Open Data Intermediaries in Developing Countries

François van Schalkwyk, Michael Caňares, Sumandro Chattapadhyay & Alexander Andrason

ABSTRACT

This synthesis research paper offers a more socially nuanced approach to open data intermediaries using the theoretical framework of Bourdieu's social model. The study is based on the analysis of a sample of cases from the Emerging Impacts of Open Data in Developing Countries Project (ODDC) project. Data on intermediaries were extracted from the ODDC reports according to a working definition of an open data intermediary presented in this paper, and with a focus on the value that intermediaries provide in linking agents in an open data supply chain. The study found that open data supply chains may comprise multiple intermediaries and that multiple forms of capital may be required to connect the supply and use of open data. The effectiveness of intermediaries can be attributed to their proximity to data suppliers or users, and proximity can be expressed as a function of the type of capital that an intermediary possesses. However, because no single intermediary necessarily has all the capital available to link effectively to all sources of power in a field, multiple intermediaries with complementary configurations of capital are more likely to connect between power nexuses. This study concludes that consideration needs to be given to the presence of multiple intermediaries in an open data ecosystem, each of whom may possess different forms of capital to enable the use of open data.

Introduction

As public institutions open up vast and complex datasets, the expectation is that our lives as citizens will improve as a consequence of the data being made publicly available. However, there are several stumbling blocks in the path of extracting the benefits from open data. On the side of the provider these barriers may include the effort and cost required to convert closed to open data; the cost of providing a user-focused context to ensure the uptake of complex datasets; poor data quality; absence of legal and policy frameworks; a lack of capacity to implement and sustain open data practices; and resistance by data custodians to opening data (Janssen 2011; Magalhaes et al. 2013). On the side of the data user, barriers include lack of access, low levels of data literacy, lack of human, social and financial capital to effectively use open data, and also to open up and combine several datasets that together can create value for citizens (Gurstein 2011; Magalhaes et al. 2013; Canares, 2014).

To remove some of these barriers and, in so doing, to unlock the potential of open data, open data intermediaries are seen as playing a crucial role in linking complex open datasets with user needs. According to Van Schalkwyk et al. (2014) their presence in the open data ecosystem stimulates the flow of open data between actors in the ecosystem. Roberts (2014) points out that:

> [C]itizens will be more dependent on third parties — groups that I will call trusted intermediaries — to assure that transparency policies are maintained, and help make sense of information that is accessible through transparency policies. […] [O]ur dependence on intermediaries will increase, and this will raise the difficult question of whether such groups can acquire the resources needed to do the job of intermediation properly.

The different roles that intermediaries may assume in an open data ecosystem and how they are able to intermediate, has received limited attention in research. This synthesis research attempts to develop a more nuanced understanding of open data intermediaries at work in developing countries with a particular focus on how intermediaries connect actors and facilitate the flow of data.

The concept of intermediaries in the context of ICT research was first mentioned in the 1980s, but more as a process of intermediating, than as a collective description of persons and organizations performing the role of intermediation (Tuula 2008). The function of intermediaries was at the time considered critical to produce, launch, scale and popularise innovations by transmitting information from one supplier to another. Thus, intermediaries are viewed as bridging organizations (Sapsed et al. 2007), brokers (Hargadon & Sutton 1997), agents (Howells 2006) or support organizations (Brown & Kalegaonkar 2002). Among these definitions and descriptions, there is one binding narrative – that an intermediary is located between two or more parties.

In terms of providing explanations for the existence of intermediaries and the functions they perform, Sein and Furoholt (2012) capture a widely held view on intermediaries: Intermediaries "help users access information that is publicly available by locating these resources", "integrating various sources on a specific topic, structuring these findings into a form understandable by interested users and disseminating it to them". Janssen and Zuiderwijk (2014) in their study on what they describe as "infomediary business models" also regard intermediaries as creators of value positioned between data providers and data users.

They also point to the fact that intermediaries are vital in systems that become ever more complex resulting in greater levels interdependency between multiple agents as specialization intensifies. Sein and Furoholt argue that, in the case of e-government and governance, intermediaries are critical in the "diffusion of services (Al-Sobhi et al. 2010), reducing corruption (Bhatnagar 2003), moderating discussion on democracy (Edwards, 2002) and providing egovernment, services of various types (Bailey 2009; Gorla 2009)." It is also in these case studies that a new intermediary function was mentioned – what is referred to as "offline intermediaries".

There is a tradition of research and advocacy organisations working with government data that pre-dates the open data movement. These organisations have traditionally facilitated access, use and communication of insights from government data among various non-governmental agencies. These prototype intermediaries include the media, civil society organisations and researchers. A study of the data practices of research and advocacy organisations working with government data has revealed the crucial potential of such organisations to enrich the supply of *open* data in the data ecosystem (Chattapadhyay 2014).

Focusing on what drives such ecosystems, Fransman (2010) draws on the work of evolutionary economist Joseph Schumpeter to describe ICT as a sectorial ecosystem within the larger socio-economic ecosystem. He identifies the dynamically interacting organisms in the ICT ecosystem including firms, non-firms, consumers and intermediaries 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. Of relevance here is research by Intarakummerd and Chaoroenporn (2013) on intermediaries and their role in innovation in a developing-country context. Their findings highlight the role of intermediaries in compensating for a lack of social capital in innovation systems.

Van Schalkwyk et al. (2014) in a study on the use of open data in the governance of South African public universities hint at intermediaries in this data ecosystem relying on personal connections to enable the flow of data to potential data users from a closed government data source. Open data 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.

With regard to the roles of intermediaries in countering injustice, Johnson (2013, p. 12) refers to as "disciplinary power" and the potential of power to entrench existing injustices in an 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.

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.

Some studies have set out to identify the types of *open* data intermediaries that may be active in a particular data ecosystem. Magalhaes et al. (2013) provide a typology of open data intermediaries consisting of three basic types: civic start-ups, open data services and infomediaries. However, their open government intermediaries' framework does little to identify the unique characteristics of each of the intermediary types; nor does their framework provide insight as to the incentives or motivations for intermediaries entering into the ecosystem. Deloitte Analytics (2012) in their open data 'marketplace' identify aggregators, developers, enrichers and enablers as playing intermediating roles. Only two organisational types populate their intermediary category: application developers and businesses.

While the Deloitte report acknowledges the complexity (and the opportunities inherent in this complexity) of open data supply and use, neither their typology nor that of Magalhaes et al. (2013) attempts to capture the full complexity of open data supply, intermediation and use. For example, neither make reference to contexts outside of the developed world where open data practice may be promoted or stifled by very different contextual factors.

It is apparent from this brief overview of the literature that few studies focus on *open data* intermediaries specifically, and that there is a lacuna in terms of empirically-based research that attempts to explain the behaviour of open data intermediaries as key actors in data ecosystems. The research questions of this synthesis study are therefore as follows: How do open data intermediaries promote and/or enable the flow of open data in developing country contexts in order to increase the probability of open data use and impact?

Conceptual framework

Given the limited amount of research on open data intermediaries and the requirement for a stable and robust understanding of what an open data intermediary is in order to operationalise the research question, we present a working definition of an open data intermediary.[1] The definition of an open data intermediary used in this paper is as follows: An open data intermediary is an agent (i) positioned at some point in a data supply chain that incorporates an **open** dataset, (ii) positioned between two agents in the supply chain, and (iii) facilitates the use of open data that **may** otherwise not have been the case.

---

[1] This definition has been formulated based on the insights gained from examining the ODDC cases. In addition, the definition benefited from the input of a broad range of stakeholders at three separate workshops, one in Berlin in July 2014, a second in Guimaeraes in October of the same year, and a third in Jakarta in February 2015.

A narrow definition of open data exists in the form of the Open Definition.[2] Broadening the definition creates space for contestation founded on contextual variances. For example, an unequivocal position on open data being machine-readable or an insistence that a lack of explicit licencing limits the uptake of open data are contested. It is therefore important that the definition presented here is not conditional on a particular understanding of what open data is (Davies 2014). The definition presented here is therefore intended to be agnostic; in other words, the definition of an open data intermediary is not dependent on the definition of open data.

It is implicit in the definition that there is a difference between an *open* data intermediary and a data intermediary. For an agent in the data ecosystem to be considered an open data intermediary, it is a requirement that open data be located at some point in a supply chain in the ecosystem (see Figure 1). The implication of this is that the end product in the chain may not necessarily be open. The defining condition is not the final data output's openness but that open data located at some point in the supply chain enabled the reuse of data.

Figure 1: Conditions for a data supply chain to be considered open

Source		End product	SUPPLY CHAIN	
			OPEN	CLOSED
Open	⇨	Open	✓	
Closed	⇨	Open	✓	
Open	⇨	Closed	✓	
Closed	⇨	Closed		✓

An open data intermediary may neither access nor supply open data but may nevertheless facilitate the flow of data in a supply chain by unblocking a process in the chain. For example, such an intermediary may broker an agreement between two agents in the supply chain without actually working with data in any way.

Theoretical framework

Intermediaries as 'keystone species' is a concept present in the emerging literature on ecosystems as a useful framework for understanding the dynamics of open data systems (Harrison et al. 20120. The function or role of keystone species bears many similarities to those of intermediaries. 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 2004:7).

However, the biological ecosystems approach fails to capture inherent power dynamics, in particular the power imbalances and capital flows that exists in a social ecosystem. While ecosystems tend to equilibrium, all actors in the system are not equal. This is particularly evident in a developing country context, where the increased imbalance in the distribution of economic,

---

[2] http://opendefinition.org/

social or cultural power may increase the importance of capital-laden intermediaries in facilitating the flow of data in the system and thus enabling data transactions between actors. It is for these reasons that the ecosystems approach, although providing some important insights into the relative positions of intermediaries in an ecosystem, may be less revealing in terms of *why* intermediaries connect actors in a social environment and what the roles and functions of intermediaries are in the dynamics of open data systems.

In this paper we adapt Bourdieu's model of space, fields, situations, habitus and capital as an alternative and, in our view, more promissory explanatory framework of interactions that characterize actors, including intermediaries, in social systems.[3]

According to Bourdieu, the social world constitutes a multidimensional structure in which individuals are positioned depending on three relational, almost symbiotic, parameters: field (and its local variation, situation), habitus and capital. The social space is divided into a number of, fairly autonomous, *fields*. A field is a structured and dynamic portion of the space that is defined by its own rules and principles of action governing relations in which the actors can engage. It is a network of all direct and indirect, close and remote connections between actors. The notion of field also includes the actor's properties and their power structure, such as hierarchy and domination patterns, and all the types of capitals possibly employed. It is a system of objective coordinates in relation to positions – it is a multidimensional grid of possible stances and moves that an individual can adopt (Bourdieu & Wacquant 1992:97, Fox 2014:207). Therefore, in general terms, the field can be understood as an environment that constrains and, to a degree, determines the existence and motion paths of actors (either those who are already present or those who entering it), as well as the types of capital which they can employ to position themselves in this space (Bourdieu 1993, Reed-Danahay 2004, Wacquant 2006).

A field can also be imagined as an incessant battle for the position in the space (for instance, their hierarchical location) in relation to costs and profits. Agents (individuals or groups of them) compete to gain advantages by means of capital, available and meaningful to this portion of the space, and for maintaining or overturning the rules of the game regulating this field and thus its special distribution (e.g. hierarchy). It is, therefore, a conflict zone between those who strive to preserve the autonomy of the field (these actors usually want to preserve the dominant and well-established position they occupy and the *status quo* of the distribution of capital that enables it) with those who try to introduce less autonomous principles, seeking influence from external, neighbouring or even intruding and clashing fields (these actors, do not occupy a dominant positon, but want to improve it by involving non-orthodox, heterogeneous and subversive forces that challenge the extant *status quo*) (Reed-Danahay 2004, Wacquant 2006).

It is evident that fields are not stable but inherently dynamic. This non-equilibrium property also stems from the inherently relational structure of a field. To be exact, a field not only determines the individuals' actions and the types of benefits' accumulation and conversion, but is also – simultaneously and in an inverse manner – conditioned by these individuals' habitus and the

---

[3] Certainly, the ecosystem narrative can introduce the notions of predator and prey, as well as control over or access to energy resources (of any type) as metaphors for power and capital. Therefore, we do not discard the possibility that some versions of an ecosystem model may be employed for studying intermediaries. However, given that Bourdieu' theory of field, habitus and capital directly incorporates the ideas or power and assets flow or disproportion, and seems to be easily transposable to the situation of intermediation in a social context, we have chosen to use in this study.

capital that penetrates and nourishes the field that jointly work towards preservation or disruption of this field.

Habitus is a complex of durable, observable and/or publicly expressed dispositions an individual develops given the field in which it exists. Habitus is usually analysed in terms of internal (cognitive) predispositions such as beliefs, perceptions, tastes, interests, opinions, etc. that characterize an individual or a collective body in response to external structures (e.g. familiar, social and educational milieus) found in the field. Habitus is thus a structure, embedded in an individual that emerges from all the actions an individual has performed or all the engagements he or she has been involved in during the whole life (Bourdieu 1990:53). It is a product of the entire past, a product that is, however, actively present as disposition (Bourdieu 1990:56). This means that habitus understood as a set of dispositions evolve by participating in a given filed and stem from the constraints imposed by it. Therefore similar experiences in a given field tend to yield similar habitus and, thus, can characterize a whole group of people. But the influence of the field on habitus – and thus the dynamic constitution of habitus – is also regulated by the habitus itself, as habitus filter the field's impact. It is also responsible for continuity and discontinuity of a field since it can both inculcate and accumulate social behaviours typical of the field in individuals, and, having been modified, produce transformation and revolutions in the field (Bourdieu 1984, Wacquant 2006). This implies that habitus is both a structured patient (it reflects the field with its capital) and a structuring agent (it moulds the influence of the field and contributes to the maintenance or disruption of the field).

Since a field constrains habitus, being at the same conditioned by it, and inversely since habitus influences the field being simultaneously determined by it, habitus and field are entirely relational phenomena in the sense that they can only be analysed in relation one to another. Both contribute one to another and depend one on another. Habitus evolves over times due to the chaining field and contributes to the changing field as well. Habitus is a manifestation of the field that embeds and constitutes it, and the field embeds the habitus by which it is expressed and constituted. The intricacy of this relationship is interminable – both are situated simultaneously one in another and no one is given as primary.

If a particular habitus is fully harmonious with the filed in which it exists and remains in a so-called doxic situation that encourages, preserve and reproduces the current structure of the field, the field operates smoothly, because the individual fully respects the running of the system as natural and commonsensical. However, in cases where there is no correspondence between the habitus and the field, in which the former has been developing, some response is expected. This response can preserve the field, modify it, or disrupt it (Bourdieu 1984, Reed-Danahay 2004, Swartz 2002, Wacquant 2004:315-319).

Individuals (or a group of them) are positioned in the social space not only in respect to the two coordinates mentioned thus far, i.e. structure of the field to which they belong and their own habitus, which records actual stances (the path in the field) and predicts potential positions these individual can adopt, by also by the volume and type of the capital they hold. To put it simply, field and habitus depend on capital (Bourdieu 1990:119, Wacquant 2006).

Capital is another crucial element of the social space that both conditions and is conditioned by the other coordinates such as field and habitus. Capitals correspond to the accumulation and/or convertability of advantages and are capable of discriminating agents because of their distinct positions in the field (Bourdieu 1984, 1986; Zhang 2010; Halford & Savage 2010). Traditionally, there are three main species of capital: economic, cultural and social. Economic capital usually

involves economic assets (e.g. monetary value), cultural capital makes reference to knowledge and experience (e.g. competencies and qualifications), and social capital is understood as institutionalised connections or social networks with which an individual is bestowed (e.g. friends, acquaintances and memberships). There is also another, fourth type, referred to as symbolic capital which corresponds to any form of capital that is not regarded as such (e.g. respect, reputation and fame) (Bourdieu 1984, 1986, 1996:148, Wacquant 2006). Recently further species have been distinguished such as technical capital (Zhang 2010) and scientific capital (Langa 2006).

If an individual who possesses a type of capital exerts his or her power on an individual who possesses less, and gets him or her to perform an action, violence can take place. Violence occurs if the dominant imposes his or her principles (usually in a doxic situation, congruous with the functioning of the field and its maintenance, which perpetuates his or her domination) on the dominated. However, the relation may be less drastic and involve a trade where both parts can exchange their assets.

It is important to emphasise that capital does not have intrinsic value in terms of being advantageous but only makes sense in relation to a field in which it is employed. In one filed – or even for some of its agents – a given form of capital is highly advantageous, while in other fields its advantage is less or even worthless. Therefore, capitals can be converted so that the assets that are not advantageous in one field or in one of its sectors (e.g. among certain individuals) be advantageous in others. Capital, thus, corresponds to assets that not only are but that also may produce advantages (Halford & Savage 2010:944-945).

Capital – by being modified for distribution (transfer) or type (conversion) – can importantly structure the field in which it has been used: the change in the capital will modify the field and thus the habitus of its actors.

Sometimes, a mediating notion between field and the actors with their habitus is posited, namely a *situation*. This stems from the fact that a field can in fact involve a great number of possible practices that constraint actors differently leading to formation of different habitus. In other words, actors never perform in fields as such but exclusively in the field's sub-section (a specific situation), which sometimes may be located in the border zone of two or more fields (Santoro 2011). Thus, a situation is a more fain-grained approach to the field-position of an individual which, contrary to autonomous fields, allows for fuzzy areas and mixing of generally independent spaces.

This exposition of Bourdieu's ideas demonstrates that an individual can be located in the social space by using three main coordinates: field (situation), habitus and capital. Not only individual, but the functioning of the entire system and, thus, the structure of power relations, depend on these three parameters. What is important is that the three are relational and interconnected, constantly influencing each other and shaping the overall system. As a result, the social space can be imagined as a complex, living body, possibly with self-organising properties and dynamic, unsettled, non-equilibrium behaviour (Fox 2014:207-210, Swartz 2002, Robbins 2002, Swedberg 2011, Wacquant 2006).

Using Bourdieu's ideas as a narrative model for intermediaries, the following can be postulated. The general environment in which data systems and their transmission take place in developing countries (with the structures found in a state, power relationships, exiting individuals, physical and social arrangements, etc.) can be viewed as a relatively autonomous field. Each particular

case of transaction constitutes a situation *s* in this general frame, in which two (or more) agents are involved: an agent α (possibly dominant due to possession of an asset in the form of data) with a particular habitus and capital (represented by a dynamic function $f(\alpha)$) and another agent β (possibly dominated due to a lack of material or symbolic resources expressed in general terms as a deficit) also with a determined habitus and capital (function $f(\beta)$). Both functions solve for the two agents, possibly predicting their most likely paths in the field and responses to its structure and possible situations in which they can actually engage. However, the relation between the two agents is possible in the situation *s* only (or principally) because an intermediary actor γ (with his or her own habitus and capital, and path $f(\gamma)$) emerges and affords for this situation in which the habitus of the agents α and β can meet and a transfer or conversion of capitals can take place. The more the path $f(y)$ intersects with the path $f(\alpha)$ and $f(\beta)$ – i.e. the more proximate it is to the both sides of a transaction – the more likely it is that such a transaction will be successful.

Research method

The study is based on the analysis of a sample of cases extracted from 17 published ODDC case studies. The sources of the data were the final reports published on the ODDC website.[4] Intermediaries were selected for inclusion in this study based (i) on the definition of an open data intermediary provided in this paper; and (ii) on the availability of sufficient data on the intermediary in question. 'Sufficient data' constituted published information on the value that intermediaries provide in order to link agents in a given open data ecosystem, in so doing promoting the use of data in the ecosystem.

Data were extracted by means of textual analysis of the ODDC case study reports. Where possible, the websites of intermediaries included in the study were consulted in order to supplement the data extracted from the case study reports. Analysis comprised of establishing the two agents between which an intermediary is located followed by an estimation of how the intermediary is able to connect between the two agents – in other words, deducing what types of capital the intermediary possesses to allow for a connection to be made. In order to make this estimation, the deficit of the recipient agent was inferred from the case study text, and a determination was made as to the value that the intermediary provided in order to connect asset-holding and deficit-exhibiting agents in the open data supply chain.

Data were captured in template tables in MSWord to allow for richer, more narrative data to be recorded as this was seen to be necessary in being able to determine what types of transaction-enabling capital intermediaries possess. Tables were classified by field and the table template was structured to capture the following data on each intermediary: Agent$_{asset}$ (Name, Type, Asset); Intermediary (Name, URL, Organisational type, Value provided, Type of capital provided to enable the transaction, Revenue model, Incentive); Agent$_{deficit}$ (Name, Type, Deficit). See Appendix 1 for a template used for each intermediary.

Data were collected on 32 intermediaries; 27 from Asia and 5 from Africa. The preponderance of Asia-based intermediaries was due to the fact that ODDC case studies focusing on Asian countries focused more narrowly on intermediaries whereas the African studies tended to focus on other aspects of open data.

---

4 http://www.opendataresearch.org/reports

Limitations

The study has relied heavily on secondary data for its analysis. This secondary data took the form of case study reports produced for the Emerging Impacts of Open Data in Developing Countries project. The case studies were not conceived or written with intermediaries in mind, although most case studies relied on a conceptual framework developed by the project and this conceptual framework acknowledged the role of intermediaries in open data ecosystems.

Intermediaries may rely predominantly on one data source or they may draw on several data sources, both open and proprietary. In this paper, because of a reliance on existing case studies, the data source used by an intermediary included in the case is often related to and confined by the focus of the case study. For example, if the case focused on budget data, then the case of a particular intermediary may only include in its description the use of a single government budget data source by the intermediary. However, this does not necessarily imply that such an intermediary does not extract and re-use data from other sources. In this sense, this paper is limited to a slice or cross-section of particular data supply chain in a specific data ecosystem as presented in the ODDC cases.

Our analysis may at times create the impression that open data supply chains are linear and/or static. This is clearly not the case. Descriptions are inevitably of a particular arrangement at a particular point in time; however this does not imply that the open data supply chains are not complex and dynamic.

Findings and discussion

The findings of the examination of 32 open data intermediaries to determine types of capital possessed in order to facilitate a situation where the intermediary actor γ (with his or her own habitus and capital, and path $f(\gamma)$) emerges and affords for this situation in which the habitus of the agents α and β can meet and facilitate a transfer or conversion of capitals, are summarised in Table 1 below.

Table 1: Types of capital possessed by open data intermediaries in order to facilitate data flow in specific fields

Type of capital (n=32)				
Economic	Cultural	Social	Symbolic	Technical
9%	6%	31%	3%	97%

The value of capital in understanding open data intermediaries

From a theoretical perspective the findings support the use of Bourdieu's theories of social interactions (and his concepts of situations, habitus and, in particular, capital) as a narrative model for open data intermediaries.

In most of the cases analysed (97%), intermediaries deployed their technical capital to collect, digitise, clean, reorganise and translate data (most often from governments) into information. There was less evidence of other forms of capital being deployed. However, there is some evidence of technical capital being used in conjunction with economic (6%), cultural (6%), symbolic (3%) and social (31%) capital.

In Kenya, both government and donors supported the Code4Kenya initiative to develop applications to increase the effective use of data. However, iHub's research shows that usage levels are paltry (Mutuku & Mahihu 2014). This could point to a preponderance and overemphasis on the value of technical capital in one field, and low levels of social capital among application developers in another field (where potential users are located). If developers/intermediaries possessed higher levels of social capital, this could make intermediaries more attuned to the needs of citizens and increase the likelihood of end-use. In other words, technical capital may enable a transaction, but the value of the transaction is limited by the fact that the transaction results in low levels of return for citizens. This finding resonates with that of Intarakummerd and Chaoroenporn (2013) who highlight the role of intermediaries in compensating for a lack of social capital in innovation systems in developing countries.

Arghyam, an Indian NGO that manages the India Water Portal, acknowledges the limits of its reach. It "works primarily through partnerships with non-profit organisations, CSR divisions of multinational corporations and the media; [… and] with volunteers from a wide range of backgrounds, disciplines and locations, who contribute their valuable time and energy to the cause of spreading awareness on and sharing solutions for India's water problems."[5]

PRS in India relies only on technology to engage with citizens but holds workshops to engage with journalists. "Engagement with citizens is facilitated through the PRS Blog, Twitter and Facebook pages. Workshops are held for journalists on tracking the activities of MPs and MLAs. In addition, PRS provides inputs to the press and electronic media on the legislative agenda in Parliament, as well as data on legislative performance. Members of the PRS team are often approached to contribute columns to provide a perspective on various key Bills."[6] This could be indicative of PRS's social capital enabling links with the media but not necessarily with citizens.

In the case of the Accountability Initiative in India, there is evidence of an intermediary using its cultural capital to make successful use of right-to-information legislation in order to extract data from government: "The data has been collected through surveys and government owned data bases ... In addition, Right to Information applications were filed to secure access to information under the control of public authorities."[7]

Our findings and the examples provided above point to the value of different types of capital in connecting data supply and use. They also point to the limits of an overreliance on technical capital in connecting users with open data.

False intermediaries

Intermediaries may assume that they are providing value, and asset holders such as government and donors may hold a similar view. However, this amounts to a perception of value rather than actual value. While this study did not focus on the uptake of data or information by end-users, some of the cases in this study show that use, let alone effective use, is low. This may indicate a gap between perceived and actual value provided by open data intermediaries.

---

[5] http://schools.indiawaterportal.org/aboutus
[6] http://www.prsindia.org/aboutus/what-we-do/
[7] http://www.accountabilityindia.in/paisa_states

The distinction between actual and perceived use is a critical one as it is a determining factor in whether an agent can in fact be described as an open data intermediary. Based on the definition of an open data intermediary presented in this paper, an intermediary must be positioned between two agents in an open data supply chain. If an intermediary is located between an agent and a purported end user in the chain, but there is no evidence of the end user actually making use of the data, then the intermediary is not in fact an intermediary. In such a case, the ostensible intermediary is in fact the end user, and represents the last instance of agency in the open data supply chain.

Layers of intermediation

In South Africa, open data on public higher education performance is made possible by the presence of two intermediaries (Van Schalkwyk et al. 2014). The first intermediary in the supply chain has a long-standing relationship with key personnel in the South African government department responsible for collecting higher education data. This intermediary therefore possesses the social capital to access closed government data; data that the second intermediary would not be able to access without recourse to legal proceedings. The first intermediary also possesses the technical capital that enables him to extract data from the complex data tables in the government database, and to reorganise, validate and repackage the data into formats usable by the second intermediary in the chain. The second intermediary has a reputation for producing high-quality research on South Africa higher education, and therefore possesses the symbolic capital to confer on the published data a high degree of reliability and confidence amongst end-users in the field.

In the case of the Karnataka Learning Partnership in India, there is an explicit acknowledgement of the limits of technology in connecting with users: "The programme data and the Share-Your-Story component ... in its current form, excludes the majority of our intended target audience – the parents of children who go to government preschools and primary schools who are mostly illiterate and do not have online access due to lack of electricity, computers, computer educators, Internet connections, local-language content etc."[8] In other words, while Karnataka has used its technical capital to consolidate and translate raw data into usable data, it concedes that this is only part of the task at hand. Given its target audience, it needs to deploy its social capital in other ways in order to connect the parents of school-going children to the information provided by the portal. Or it may lack the social capital in this field and will have to resort to soliciting a new intermediary with sufficient in the users' field in order to connect Share-Your-Story to parents.

These cases point to what we believe is an often overlooked and critical dimension in open data intermediation: intermediation does not only consist of a single agent facilitating the flow of data in an open data supply chain; multiple intermediaries may operate in an open data supply chain, and the presence of multiple intermediaries may increase the probability of use (and impact) because no single intermediary is likely to possess all the types of capital required to unlock the full value of the transaction between the provider and the user in each of the fields in play.

Based on our findings, and in line with the theory that influence is increased the closer an intermediary is to the source of power, we would suggest that proximity is an indicator of the extent to which open data intermediaries are able to intermediate effectively (Lorenzen 2006; Barnett & Duvall 2005). And proximity can be expressed as a function of the type of capital

---

[8] https://klp.org.in/about/

that an intermediary possesses. However, because no single intermediary necessarily possesses all the types of capital available to link effectively to all sources of power in or across fields, multiple intermediaries with complementary configurations of capital are more likely to connect between power nexuses. For example, an advocacy group may possess the symbolic or cultural capital required to apply effective pressure on government to release open data. However, the advocacy group will most likely lack the technical capital required to facilitate the publication of the data in useful formats. The tech community may lack the cultural or symbolic capital to negotiate effectively the publication of government data, but it is likely to have the technical capital required to develop applications or to interpret large datasets, i.e. to make the data usable. Neither the advocacy group nor the tech community may be well connected to the potential users of open data because both lack the requisite social capital in that field. Community-based organisation or professional bodies may possess the social capital required to access possible user groups and, as such, may function as effective user aggregators in linking open data to users.

The model in Figure 2 presents the multiple layers of intermediation between a data source and end-use, with each intermediary deploying its own relatively strengths as expressed by the type of capital it possesses in order to connect actors and to facilitate the effective reuse of open data.

Figure 2: A model of layers of intermediaries connecting a data source with users
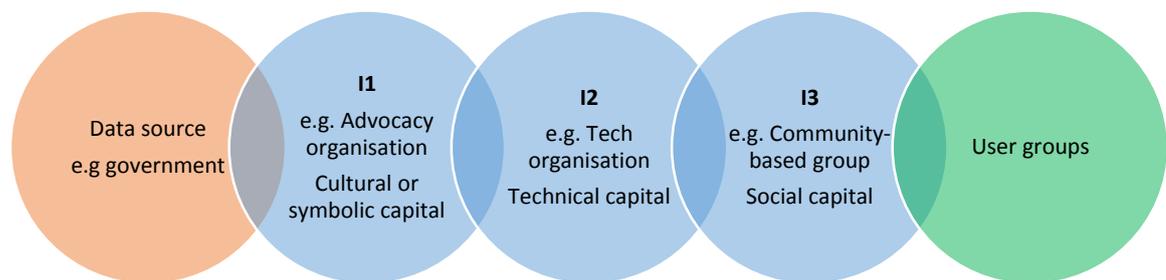
In Bourdieu's terms, given the complexity of the field, it is unlikely that the function *f(x)* of the intermediary intersects ideally both with the function *f(α)* of the initial data supplier and the function *f(β)* of the final data receptor(s) as these two latter may be topologically distant. Rather, it will intersect with the functions *f(y₁) and f(y₂)* of individuals that are proximate to it, in this field or situation. The more topologically adjacent the functions of the interacting individuals are, the more likely the transaction between them – in this case, the flow of data – will be. Therefore, in order to ensure the transaction between the individuals *α* and *β*, a chain of topologically proximate functions *f(yₙ)* should be established, where each transactional link involves functions that typologically intersect. In short, the presence of multiple open data intermediaries may improve use and impact of open data.

We may further postulate that since different parts of the field (or different situations) are inhabited by distinct individuals with dissimilar habitus, they (these situations) may require different types of capital for transactions to be successful. The more remote the regions of the field are the more probable is that they will be governed by distinct forms of capital. What constitute assets at the initial portion of the data-flow chain (connection between the

government and an intermediary), may not be so at its final fragment (connection between an intermediary and the receptor(s)). Inversely, the more proximate the individuals are, the more likely it is that their capitals will be similar or at least convertible. Thus, the fact of possessing a similar form of capital can be a tangible measure of proximity between the actors involved in the data flow. Accordingly, ensuring that two parts of each transactional link employ or are interested in an analogical type of capital (and hence ensuring their topological proximity) can importantly improve the data flow from the source to the final receptor, as the model directly relates the success of a transaction to the proximity of the actors involved in it.

In order to enhance the transaction between these technically specialised agencies (intermediaries) with the data suppliers (e.g. governments) and data users (e.g. individuals), who are less responsive to technicalities, but who are more concerned with symbolic and/or social forms of capital as they directly interact with humans, at least two further links (apart from the intermediary with its technical capital) seem to be necessary. One will connect the technical intermediary with the data suppliers, while the other will relate this technical intermediary to the data receptors. Such two additional intermediaries must have a type of capital that is attractive both to the supplier/receptor and the technical intermediary. However, the fact that the field can be extremely complex and consists of a great number of situations, in which distinct capitals play a crucial role, may necessitate a number of proximate intermediaries with similar capital types so that the transaction between the distant supplier and receptor can take place and be successful.

Technical capital as a distinct form of capital

Given that the technical capital is especially pertinent to data treatment and processing, it is crucial competence for individuals who directly deal with the data in a professional way, by opening it, retrieving, reproducing, etc. What is less clear from this study is the value of introducing technical capital as a new type of capital into a field or situation. Perhaps technical capital is no more than a proxy for other established forms of capital. If, for example, one were to determine the qualifications, memberships and social status of the founders and/or directors of the intermediary organisations included in this study, it is conceivable that these intermediaries entered into a field not because of their technical capital but rather because of their cultural or social capital. In the same way, intermediaries may be using their economic, social, cultural or symbolic capital to connect with and attract actors with the requisite technical competences to enter a field of situation. In effect, the transaction between those who own data assets and those who exhibit a deficit with regard to data is made possible, in the first instance, by an intermediary's cultural, social and/or symbolic capital, and only in the second instance by acquiring the technical skills required to connect the data asset with communities who do not possess the skills or resources (economic or symbolic) to do so themselves. In other words, after the means or social mechanism for transacting had been secured, the acquisition of technical skills follows.

Further research that explores the qualifications, memberships and social status of the founders and/or directors of open data intermediary organisations, and which attempts to determine whether these attributes, rather than their technical capacities, enabled them to enter the field and to play and intermediating role, could go some way to provide greater clarity on the extent to which technical capital is a distinctive and useful type of capital in explaining why intermediaries enter specific data ecosystems.

Conclusion

The adaptation of Bourdieu's ideas to the problem of intermediaries is not a panacea likely to model all the situations in which intermediaries are involved. Nevertheless, the theory of field, habitus and capital can provide a uniform framework for explaining certain robust characteristics of intermediaries and their agency. Furthermore, it may unveil traits previously imperceptible, and may predict the behaviours of intermediaries and their most likely impact in a data ecosystem, thus leading to development of predictive falsifiable hypotheses which can subsequently be tested.

Notwithstanding the limits of any framework seeking to reflect social reality combined with a reliance in this study on secondary data that did not always reveal the richness of data required to conduct more an in-depth analyses of open data intermediaries, the field, habitus and capital framework has revealed two valuable insights on open data intermediaries. First is the value of different types of capital in connecting data supply and use, and the concomitant acknowledgment of the limits of an overreliance on technical capital in connecting users with open data. Second is that intermediation does not only consist of a single agent facilitating the flow of data in an open data supply chain; multiple intermediaries may operate in an open data supply chain, and the presence of multiple intermediaries may increase the probability of use (and impact) because no single intermediary is likely to possess all the types of capital required to unlock the full value of the transaction between the provider and the user. It is hoped that these two insights alone not only provide fertile ground for further research but that they will make funders, policy-makers and advocates work in the area of open data more attuned to the important contribution open data intermediaries have to make in ensuring the realisation of the oft-lauded benefits of open data.

References

Al-Sobhi F, Weerakkody V & Kamal MM (2010) in Sein M & Furuholt B (2012) Intermediaries: bridges across the digital divide. *Information Technology for Development* 18(4): 332-344. DOI: 10.1080/02681102.2012.667754

Bailey A (2009) in Sein M & Furuholt B (2012) Intermediaries: bridges across the digital divide. *Information Technology for Development* 18(4): 332-344. DOI: 10.1080/02681102.2012.667754

Barnett M & Duvall R (2005) Power in International Politics. *International Organization* 59(1): 39-75.

Bhatnagar S (2003) in Sein M & Furuholt B (2012) Intermediaries: bridges across the digital divide. *Information Technology for Development* 18(4): 332-344. DOI: 10.1080/02681102.2012.667754

Bourdieu P (1984) *Distinctions*. London: Routledge.

Bourdieu P (1986) The forms of capital. In *Handbook of Theory and Research for the Sociology of Education*, ed. J. Richardson, 241-258. New York: Greenwood.

Bourdieu P (1996) *The Rules of Art: Genesis and Structure of the Literary Field*. Stanford CA: Stanford University Press.

Bourdieu P (1990a) *The Logic of Practice.* Stanford CA: Stanford University Press.

Bourdieu P (1990b) *In Other Words: Essays Towards a Reflexive Sociology*. Stanford CA: Routledge.

Bourdieu P (1993) *The Field of Cultural Production*. New York: Columbia University Press.

Bourdieu P & L J D Wacquant (1992) *An Invitation to Reflexive Sociology*. Cambridge: Polity Press.

Brown LD & Kalegaonkar A (2002) Support organizations and the evolution of the NGO sector. *Nonprofit and Voluntary Sector Quarterly* 31(2):231-258. http://nvs.sagepub.com/content/31/2/231.full.pdf

Canares M (2014) Opening the Local: Full Disclosure Policy and its Impact on Local Governments in the Philippines. Accepted for presentation at the Eighth International Conference on Theory and Practice of Electronic Governance, October 27-30, 2014, pp. 89-98. doi: 10.1145/2691195.2691214

Chattapadhyay S (2014) Access and Use of Government Data by Research and Advocacy Organisations in India: A Survey of (Potential) Open Data Ecosystem. Accepted for presentation at the Eighth International Conference on Theory and Practice of Electronic Governance, October 27-30, 2014.

Davies T (2014) Open Data in Developing Countries - Emerging Insights from Phase I. ODDC Report. http://www.opendataresearch.org/content/2014/704/open-data-developing-countries-emerging-insights-phase-i

Davies T, Perini F & Alonso J (2013) Researching the Emerging Impacts of Open Data: ODDC Conceptual Framework. ODDC Working Papers #1.

Deloitte Analytics (2012) Open Growth: Stimulating Demand for Open Data in the UK.

Edwards AR (2002). The moderator as an emerging democratic intermediary: The role of the moderator in Internet discussions about public issues. *Information Polity* 7(1): 3-20.

Fox E (2014) Bourdieu's relational view of interactions: a reply to Bottero and Crossley. *Cultural Sociology* 8(2): 204-211.

Fransman M (2010) *The New ICT ecosystem: Implications for policy and regulation*. Cambridge: Cambridge University Press.

Gorla N (2009) in Sein M & Furuholt B (2012) Intermediaries: bridges across the digital divide. *Information Technology for Development* 18(4): 332-344. DOI: 10.1080/02681102.2012.667754

Gurstein M (2011) Open data: Empowering the empowered or effective data use for everyone? *First Monday* 16:2. http://journals.uic.edu/ojs/index.php/fm/article/view/3316. DOI: 10.5210/fm.v16i2.3316

Halford S & Savage M (2010) Reconceptualizing Digital Social Inequality. *Information, Communication & Society* 13(7): 937-955.

Hargadon A & Sutton RI (1997) Technology brokering and innovation in a product development firm. *Administrative Science Quarterly* 42: 716-749.

Harrison C, Pardo TA & Cook M (2012) Creating open government ecosystems: A research and development agenda. *Future Internet* 4: 900-928. doi:10.3390/fi4040900

Howells J (2006) Intermediation and the role of intermediaries in innovation. *Research Policy* 35: 715-728.

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 & Zuiderwijk A (2014) Infomediary business models for connecting open data providers and users. *Social Science Computer Review* 32(5): 694-711.

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 JA (2013, April) From open data to information justice. Paper presented at the Annual Conference of the Midwest Political Science Association, 13 April 2013, Chicago.

Langa PV (2006) The constitution of the field of higher education institutions in Mozambique. Unpublished masters thesis, University of Cape Town. http://hdl.handle.net/11427/11301

Lorenzen M (2007). Social Capital and Localized Learning: Proximity and Place in Technological and Institutional Dynamics. *Urban Studies* 44(4): 719-887.

Magalhaes G, Roseira C & Strover S (2013) Open Government Data Intermediaries: A Terminology Framework. Paper presented at ICEGOV2013, October 22-25, 2013, Seoul, Korea.

Mejabi OV, Azeez AA, Adedoyin A & Oloyede MO (2014) Case Study Report on Investigation of the Use of the Online National Budget of Nigeria. http://www.opendataresearch.org/sites/default/files/publications/Investigation%20of%20Open%20Budget%20Data%20in%20Nigeria-print.pdf

Mutuku L & Mahihu CM (2014) Understanding the Impacts of Kenya Open Data Applications and Services. http://www.opendataresearch.org/content/2014/731/understanding-impacts-kenya-open-data-applications-and-services

Nardi B & O'Day VL (1999) *Information Ecologies: Using technology with heart*. Cambridge, MA: MIT Press.

Reed-Danahay D (2004) *Locating Bourdieu*. Bloomington: Indiana University Press

Robbins D (2002) Sociology and philosophy in the work of Pierre Bourdieu, 1965-75. *Journal of Classical Sociology* 2(3):299-328

Roberts A (2014) Making Transparency Policies Work. http://www.freedominfo.org/2014/10/making-transparency-policies-work/

Santoro M (2011) From Bourdieu to cultural sociology. *Cultural Sociology* 5(1) :3-23.

Sapsed J, Grantham A & DeFillippi R (2007) A bridge over troubled waters: Bridging organizations and entrepreneurial opportunities in emerging sectors. *Research Policy* 36: 1314-1334.

Sein M & Furuholt B (2012) Intermediaries: bridges across the digital divide. *Information Technology for Development* 18(4): 332-344. DOI: 10.1080/02681102.2012.667754

Swartz DL (2002) The sociology of habit: the perspective of Pierre Bourdieu. *The Occupational Therapy Journal of Research* 22: 61-69

Swedberg R (2011) The economic sociologies of Pierre Bourdieu. *Cultural Sociology* 5(1) :67-82

Tuula M (2008) "Insightful Encounters – Regional Development and Practice-Based Learning". Conference on Regional Development and Innovation Processes. 5-7 March 2008, Porvoo - Borgå, Finland. Intermediary Organisation in a Regional Development Network. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.494.126&rep=rep1&type=pdf

Van Schalkwyk F, Willmers M & Czerniewicz (2014) Case study: Open data in the governance of South African higher education http://www.opendataresearch.org/content/2014/710/case-study-open-data-governance-south-african-higher-education

Van Schalkwyk F, Willmers M, Czerniewicz L & Musundwa S (2013) Viscous open data: The flow of data in a public university governance ecosystem. Paper presented at ICT for Development Conference, December 2013, Cape Town.

Wacquant, L (2006) Pierre Bourdieu. In: Rob Stones (ed.), *Key Contemporary Thinkers.* London and New York: Macmillan. 1-15.

World Economic Forum (2015) The Local Data Revolution: The Impact of Human Centred Design. Report from the 2015 Mobile World Congress, Barcelona, Spain, 4 March 2015.

Zang (2010) Technical Capital and Participatory Inequality in eDeliberation: An actor–network analysis. *Information, Communication & Society* 13(7): 1019-1039.