

Open Data Intermediaries in Developing Countries

François van Schalkwyk

Stellenbosch University, South Africa

Corresponding Author.

francois@compressedsl.com

Michael Cañares

Step Up Consulting, Phillipines

mikocanares@steupconsultants.com

Sumandro Chattapadhyay

The Center for Internet and Society, India

sumandro@cis-india.org

Alexander Andrason

Stellenbosch University, South Africa

andrason@sun.ac.za

This paper explores the concept of open data intermediaries using the theoretical framework of Bourdieu's social model, particularly his species of capital. Secondary data on intermediaries from Emerging Impacts of Open Data in Developing Countries research was analysed according to a working definition of an open data intermediary presented in this paper, and with a focus on how intermediaries are able to link agents in an open data supply chain, including to grassroots communities. 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.

Van Schalkwyk, F., Cañares, M., Chattanpadhyay, S., Andrason, A. (2016). Open Data Intermediaries in Developing Countries. *The Journal of Community Informatics*, 12(2) (Special Issue on Open Data for Social Change and Sustainable Development), 9-25.

Date submitted: 2015-08-08. Date accepted: 2016-05-09.

Copyright (C), 2016 (the authors as stated). Licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 2.5. Available at: www.ci-journal.net/index.php/ciej/article/view/1146

Introduction

As public institutions open up vast and complex datasets, the expectation is the lives of all citizens will improve as a consequence of 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 all 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 for all, open data intermediaries are seen as playing a crucial role in linking complex open datasets with user needs, particularly in cases when those users come from grassroots communities. However, 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 Zuidervijk (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 (2010, p. 334) argue that, in the case of e-government and governance, intermediaries are critical in the ‘diffusion of services [...], reducing corruption

[...], moderating discussion on democracy [...] and providing e-government, services of various types [...].’

Focusing on what drives ICT ecosystems, Fransman (2010) draws on the work of 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 Fransman’s 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 cooperate 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 the contribution of intermediaries to innovation in developing-country contexts. They suggest that intermediaries in developing countries are able to compensate for a lack of social capital in innovation systems. By using their social capital to link actors in the ecosystem, intermediaries are able to trigger the innovation that propels the ecosystem. Intarakummerd and Chaoroenporn (2013) also differentiate between public and private intermediaries. Their research therefore not only identifies the critical contribution that intermediaries with a particular type of capital play in driving innovation, but also that different types of intermediaries working in concert are required¹.

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). Through the regular use of government data, these organizations are able to demand disclosure of additional data where they encounter gaps in the supply of data, validate the quality of government data and reinforce the value of data sharing, not only between government and other stakeholders, but also between data users within government.

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 (or social capital) 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.

In a developing-country context, several authors have made mention of the importance of trust in enabling intermediaries to connect actors – in open data in particular (Davies 2014) and, more broadly, in ICTs (Srinavas 2007, Diaz Andrade & Urquhart 2010, Sambasivan 2010, all in Bailur & Masiero 2012) and development (Fukuyama, 2001; Lyon, 2000).

¹ See research by Maghalhaes (2013) and Jansen & Zuiderwijk (2014) on different types of open data intermediaries, and Sein (2011) and Sein & Furuholt (2012) on intermediaries in e-government (which can be assumed to include open data intermediaries).

Interpersonal trust can be interpreted as a proxy for social capital – trust is the bonding agent in social networks. Intermediaries that are trusted by actors in different networks are able to bridge discreet networks thereby creating new linkages.

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. There is, however, a suggestion in the literature of the importance of the ability of intermediary actors to link other actors in and between networks. The research question of this synthesis study is therefore as follows: How are open data intermediaries able to facilitate 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 intermediary². 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.

A narrower definition of open data exists in the form of the Open Definition³. 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 is deliberately 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.

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.

² 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.

³ <http://opendefinition.org>

The distinction between actual and perceived use of open data is an important one to make as it is a determining factor in whether an agent in a data chain 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 use, 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.

| Source | ⇒ | End product | SUPPLY CHAIN | |
|--------|---|-------------|--------------|--------|
| | | | OPEN | CLOSED |
| Open | ⇒ | Open | ✓ | |
| Closed | ⇒ | Open | ✓ | |
| Open | ⇒ | Closed | ✓ | |
| Closed | ⇒ | Closed | | ✓ |

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

Theoretical Framework

Intermediaries in open data systems have been likened to biological ecosystems, in which intermediaries are viewed as ‘keystone species’ (Harrison et al., 2012; Nardi & O’Day, 1999; Iansiti & Levin 2004; Van Schalkwyk et al. 2015). However, the biological ecosystems approach is, we believe, predisposed to under-representing the inherent power dynamics in ecosystems, in particular the power imbalances and the capital flows that exist in a social ecosystem. In this paper we adapt Bourdieu’s model of field, habitus and capital as an alternative and, in our view, more promising explanatory framework of interactions that characterize actors, including intermediaries, in social systems.⁴

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.

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 multidimensional grid of

⁴ We do not discard the possibility that some versions of an ecosystem model may be employed for studying intermediaries. However, given that Bourdieu’s theory of field, habitus and capital directly incorporates the concepts of power and assets flow (or disproportion), and seems to transpose effortlessly to the situation of intermediation in a social context, we have chosen to explore Bourdieu’s social theory in this study. We also note that authors such as Helbig et al. (2012) opt for an alternative heuristic – the ‘information polity’ – because it allows ‘for understanding how the creation or modification of data sources, flows, or governance relationships affects interests of various stakeholders’ (Helbig et al. 2012: 11, emphasis added).

possible stances and moves that an individual can adopt (Bourdieu & Wacquant, 1992; Fox, 2014).⁵ Habitus is a complex of durable, observable and/or publicly expressed dispositions an individual develops given the field in which they exist. Habitus is 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 their lifetime (Bourdieu 1990:53).

Fields are not bounded by impenetrable borders. 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). A situation is therefore a more granular approach to the field-position of an actor which, contrary to autonomous fields, allows for fuzzy areas and mixing of generally independent spaces.

Individuals (or a groups of them) are positioned in the social space not only in respect to the structure of the field to which they belong and their own habitus, but also by the volume and type of the capital they hold (Bourdieu, 1990; Wacquant, 2006). Capitals correspond to the accumulation and/or convertibility of advantages, and are capable of discriminating agents because of their distinct positions in the field (Bourdieu, 1984; Halford & Savage, 2010. Bourdieu (1984, 1986) proposes the notion of capital to explain the interaction of actors in social systems.

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, 1986; Bourdieu, 1996; Wacquant, 2006).

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 field a given form of capital is highly advantageous, while in other fields its advantage is less or even worthless. 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).

It is evident that fields are not stable but inherently dynamic. This non-equilibrium stems from the inherently relational structure of the social world whose three main coordinates (field, habitus and capital) are interconnected, constantly influencing each other and shaping the overall system. A field not only determines the individuals' actions (habitus) and the types

⁵ 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. The habitus can be imagined as a collection of paths that reflect the actions of an individual in the field (Bourdieu, 1993). It is analyzed in terms of internal (cognitive) predispositions that characterize an individual in response to external structures found in the field.

of benefits' accumulation and conversion (capitals), but is also conditioned by these individuals' habitus and the capital that penetrates and nourishes the field that jointly work towards preservation or disruption of this field. Inversely, habitus and capital are not only conditioned by the field, but can in a feed-loop manner contribute to it, thus structuring the field in which they have been embedded: the change in the capital or habitus will modify the field (Fox, 2014).

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, existing 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 $f(\gamma)$ 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 ODDC case studies. The sources of the data were the final reports published on the ODDC website (<http://www.opendataresearch.org/reports>). 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.

From the ODDC case studies, 32 intermediaries were identified: 27 from Asia and 5 from Africa.⁶ 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.

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

⁶ For a full list of cases included in the study, see <http://dx.doi.org/10.5281/zenodo.45181>

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 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: Agentasset (Name, Type, Asset); Intermediary (Name, URL, Organisational type, Value provided, Type of capital provided to enable the transaction, Revenue model, Incentive); Agentdeficit (Name, Type, Deficit).⁷

Limitations

The study 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 acknowledges 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 intermediaries⁸ 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.

⁷ See <http://dx.doi.org/10.5281/zenodo.45181> to access the data tables.

⁸ For a list of all 32 intermediaries and the raw data, refer to <http://dx.doi.org/10.5281/zenodo.45181>.

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 likely to be able to connect with 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.

The case of Arghyam, an Indian NGO that manages the India Water Portal, is an example of an open data intermediary acknowledging the limits of its reach ostensibly because it is cognizant that it lacks the capital to connect effectively to users. 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' (<http://schools.indiawaterportal.org/aboutus>).

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' (<http://www.prsindia.org/aboutus/what-we-do/>). 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' (http://www.accountabilityindia.in/paisa_states).

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. Stated differently, the importance of other forms of capital, such as social capital, plays a critical role in connecting data to prospective users. This finding supports the findings of Intarakummerd and Chaoroenporn (2013).

Multiple Intermediaries

In the case study from South Africa, open data on public higher education performance is made possible by the presence of two intermediaries (Van Schalkwyk et al., 2015) (see Figure 2). The first intermediary in the supply chain (CS) has a long-standing relationship with key personnel in the South African government department responsible for collecting higher education data. This intermediary possesses the social capital to access closed government data; data that the second intermediary (CHET) 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 first intermediary combines his social and technical capital to connect government data to the second intermediary. 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 open data a high degree of reliability and confidence amongst end-users in the field. It has also secured funding from an international philanthropic organisation providing it with the economic capital required to retain the services of the first intermediary.

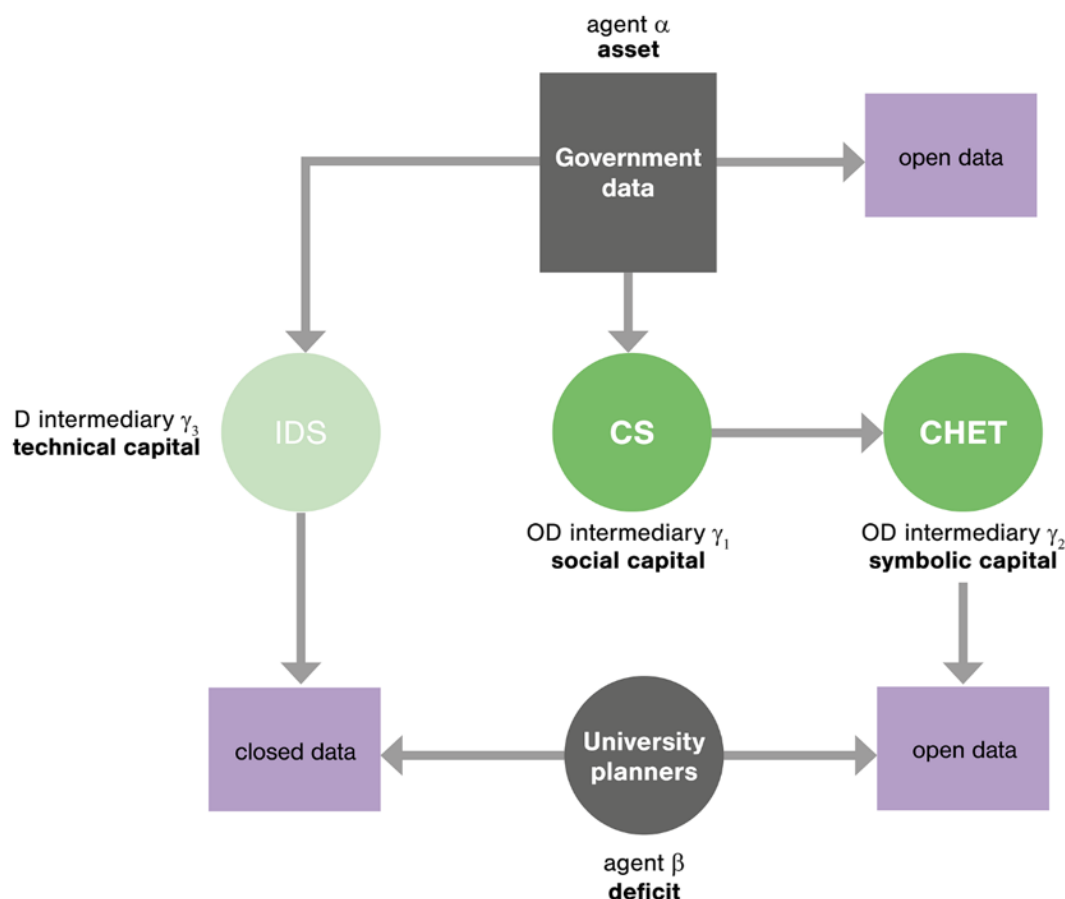


Figure 2: Multiple open data intermediaries in the South African higher education data supply chain⁹

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.’ (<https://klp.org.in/about/>). In other words, while Karnataka has used its technical capital to consolidate and translate raw data into usable open 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 dimension in open data intermediation: intermediation does not only consist of a single agent facilitating the flow of

⁹ Figure 2 has been simplified to illustrate the points made in the text. A more detailed version of this supply chain is available at <https://dx.doi.org/10.6084/m9.figshare.2067318> . For a diagrammatic representation of how this supply chain fits into the larger ecosystem, see <https://dx.doi.org/10.6084/m9.figshare.1040414> .

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 (Lorenzen, 2006; Barnett & Duvall, 2005), we would suggest that proximity is an indicator of the extent to which open data intermediaries are able to intermediate effectively. And proximity can be expressed as a function of the type of capital 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 3 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.

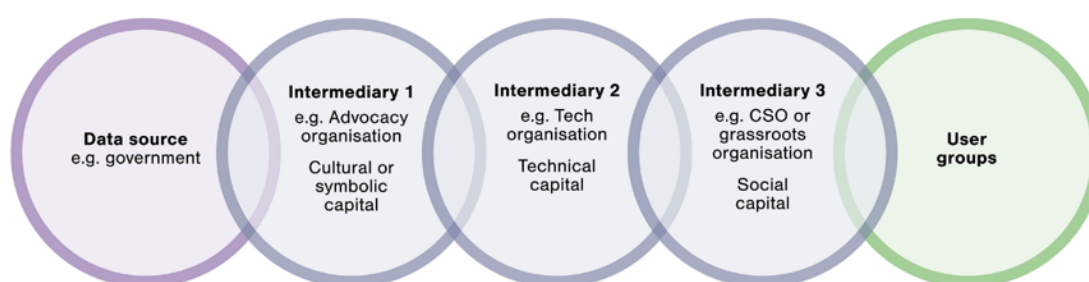


Figure 3: A model of layers of intermediaries connecting a data source with users

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 $f(\alpha)$ of the initial data supplier and the function $f(\beta)$ of the final data receptor(s) as these two latter may be topologically distant. Rather, it will intersect with the functions $f(\gamma_1)$ and $f(\gamma_2)$ 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 $f(\gamma_n)$ should be established, where each transactional link involves functions that topologically 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 it 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 a crucial competence for individuals who deal directly 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 – that is, those of the 97% instances where technical capital was used to link actors in open data supply chains – 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 or 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 an 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.

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 than an in-depth analysis 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, particularly those located in grassroots communities, 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 who work in the area of open data more attuned to how open data intermediation needs to be arranged for the actualization of the oft-lauded benefits of open data for all communities.

Acknowledgments

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

References

- Bailur, S., & Masiero, S. (2012). The complex position of the intermediary in telecenters and community multimedia centers. *Information Technologies & International Development*, 8(1), 27-42.
- Barnett, M., & Duvall, R. (2005) Power in international politics. *International Organization*, 59(1), 39-75.
- Bourdieu, P. (1984). *Distinctions*. London: Routledge.
- Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 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., & Wacquant, L. J. D. (1992). *An invitation to reflexive sociology*. Cambridge: Polity Press.
- Brown, L. D., & Kalegaonkar, A. (2002). Support organizations and the evolution of the NGO sector. *Nonprofit and Voluntary Sector Quarterly*, 31(2), 231-258. Retrieved from <http://nvs.sagepub.com/content/31/2/231.full.pdf>
- Canares, M. (2014, October). Opening the local: Full disclosure policy and its impact on local governments in the Philippines. Paper presented at the Eighth International Conference on Theory and Practice of Electronic Governance, Guimaraes, Portugal. doi: 10.1145/2691195.2691214
- Chattapadhyay, S. (2014, October). Access and use of government data by research and advocacy organisations in India: A survey of (potential) open data ecosystem. Paper presented at the Eighth International Conference on Theory and Practice of Electronic Governance, Guimaraes, Portugal.
- Davies, T. (2014). Open data in developing countries: Emerging insights from phase I (ODDC Report). Retrieved from <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 No. 1).
- Edwards, A. R. (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.

- Fukuyama, F. (2001). *Social capital, civil society and development*. *Third World Quarterly*, 22 (1), 7-20.
- Gurstein, M. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday*, 16 (2). 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, R. I. (1997). Technology brokering and innovation in a product development firm. *Administrative Science Quarterly*, 42, 716-749.
- 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: A White Paper*. Albany, NY: Center for Technology in Development.
- 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*, 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.
- Lorenzen, M. (2007). Social capital and localized learning: Proximity and place in technological and institutional dynamics. *Urban Studies*, 44 (4), 719-887.
- Lyon, F. (2000). Trust, networks and norms: The creation of social capital in agricultural economies in Ghana. *World Development*, 28 (4), 663-681.
- Magalhaes, G., Roseira, C., & Strover, S. (2013). Open government data intermediaries: A terminology framework. Paper presented at International the 7th Conference on Theory and Practice of Electronic Governance ICEGOV, Seoul, Korea.
- Mejabi, O. V., Azeez, A. A., Adedoyin, A., & Oloyede, M. O. (2014). *Case study report on investigation of the use of the online national budget of Nigeria* (Open Data Research Report). Retrieved from <http://www.opendataresearch.org/sites/default/files/publications/Investigation%20of%20Open%20Budget%20Data%20in%20Nigeria-print.pdf>
- Mutuku, L., & Mahihu, C. M. (2014). *Understanding the impacts of Kenya open data applications and services* (Open Data Research Report). Retrieved from <http://www.opendataresearch.org/content/2014/731/understanding-impacts-kenya-open-data-applications-and-services>

- Nardi, B., & O'Day, V. L. (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.
- 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. (2011). The 'I' between G and C: E-government intermediaries in developing countries. *The Electronic Journal on Information Systems in Developing Countries* 48 (2):1-14.
- 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, D. L. (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, March). Insightful Encounters – Regional Development and Practice-Based Learning. Paper presented at the Conference on Regional Development and Innovation Processes, Porvoo – Borgå, Finland. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.494.126&rep=rep1&type=pdf>
- Van Schalkwyk, F., Willmers, M., & Czerniewicz, L. (2014). *Case study: Open data in the governance of South African higher education* (Open Data Research Report). Retrieved from <http://www.opendataresearch.org/content/2014/710/case-study-open-data-governance-south-african-higher-education>
- Van Schalkwyk, F., Willmers, M., & McNaughton, M. (2015). Viscous open data: The roles of intermediaries in an open data ecosystem. *Journal of Information Technology for Development*. doi: 10.1080/02681102.2015.1081868
- Wacquant, L. (2006). Pierre Bourdieu. In R. Stones (Ed.), *Key Contemporary Thinkers* (1-15). London and New York: Macmillan.
- 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).