University planners were identified as central actors in the (re)use of open data in university governance as they are tasked with providing university executives and councils with accurate data reflecting the state of their institution (Bunting 2004). Seven of the 23 public universities in South Africa were included in this study. The governance structures of South African public universities are fairly homogenous; a selection matrix was therefore devised to ensure a representative sample in terms of size, university type and location.

Acknowledging their role in the research–policy nexus, the research also considered the use of open data by 12 higher education studies researchers. These researchers were identified by analysing journal articles, books and other publications on South African higher education published post-2009 and which made reference to empirical data related to the South African higher education system.

Staff in the employ of the Department of Higher Education and Training (DHET), the sole aggregator and supplier of system-wide, institutional-level data, were also included as key actors in the public university governance ecosystem.

Data on actors in the ecosystem was collected primarily via semi-structured interviews. While time-consuming and therefore limiting in terms of the sample size, semi-structured interviews were preferred as they provided the flexibility to explore unanticipated issues while at the same time providing the structure required to collect comprehensive and comparable data. This flexibility was important given that research was being conducted on *emerging* dynamics in a new and under-researched area.

## *Data*

For the purposes of this study, public university governance data is presumed to include data on students (enrolment data, graduation data and demographic data such as gender and race), staff (number, type, level and qualifications, as well as demographic data), knowledge production (number, type, frequency of publications), curriculum (number and types of courses, and qualifications offered), space (infrastructure, facilities, equipment) and finances. This is the data currently stored in the South African Department of Higher Education and Training's Higher Education Management Information System (HEMIS). These data represent, in our view, some of the key data that universities would use to inform decision-making and implementation at institutional level.

We recognize that defining “open data” represents a significant challenge in an era where internet-enabled information exchange has introduced varying legal and social interpretations of openness, and data manifests in myriad forms. For the purposes of this investigation, the focus is on data contained in databases in both “raw” and “processed” or “shaped” forms. Our definition of open data is that formulated by the Open Knowledge Foundation: “A piece of data or content is open if anyone is free to use, reuse, and redistribute it – subject only, at most, to the requirement to attribute and/or share-alike.”<sup>5</sup>

It is acknowledged that data can be interpreted and classified along a continuum ranging from completely open to completely closed depending on the criteria used to assess openness and, ultimately, on how the authors of the assessment method expect open data to make an impact in a particular context (Van Schalkwyk, 2013). We used

---

5 <http://opendefinition.org/>

the 8 Principles of Open Government as well as the Exploring the Emerging Impacts of Open Data in Developing Countries project's 10-point assessment framework in order to evaluate whether a data source is open or closed (ODDC, 2013). Based on our experience in working with open data in South Africa, we are aware of extremely low levels of interoperability and general confusion in terms of open data licensing. We therefore determined that a 100% compliance rate was unrealistic, and devised a "handicap" system in which a data source had to score at least 80% on both of the open data assessment tools used for it to be considered open.

The identification of data sources was done by means of desk research combined with data collected from the interviews with higher education studies researchers and university planners.

The sample of data sources was further refined by excluding data sources that contained data on the public university sector but which were deemed to be irrelevant in the context of university governance, that is, they provided data that were unlikely to be used in university-level decision-making, planning and implementation.

### *Policies and legislation*

South African higher education is regulated by the provisions of the Higher Education Act of 1997. The Department of Higher Education and Training (DHET) is the government department responsible for the public higher education sector. Universities are largely autonomous and government steers the system by setting goals at the system and institutional levels, and by monitoring the performance of the system and of individual institutions against these goals (Bunting et al., 2010). The primary state steering lever is government funding in the form of annual block grants and earmarked funds for designated projects (Pillay, 2010). In order to monitor performance, universities are required to submit data for inclusion in the Higher Education

Management Information System (HEMIS), which is managed by DHET. Universities are required to capture and submit HEMIS data on students, on staff and on building space data.<sup>6</sup> HEMIS forms the basis for annual state funding allocations as well as system-level policy/steering decisions.

At the level of the university, governance structures are largely homogenous across the system of 23 public universities. The Act outlines the functions of university governance structures and asserts the supremacy of the council as the final authority in university-level governance (Ncayiyana & Hayward, 1999). Council is advised by university executives who head the various organizational units. Typically, it is the function of the institutional planning unit (or its equivalent) to provide strategic support services to the university executive, including the provision of management information that is both relevant and timely for strategic decision-making.

The identification of relevant policies and legislation was done by means of desk research. Rens' (2013) "Opening public data in South Africa: Legal complications" provided a useful starting point in identifying relevant legislation. In our case, relevant policy and legislation was taken to mean those policies and laws which have a direct bearing on the provision and use of data in the governance of South African public universities, with a particular emphasis on any such policies or laws that have a bearing on open data.

#### **4. Findings**

Our modelling of the South Africa public university governance data ecosystem is

---

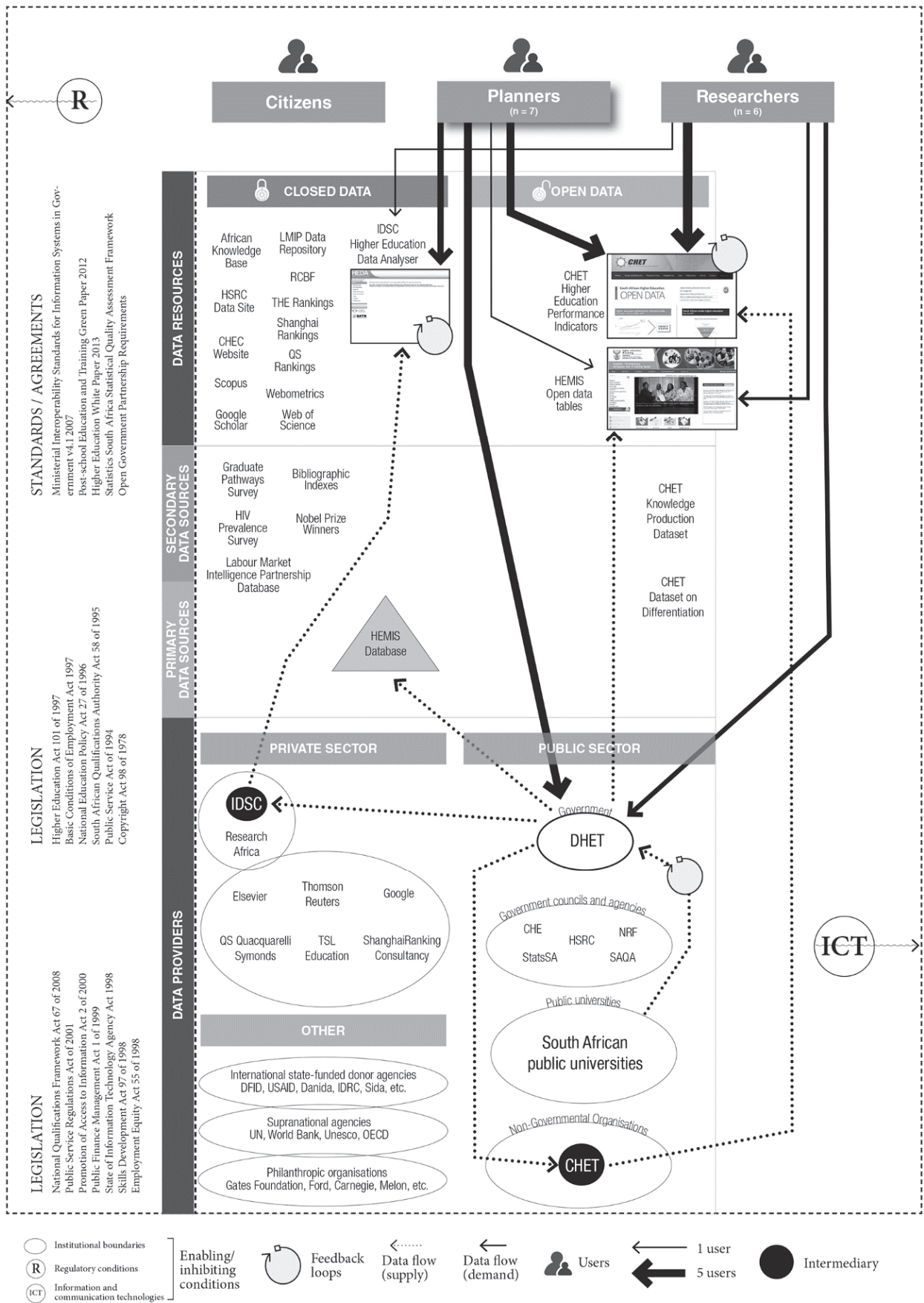
6 Universities are also required to submit finance and research output data but this data is not stored in the HEMIS database. Finance and research data are submitted in the form of annual reports.

presented in Figure 1. It assumes an open system in which new actors operate in a previously closed, government-controlled information system. Traditional boundaries have been displaced and non-public actors have entered into the data system (Janssen et al., 2012). Within the contextual forces, both enabling and restrictive, exerted by ICTs, institutions and the regulative environment, the ecosystem analysis locates the relative positions of the actors in the ecosystem: data providers, sources, resources and users.

Presenting an ecosystem tends towards a simplistic view, the inevitable result of analysis which attempts to create order out of a complex, non-linear set of processes, especially through the lens of a particular conceptual framework. Nevertheless, the representation in Figure 1 is revealing: it shows that while there is a relatively even distribution of data providers in the ecosystem between public and private sectors, the number of data sources is unevenly distributed between those which are open and those which are closed. It is therefore not possible to describe an exclusively open data ecosystem in this case, and, in general, it is therefore perhaps more useful to reflect on the *degree of openness* as an attribute of data ecosystems; or to accept that ecosystems of this type will inevitably consist of data that varies along a continuum of openness.

Because there are many connections and interdependencies between closed and open data sources in the ecosystem one cannot capture the full range of dynamics at play in a data ecosystem that is defined solely by openness. Innovators in the ecosystem may rely on data from both closed and open resources. A by-product of this blended approach on the part of innovators is that it muddies the waters when trying to measure the impact of open data per se in innovation systems, on transparency and, ultimately, on social or economic development.

**Figure 1: A representation of the South African public university governance [open] data ecosystem**



The preponderance of datasets in the ecosystem remains closed, despite the healthy representation of public sector data on the supply side. Datasets indicated as closed are not necessarily inaccessible; they simply do not meet the criteria set out by the Sebastopol Principles and the ODDC. Data that are not machine readable, interoperable, openly licensed, etc. – criteria which many of the datasets in our ecosystem do not meet – are more limited in their potential uptake and reuse. Planners (as the primary users in our ecosystem) draw on both open and closed datasets in the execution of their governance tasks, but their activities in accessing, collating and interpreting the data for the purposes of informed decision-making could be much more efficient and could yield new relational insights if more data was open, linked and licensed.

### ***Findings on the flow of data in the ecosystem***

Interviews with university planners revealed that all universities in the sample extract data from a central management information system, validate the data, and upload the data to HEMIS (via a private IT company appointed on contractual basis by DHET). In all cases, universities have dedicated administrative personnel responsible for collecting, preparing, validating and submitting data to DHET.

HEMIS is a SQL database hosted and managed by DHET and currently maintained by a private IT company, Praxis. Direct public access to the data is restricted as the unit records contain personal data. Public access would be in contravention of privacy rights.

From the HEMIS database, two *open* datasets are published. The first of these is the DHET's own open dataset, which is available on its website as eight *anonymized* data tables (on enrolments, graduates and staff) in Microsoft Excel format.

The second open dataset is supplied by CHET. CHET publishes 18 data tables (each related to a specific performance indicator) in two formats: (i) as downloadable csv files and (ii) through an interface which allows users to generate custom graphs and data tables per indicator with the possibility of comparisons across a maximum of four universities.

A third data provider was identified in the supply of data from HEMIS. The private IT-company IDSC, also extracts data from HEMIS, and makes data available to universities via its own platform, the Higher Education Data Analyzer (HEDA). However, IDSC is not a supplier of *open* data.

No evidence of interoperability was found to exist. In other words, there are no connections between government's HEMIS database and other data sources in the ecosystem.

### ***Findings on the use of open data in the ecosystem***

It was found that CHET's open data is being used by university planners (5 out of the 7 planners interviewed use CHET's open data) and by higher education studies researchers (all 6 researchers use the CHET data, albeit infrequently). It also found that university planners found CHET's performance indicator data useful and some planners expressed interest in additional indicators being made available as open data.

Researchers expressed the need for richer, more granular data. Both planners and researchers expressed the value of the comparative, institution-level open data made available by CHET.

The findings show that DHET's open data is often not the first port of call for data users. In fact, only one university planner and two researchers indicated that they make use of the online DHET HEMIS data tables. One of the researchers indicated that his use of the DHET data tables is a frustrating experience. While users seldom use the



DHET open data tables, the majority planners (5) and the researchers (3) approach DHET by email or telephonically to request HEMIS data.

Users also make use of the HEDA platform; in the case of planners, 4 of them do so, and one researcher was found to use the HEDA platform. This researcher gained access to the HEDA platform by using an institutional user's login details.

## **5. Discussion**

Our discussion and analysis of the findings as presented above and in Figure 1 focus on that part of the ecosystem described by Fransman (2010:24) as the “economic–institutional” component of the ecosystem.

### *Governance domains broader than government*

Helbig et al.'s (2012) information polity heuristic in which the primary source and resource are presumed to reside within government does not hold in our case. The primary data source (where data is collected and processed) is located outside of government (in the universities) while the corresponding primary data resource (HEMIS) is hosted and maintained by government. In other words, when extending the field of analysis from government to the public sector, the location of primary data sources could be in autonomous public bodies (such as universities). It is therefore suggested that analyses of open data ecosystems in relation to governance not be conflated with government – additional governance domains are likely to exist in the broader ecosystem and may have a bearing on how government open data is supplied and (re)used.

### *Determinants of the shape of open data supply*

The fact that two *open* datasets exist from the same source raises the question of why this apparent duplication of open data provision exists in the first place.

We would suggest that the data supplied by government and by the intermediary are different because (i) each provider has different motivations for opening up the HEMIS data, and (ii) there are implicit differences inherent in being a central data source in the ecosystem and in being an intermediary provider.

The fact that DHET provides open HEMIS data could be attributed to a government-wide pledge to open data provision in order to validate its commitment to transparency and accountability.<sup>7</sup> However, based on an interview conducted with a senior official at DHET, the more likely reason for the provision of open data is to redirect those who approach DHET for HEMIS data to the online open data tables. DHET has limited capacity to deal with requests for data – only four DHET staff have access to the full dataset – and sharing the dataset online is therefore an attempt to reduce the burden placed on the Department by external requests for data. The fact that the dataset is difficult to locate on the DHET site seems to support this finding from the interview. If the motivation for opening up the data was transparency or (re)use, one would expect the data to be easier to locate on the DHET website. However, if the motivation is the ability to redirect data queries, then all that is needed is a hyperlink to the data (however obscure that link may be) that can easily be shared electronically.

In the case of CHET, the supply of open data is premised on a clearly identified governance need: “The requirement that each higher education institution must confirm its acceptance of planning targets makes it essential that councils understand (a) what is implied by the targets, and (b) how their institution is performing relative to these targets. CHET’s experience has been that this has not been an easy task for councils.

---

<sup>7</sup> South Africa was a founding member of the Open Government Partnership launched in 2011.

Currently very few institutions produce datasets which would enable council members to engage meaningfully in discussions about the performance of the institution which they are entrusted to govern. CHET decided [...] to produce data profiles which should enable university councils to make assessments of the performance of their institution relative to the targets” (Bunting et al., 2010). CHET’s data is both licensed and easier to locate than the DHET open data, confirming a more user-orientated approach on the part of CHET.

The reason for the dual provision of *open* data on South African public universities appears to be that each supplier is driven by different priorities. In the case of DHET, the supply is driven by internal factors – a lack of capacity – and with no particular reference to users and what their data needs may be. A government-wide commitment to improved system-level governance through transparency and accountability therefore appears to have little bearing on the provision of open data. In the case of CHET, supply is driven by a perceived need for improved governance at institutional level in the light of government-set targets through evidence-based decision-making in relation to such targets.

These differences in priority determine how each organization supplies its open data. CHET provides “shaped” data in the form of indicators because it believes that it will have the greatest chance of effecting change in this format. DHET, on the other hand, effectively “dumps” data on its website with little by way of contextual information to guide the user on how to use or interpret the data. That the data is as complete as possible is important to DHET as it increases the likelihood of the dataset covering the full range of possible types of data requested by those approaching DHET.

Supply of open data is also, however, shaped by different positions in the supply chain. The DHET open dataset complies with the requirements of the data being timely,

complete and primary (albeit that the DHET data is not strictly primary but an anonymized version of the primary data). In the case of the CHET open data, the relevant data is extracted from HEMIS and supplied as indicators. As intermediary, timeliness is a condition over which CHET can never exercise full control unless DHET were to provide real-time access to the HEMIS database. CHET will always be dependent on the release of data by DHET, and will lag behind DHET in the release of open data. Compounding the time lag in the release of data by CHET are the resources (both financial and human) required to convert the raw HEMIS data to performance indicators.

#### *Issues of access to government data in developing countries*

Two intermediaries (CHET and IDSC) play an intermediating role in the flow of data – they are positioned between the government data source (HEMIS) and data users. *How* CHET and IDSC access and provide data from the same data source raises questions to be explored around parallel public and private data flows, the role of social capital and trust in accessing government data and the unequal provision of (re)use rights from a common government data source. Intarakummerd and Chaoroenporn (2013) in their research on the role of intermediaries in innovation in developing countries, highlight the role of intermediaries in compensating for a lack of social capital in innovation systems. They also point to the importance of government initiating and coordinating the activities of both public and private intermediaries.

#### *Keystone species*

“Keystone species” are considered crucial because their presence performs some vital enabling function in the ecosystem (Nardi & O’Day 1999: 53), either as mediators, as actors who bridge institutional boundaries and translate across disciplines, or as creators

of value in ecosystems by creating platforms, services, tools or technologies that offer solutions to other actors in the ecosystem (Iansiti & Levin, 2004, p. 7).

CHET enables new connections and solutions within the ecosystem. For example, while university planners can access the anonymized HEMIS data tables from the DHET website, the CHET open data platform enables planners to compare themselves with other universities across a set of indicators using the tools developed for doing so. CHET enables researchers to access data that would otherwise remain inaccessible and difficult to interpret. CHET is also located outside of the two primary institutions – the state and the university – thus enabling it to play a mediating role. CHET as intermediary therefore plays a vital role in the ecosystem in stimulating the flow of data between government data and users, and could be described as a keystone species within the South African public university open data ecosystem.

Keystone species are enablers, not necessarily drivers in the ecosystem; they can be useful but they are not essential to the sustained functioning of an ecosystem. The public university system is a competitive landscape in which public universities compete for finite resources (such as fee-paying students, government block grants, research project funding, etc.). In this context, new knowledge has value in that it may inform decisions that could give a university a competitive advantage over its rivals. In keeping with Fransman's (2010) concept of ICT ecosystems, innovation can therefore be seen as a key driver in the ecosystem as there is virtuous circle between data production, data supply (open or closed) and consumption.

### *Sustainability*

Whether this virtuous cycle in the case of open data supply and consumption is sustainable, is uncertain. The collection, repackaging and provision of data in a format and context that ensures greater probability of use and impact (Helbig et al., 2012)

requires the investment of resources (Iansiti & Levin, 2004). External funding, predominantly from foreign philanthropies, ensures the ability of the intermediary (CHET) to provide *open* data. The feedback loop is reinforced if there is evidence of use/impact as this increases the likelihood of future external funding but does not guarantee it. The provision and impact of open data by the intermediary in the ecosystem is therefore not inherently sustainable.

The issue of sustainability is relevant given the presence of a second intermediary in the ecosystem. As O’Neil (2013, p. 33) states unequivocally: “Without money, there is no sustainability.” IDSC is a commercial supplier of public university governance data; South African public universities pay annual subscription fees to access IDSC’s data platform. IDSC’s presence in the ecosystem suggests that in the case of public university governance in South Africa, data users (universities) derive value from the data and that they are prepared to enter into an exchange relationship for the provision of this data. Similar business models exist in other countries (e.g. Academic Analytics in the US). For as long as the users perceive value in the provision of data, and are adequately resourced to enter into an exchange relationship,<sup>8</sup> this part of the ecosystem appears to be more stable and sustainable.

An unfortunate irony would be if the less-sustainable actor – CHET – disrupted the closed data system only for the vacuum in the ecosystem to be filled by a more financially sustainable actor – IDSC. Stated differently, a less sustainable virtuous cycle premised on openness could potentially facilitate a more sustainable “vicious” cycle in

---

8 It is worth noting from our findings on the use of higher education data, that higher education researchers do not make use of HEDA. Researchers, some of which are independent or employed by non-governmental organizations, most likely do not have the financial means to subscribe to the HEDA platform.

which public data is not only used for private gain, but in which access to data on higher education performance becomes increasingly restricted. This raises questions about the viability of not-for-profit, civic-minded intermediaries and, ultimately, the sustainability of *open* data ecosystems; particularly those ecosystems in which the state plays a weak supportive and co-ordinating role.

For the time-being both intermediaries co-exist in the ecosystem. However, in a developing country context where markets are both large and untapped, the sustainability of intermediaries is a reason for concern if the value of open data is to be realized. Philanthropic support may be providing the impetus, but capital will have to flow, and governments will have to provide the conditions for both the economic and social benefits of open data to be realized.

#### *Capacity and evolution*

In the case of DHET, the provision of open data is not driven by financial incentives or rewards, nor by civic-mindedness. As indicated earlier, data appears to be driven by a need for efficiency due to limited capacity at departmental level.

While government has opened up the HEMIS data by providing the data tables on its website, the impact of this action in creating new connections between the open data and potential users is minimal. Interviews confirm that university planners, industry, supranational agencies, researchers, the media and other stakeholders still approach DHET directly for data, rather than downloading the data from the DHET website. The provision of open HEMIS data appears to have had little impact in disrupting this behaviour on the part of data users in the ecosystem.

Granting access to HEMIS by third-parties (such as CHET) under controlled conditions to protect personal data could further stimulate the provision of open data and relieve pressure on the capacity-constrained government department. This could

bolster the impact of open data on the governance of South African public universities. Discussions between DHET, CHET and other stakeholders on how to share HEMIS data and how to improve the interpretability of the current DHET open data tables, could stimulate a new phase in the evolution of the higher education governance open data ecosystem.

Using Ding et al.'s (2012) roadmap of linked open data, DHET still has some way to go before proceeding from its current "open stage" to the "link stage". However, before embarking on a strategy to link its data more effectively, a more fundamental step may be needed to increase the use of DHET's open data. Ding et al.'s (2012) open stage stipulates not only that governments place datasets online but that they assist citizens in finding relevant datasets. And Helbig et al. (2012) in their case study research, highlight the importance of the context in which data is shared as a determinant in the uptake of open data by consumers. Context is a determinant factor in avoiding conflicts of meaning, misinterpretation and user frustration. Unless data providers (including government) not only pay attention to but invest resources to create contexts in which open data is easy to interpret and consume, open data initiatives risk reducing their impact on governance as well as their contribution to innovation and socio-economic development.

Returning to DHET's unique position in the ecosystem as the central data source, it is noticeable from the ecosystem analysis that it is isolated from all other data resources in the ecosystem. Data from the HEMIS database shared with secondary data providers such as IDSC and CHET, are supplied by only four senior staff at DHET who have access to HEMIS. No external database or system draws data from HEMIS. This seems surprising given government's political commitment to electronic and open government.



It seems that while there is intent on the part of government, until government open data is supplied in an information context that meets the needs of its citizenry (constituted of a range of user types, both in terms of needs and levels of access) and is made available via platforms that allow for interoperability, the reuse of government with the open data ecosystem will remain limited. As Gurstein (2011) cautions: “Any critical analysis of ‘open data’ use has to include how and under what conditions the data that is being made available is contextualized and given meaning.”<sup>9</sup>

Whether it is providing a richer information context to ensure greater interpretability or creating a data source that is interoperable, intermediaries have a valuable contribution to make in providing capacity and flexibility to resource and institutionally constrained government departments such as DHET.

Open government data initiatives are often linked to the notion of “government as a platform” (O’Reilly, 2010) in which government acts as the primary data provider and innovative actors external to the state (re)use open government data to provide better, more efficient or more customized public services. In the case of this research project, these actors are the two intermediaries in the ecosystem and their presence attests to their greater (innovative) capacity over the public sector. However, to realize and to maximize fully the contribution of these intermediaries in the evolution of the

---

9 Here we are only referring to the context in which the data is provided (a website, an online platform, a dashboard, etc. and the content they contain). Equally important, according to Gurstein (2011) are the variable contexts in which the spectrum of open data consumer finds themselves. A point echoed by the political commentator Steven Friedman when commenting on the implementation failures of the South African government’s Open Government Partnership action plan – “Democratic government is meant to serve the people. This possibility is restricted when government alone decides the forums in which citizens should talk to it” (Friedman, 2013).

ecosystem, government needs to interact with the intermediaries in the ecosystem. In our case, it is clear that there is little interaction between DHET and the two intermediaries, and that this is inhibiting the evolutionary pace of the ecosystem.

### *Intermediaries and information injustice*

In the ecosystem under analysis, funding incentives to universities ensure the provision of data from universities to government. In other words, an incentive is already built into the system to ensure data capturing at institutional level and its supply to a central point in government (i.e. DHET); a condition which is absent in many other African public university governance ecosystems (Cloete et al., 2011), and may be absent in other developing-country contexts. A lack of capacity could be the primary constraint on open data provision in these contexts, and a lack of incentives or rewards is likely to maintain the status quo. However, the introduction of incentives and rewards to stimulate the flow of data may introduce unintentional bias in data collection and interpretation.

DHET indicated that a process was underway to ensure that the Department's data complied with Statistics South Africa's South African Statistical Quality Assessment Framework (SASQAF). Compliance with SASQAF would mean that the DHET's higher education data would be elevated to official, national data status. Such compliance with SASQAF as well as the incentives present in the ecosystem for universities to capture data according to HEMIS specification, highlights the potential danger of reinforcing social injustices predicated on what Johnson (2013, p. 12) refers to as "disciplinary power" in the ecosystem:

The opening of data can function as a tool of disciplinary power. Open data enhances the capacity of disciplinary systems to observe and evaluate institutions' and individuals' conformity to norms that become the core values and assumptions

of the institutional system whether or not they reflect the circumstances of those institutions and individuals. [...] [T]he surveillers and sousveillers evaluate all institutions according to the norm [...] and the institutions internalize the norms and orient their actions to them. With the norms reflecting the power structure of the society in which they developed, they reiterate the injustices that open data set out to ameliorate.

Implicit in understanding the functioning of open data in society is a sensitivity to institutions as sites of shared norms and values where conformity is prized. The open data movement needs to take heed of the institutional context when evaluating the effects and impacts of opening up data. A more nuanced appreciation of institutional contexts will allow the open data movement to predict with greater certainty the possible strategic responses of those institutions (such as government) being pressured to open their datasets.

Johnson (2013) refers to the US Integrated Postsecondary Education Data System (IPEDS) in imposing institutional conformity through disciplinary surveillance; HEMIS has a similar impact on South African public universities. This condition seems to highlight the importance of intermediaries in curtailing the “de-ameliorating” effects of disciplinary surveillance on open data. In other words, intermediaries, as actors who may well operate outside of the boundaries of the state apparatus and of the institution of the university, have the propensity to challenge how data is collected, interpreted and shared. Their role as de-institutionalized actors could go some way in restoring the democratic value of open data. In addition, intermediaries are in a position to add to existing datasets, thereby extending both the corpus of open data on higher education and the possibilities for new interpretations of the data. In our case, there is evidence of this expansion in two forms. First, CHET’s open data platform provides financial data which is neither collected by HEMIS nor shared by DHET. Second, because CHET provides indicator data, it has introduced new forms of analysis (e.g. on success rates or

cost per graduate) which challenge the normative assumptions inherent in the DHET's construction of the HEMIS database.

It goes without saying that CHET, in the process of representing the HEMIS data and by adding its own data, is not immune to embedding its own values into the open data presented. However, as Johnson (2013) suggests, pluralism is one approach to countering information injustices. By promoting multiple, even conflicting, information systems, by including multiple stakeholders in the design of such systems and by broadening the range of data analysers, the undesirable effects of embedded norms and values are more likely to be ameliorated. Intermediaries, it would appear, have an important role to play in this regard.

#### *Limitations of the ecosystems approach*

What Johnson's (2013) analysis reminds us of and what Helbig's (2012) information polity retains, and which we would suggest the ecosystem presented here lacks, are the power dynamics at play between actors. The marketization of higher education and the fears of primary data providers related to how their data will be used was evident in interviews conducted with DHET and with university planners. This points to the relatively powerful position of primary data providers in the ecosystem. The ecosystem does not capture the power relations between primary providers and other actors in the ecosystem, nor does it reflect the extent to which citizens are able to mobilize in order to counter the power of primary providers of data. If injustice is seen as a potential outcome of an imbalance in power, then Johnson (2013) is correct to caution about the possible injustices inherent in open data provision. Johnson (2013), approaching the use of open data from the demand-side, expresses concerns over the correlation between open data and information injustice, a scenario premised on the differential capabilities of users; capabilities which result in an uneven distribution of power between users.

Again, the ecosystem does not capture the power dynamics at play between users – a condition which may well determine the (re)use and, more importantly, the impact of open data.

The application of the ecosystem analysis in this case has been primarily descriptive and has provided a useful metaphor for characterizing and explicating the roles of various actors within the emerging open data ecosystem. The invisibility of the internet as a key facilitator of change in the ecosystem modelling approach is, however, one of the limitations of the ecosystems approach. An alternative approach such as actor network theory contemplates the agency that both human and non-human entities have by virtue of being represented as actors with similar ontological positions in the network. This provides a means of considering the impact of object entities such as the internet and data, and to use more robust analytic techniques such as social network analysis to further explore the roles, relationships and influence of these various actors within the emerging ecosystem. Future research could consider the integration of these established theoretical approaches and analytic techniques with the ecosystem method for examining open data as a relatively new political, social and economic phenomenon.

## **6. Conclusion**

The ecosystems analysis set out to establish whether government-supplied open data is viscous or fluid, whether its data is contributing to the evolution of the ecosystem, and what role intermediaries are playing, if any, in the evolution of the South African public higher education governance open data ecosystem.

From the interviews conducted, the South African government (in the form of DHET) makes open data available because of capacity constraints. Rather than imperatives of transparency or accountability, it is an efficiency imperative which has opened up HEMIS data. But uptake of the HEMIS open data on the DHET website

appears to be minimal; an array of data users still approach DHET directly for data from HEMIS.

In parallel to the DHET's provision of open data, data is provided by two intermediaries in the ecosystem. These intermediaries follow different modes of data provision – one open and one closed. Both have a very clearly defined target audience in the form of university planners and both provide relatively elaborate information contexts with these users in mind. In contrast, it would be difficult to describe the DHET open data information context, as it currently exists on its website, as user-centric. The open data provided by DHET is viscous rather than fluid.

And while it is true that the ecosystem has evolved due to the activities of the intermediaries and in spite of the DHET open data's poor information context, DHET would do better (i) by improving the information context in order to facilitate the uptake and (re)use of its open data; (ii) by making the HEMIS data interoperable, thereby allowing more interest and activity from existing and new intermediaries; and (iii) by engaging with intermediaries in the ecosystem in order to make the most of their innovation, capacity and flexibility.

Such moves will contribute to the further and continued evolution of the public university governance ecosystem by decreasing the viscosity of government-supplied open data and increasing the fluidity of open data between actors in the ecosystem. This will sustain and promote plurality in the supply of open data and increase responsible use by university planners, higher education studies researchers and other stakeholders in their efforts to ensure good governance in South Africa's public universities.

Intermediaries, driven by civic or financial imperatives, and using their social capital to access data directly from government, are nevertheless creating their own platforms to make data (open or closed) pertinent to the governance of South Africa

public universities available to clearly identified user groups within the public university governance ecosystem. It is in this part of the ecosystem that we see expansion. Time will tell if these new platforms are effective, efficacious and sustainable.

## References

- Annan, K (2015) Special Keynote Address at the African Higher Education Summit, 10-12 March, Dakar, Senegal. Retrieved from <http://livestream.com/WorldBankAfrica/HigherEducationSummit/videos/79849399>
- Bauer, F. & Kaltenböck, M. (n.d.). *Linked open data: The essentials*. Vienna: Edition mono/monochrom.
- Bunting, I. (2010). *Performance indicators: The South African higher education system 2000-2008*. Cape Town: CHET.
- Bunting, I. & Cloete, N. (2004). *Developing performance indicators in higher education*. Cape Town: CHET
- Bunting, I., Sheppard, C., Cloete, N. & Belding, L. (2010). *Performance indicators: South African higher education 2000-2008*. Cape Town: CHET.
- Castells, M. (2001). Universities as dynamic systems of contradictory functions. In J. Muller, N. Cloete & S. Bada (Eds), *Challenges of globalization: South African debates with Manuel Castells*. Cape Town: Maskew Miller Longman.
- Clark, B. R. (2008). The advantages of case study narratives in understanding continuity and change in universities. In B. R. Clark, *On higher education: Selected writings, 1956-2006* (pp. 549-554). Baltimore, Johns Hopkins University Press.
- Cloete, N., Bailey, T., Pillay, P., Bunting, I. & Maassen, P. (2011). *Universities and economic development in Africa*. Cape Town: Centre for Higher Education Transformation (CHET).
- Cloete, N., Maassen, P. & Bailey, T. (2015a) *Knowledge production and contradictory functions in African higher education*. Cape Town: African Minds.
- Cloete, N., Maassen, P., Bunting I., Bailey, T., Wangenge-Ouma, G. & Van Schalkwyk, F. (2015b) Managing contradictory functions and related policy issues. In N.

- Cloete, P. Maassen & T Bailey (Eds), *Knowledge production and contradictory functions in African higher education*. Cape Town: African Minds.
- Davies, T. (2013). *Assessing open data supply: Challenging questions*. Retrieved from <http://www.opendataresearch.org/news/2013/assessing-open-data-supply-challenging-questions>
- Davies, T., Perini, F. & Alonso, J. M. (2013). *Researching the emerging impacts of open data: ODDC conceptual framework*. July 2013 – ODDC Working Papers #1.
- Department of Higher Education and Training (Republic of South Africa). (2013). *White paper for post-school education and training: Building an expanded, effective and integrated post-school system*. Pretoria: Department of Higher Education and Training.
- Ding, L., Peristeras, V. & Hausenblas, M. (2012). Linked open government. *IEEE Intelligent Systems*, May/June 2012, 11-15.
- Exploring the Emerging Impacts of Open Data in Developing Countries (ODDC). (2013). Draft framework: Assessing country level open data supply. Framework and research strategies. Retrieved from [https://docs.google.com/document/d/1z-T3QmmZTmWkFrKySi-x\\_EBDVqIt6Orr6z-C58gPGeg/edit](https://docs.google.com/document/d/1z-T3QmmZTmWkFrKySi-x_EBDVqIt6Orr6z-C58gPGeg/edit)
- Fransman, M. (2010). *The new ICT ecosystem: Implications for policy and regulation*. Cambridge: Cambridge University Press.
- Gurstein, M. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday* 16(2).
- Harrison, C., Pardo, T.A. & Cook, M. (2012). Creating open government ecosystems: A research and development agenda. *Future Internet*, 4, 900-928.  
doi:10.3390/fi4040900
- Helbig, N., Cresswell, A. M., Burke, G. B. & Luna-reyes, L. (2012). *The dynamics of opening government data*. New York: Center for Technology in Government.
- Iansiti, M. & Levin, R. (2004). Strategy as ecology. *Harvard Business Review*, 1 March 2004, 68–78.
- Intarakumnerd, P. & Chaoroenporn, P. (2013). The roles of intermediaries in sectoral innovation system in developing countries: public organizations versus private organizations. *Asian Journal of Technology Innovation*, 21(1), 108-119.



- Janssen, M., Charalabidis, Y. & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, 29, 258–268.
- Johnson, J. A. (2013, April). *From open data to information justice*. Paper presented at the Annual Conference of the Midwest Political Science Association, 13 April 2013, Chicago.
- Magalhaes, G., Roseira, C. & Strover, S. (2013). Open government data intermediaries: A terminology framework. In T. Janowski, J. Holm & E. Estevez (Eds), *Proceedings of the 7th international conference on theory and practice of electronic governance*, Seoul, Republic of Korea, 22-25 October 2013. (ICEGOV2013). ACM Press, International Conference Proceedings Series.
- Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S. & Doshi, E. A. (2013). Open data: Unlocking innovation and performance with liquid information. McKinsey Global Institute. Retrieved from [http://www.mckinsey.com/~media/McKinsey/dotcom/Insights/Business%20Technology/Open%20data%20Unlocking%20innovation%20and%20performance%20with%20liquid%20information/MGI\\_OpenData\\_Full\\_report\\_Oct2013.ashx](http://www.mckinsey.com/~media/McKinsey/dotcom/Insights/Business%20Technology/Open%20data%20Unlocking%20innovation%20and%20performance%20with%20liquid%20information/MGI_OpenData_Full_report_Oct2013.ashx)
- Mathekga, R. (2013). *Independent reporting mechanism: South Africa progress report 2011–2013*. Open Government Partnership. Retrieved from <http://www.opengovpartnership.org/country/south-africa>
- Nardi, B. & O'Day, V.L. (1999). *Information ecologies: Using technology with heart*. Cambridge, MA: MIT Press.
- Ncayiyana, D. J. & Hayward, F. M. (1999). *Effective governance: A guide for council members of universities and technikons*. Cape Town, CHET.
- O'Neil, D. (2013). Building a smarter Chicago. In B. Goldstein & L. Dyson, *Beyond transparency: Open data and the future of civic innovation*. San Francisco: Code for America Press.
- O'Reilly, T. (2010). Government as a platform. In D. Lathrop & L. Ruma (Eds), *Open government: Collaboration, transparency, and participation in practice*. Sebastopol, CA: O'Reilly.
- Open Definition. Retrieved from <http://opendefinition.org/>
- Open Government Data. Retrieved from <http://www.opengovdata.org/home/8principles>
- Pillay, P. (2010). South Africa. In P. Pundy (Ed.), *Higher education financing in East and Southern Africa* (pp. 153-172). Cape Town: African Minds.

- Rens, A. (2013, April 15) Opening public data in South Africa: Legal complications. Retrieved from <http://aliquidnovi.org/opening-public-data-in-south-africa-legal-complications/>
- Republic of South Africa. (2011). Open Government Partnership Action Plan: Republic of South Africa. Retrieved from <http://www.opengovpartnership.org/country/south-africa/action-plan>
- Scott, W. R. (2014). *Institutions and organizations* (4<sup>th</sup> edn). Thousand Oaks, CA: Sage.
- Statistics South Africa. (2010). South African Statistical Quality Assessment Framework (SASQAF). 2<sup>nd</sup> Edition. Retrieved from [http://beta2.statssa.gov.za/standardisation/SASQAF\\_Edition\\_2.pdf](http://beta2.statssa.gov.za/standardisation/SASQAF_Edition_2.pdf)
- Sharif, R. M. (2013). *Utilization and value of public sector information for knowledge development: The case of South Africa*. Unpublished doctoral thesis. Syracuse University, Syracuse, New York.
- UN ESCAP. (n.d.). What is Good Governance? Retrieved from <http://www.unescap.org/pdd/prs/ProjectActivities/Ongoing/gg/governance.asp>
- Van Schalkwyk, F. (2013). Supply-side variants in the supply of open data in South African public university governance. In T. Janowski, J. Holm & E. Estevez (Eds), *Proceedings of the 7th international conference on theory and practice of electronic governance*, Seoul, Republic of Korea, 22-25 October 2013 (ICEGOV2013). ACM Press, International Conference Proceedings Series.
- Yu, H. & Robinson, D. G. (2012). The new ambiguity of “open government”, *UCLA Law Review Discourse*, 178, 178-208.