The Social Shaping of Open Data through Administrative Processes

Sirko Hunnius
IfG.CC – Institute for eGovernment
Kutschstall-Karree, Am Neuen Markt 9c
D-14467 Potsdam, Germany
+49(0)331.7403.6739
shunnius@ifg.cc

Bernhard Krieger
IfG.CC – Institute for eGovernment
Kutschstall-Karree, Am Neuen Markt 9c
D-14467 Potsdam, Germany
+49(0)331.7403.6739
b.krieger@cantab.net

ABSTRACT
Many models have been provided in the last years that aim at describing an optimal open data publication process. However, they fail to explain the different outcomes of open data initiatives. Based on qualitative research this paper conceptualises the open data phenomenon as a set of techno-political arenas in which different interests of a variety of actors potentially and actually collide. The micro-political arena model constitutes an instrument to delineate the social and institutional context of open data that can be employed to explain the successes, as well as the failures of individual open data projects.

Categories and Subject Descriptors
Management of Computing and Information Systems, Data Management Systems, Data Modeling, Information Integration, Digital Libraries and Archives, Resource Description Framework (RDF)

General Terms
Management, Human Factors, Standardization, Theory.

Keywords
Open data, process models, techno-political arenas, social and institutional context.

1. PROBLEM STATEMENT AND RESEARCH QUESTION
The topic of open data is generating considerable interest among researchers, technology developers and practitioners in public administration. The conversation so far often circles around the potentials of open data [8, 20, 26]. However, up until now there is little evidence of any significant economic or societal impact [12]. Regarding the sheer amount of open data anglo-american governments seem to provide far more data and render the latter in a more sophisticated way than governments in continental European countries. This difference is puzzling as open data seems to be an international trend, fostered by an international community and pushed by international advocacy groups (e.g. Open Knowledge Foundation). Multinational initiatives such as the Open Government Partnership are taken up by countries as diverse as the United States of America, Chile, Austria, Russia, Kenya and Malaysia. Considering this heterogeneity of actors on the one hand and the differences in the implementation of open data activities on the other institutional factors appear to play an important role in how open data is perceived and adopted in the different public sector organisations.

This paper investigates the administrative practices involved in the provision of open data by public sector institutions. In particular it analyses the impact of the former on open data, especially on data and meta data quality. Therefore, the focus of this paper is limited on processes in the context of open data within the public administration. Such focus excludes the usage of open data for this endeavour. The research question thus concerns the relationship between institutional arrangements of the involved public sector actors and actual publication activities of the latter. An in-depth scrutiny of open data processes provides a deeper understanding of the kind of data, as well as of its structure, vocabulary, meta data, license etc. employed by governmental institutions.

OpenSym '14, August 27 - 29 2014, Berlin, Germany

ACM 978-1-4503-3016-9/14/08…$15.00.
http://dx.doi.org/10.1145/2641580.2641601

2. OPEN DATA TRAJECTORIES
Various models of (linked) open data have been put forward under different headings. They have been termed the open data life cycle, the open data value chain or plain open data process [30]. The different terminologies illustrate different purposes – practical guidance [13] or analytical separation – and foci. Whereas value chain models focus more on the creation of value during open data usage [15], the life cycle models aim to structure the handling of the data itself. Existing process models focus on activities within public administration, such as generating, editing and publishing the data without paying too much attention on the outside-use.

Most models contain similar elements and differ only regarding semantics, granularity or the extension of the process. Hyland et al. [13] provide a six-step guidance model that contains the steps to (1) identify, (2) model, (3) name, (4) describe, (5) convert, (6) publish the data and the reverse activity to maintain it, similar to Villazon-Terrazas et al. [28]. Another model by Hausenblas et al. [11] also includes the user perspective, adding the steps "discovery", "integration" and "use cases". With the ambition to

build tools to support creating linked data, the LOD2 project developed a more granular 8-step lifecycle model [1]. Synthesizing various models, van den Broek et al. [3] derive a lifecycle model comprising the steps (1) identification, (2) preparation, (3) publication, (4) re-use and (5) evaluation.

All of these models have in common that they describe a consecutive, one-dimensional arrangement of activities that an unspecified set of actors repeatedly undertake in order to provide a formerly unexposed amount of data to an abstract general public. Furthermore, these models incorporate only one analytical level. They exclusively take the operational day-to-day processes into account (such as extracting, cleaning, publishing and maintaining data), while at the same time neglecting the strategic processes (such as policy production, decision making and administrative enforcement). However, for the research question discussed in this paper the operational open data processes appear to be less relevant than the meta process of negotiating and designing the latter. Thereby the decision making processes of which data will be published, who extracts data, how are data edited, how data can be accessed, which licenses are available, how data privacy and liability issues are treated, who is involved in these decisions etc. become central. These more general strategic processes about open data refer to the governance structure, likely to be connected to an organisation's ICT and data governance.

The issues outlined point to another deficiency of most open data process models: These process models are actor-blind. If at all, institutional characteristics and actor-interests are considered as "impediments" [30] or restrictions hindering an implicit normative "open data ideal". However, as all science and technology studies during the last 80 years emphasize the relevance of the social context for the development of technical endeavours it appears to be negligent to entirely ignore the stakeholders of open data during the analytical effort of process modelling. This paper therefore aims to contribute to the discussion around open data provision by integrating both research on the institutional arrangements of data providers, as well as investigation of the situational perspective of the involved actors.

3. METHOD

In order to analyse this exploratory question, a qualitative research strategy appears adequate, because of it adds 'thickness' and conceptual openness to the phenomenon [2, 10]. Thereby, numerous potentially relevant aspects, like actor's perceptions and constructions, as well as institutional aspects can be apprehended [9, 17]. The research goal is to gain a deeper understanding of the processes around open data within public administration.

For this research effort 20 semi-structured interviews were carried out with public administrators involved with open data in their organisation. In order to grasp open data as a broader, multi-governmental and multinational phenomenon interviewees from four continental European countries (Spain, Denmark, Finland and Germany) were selected. In order to reflect the whole spectrum of governmental levels, we interviewed ten participants working in municipalities, two at the regional and eight at the national level. The interviews lasted around one hour each and included three topics: (1) perception of open data, (2) governance structure around open data and (3) actual process of data publication. The interviews were recorded, coded and analysed with the help of the qualitative data analysis software Nvivo. The transcripts were coded based on a coding scheme derived from the relevant literature on the topic which was integrated into the interview guideline. This coding scheme was refined during the coding process considering the empirical data.

4. EMPIRICAL FINDINGS

The interviewees generally perceive open data as an innovative and catchy topic that therefore attracts political, as well as administrative attention. "It [drive for open data] comes from the politicians. [...] XYZ as a city presents itself to the citizens, shows itself as a service provider, [...] as attractive and thus modernity always plays a role. E-participation is also a topic - everywhere en vogue - [...] around the elections. A topic that you can catch attention with." (interview participant, municipal level Germany) In addition, it is vocally supported by national governments, the European Commission and outside advocacy groups mainly from civil society. However, differences are notable in how public sector organisations frame open data. The latter can be distinguished in a stronger emphasis on economic potential, transparency or internal efficiency. This seems to play a role in the decision which administrative unit gains responsibility for open data. Also, the established mode of ICT and data governance influences how open data is adopted and who can design the process of data publication. In the observed cases, mostly central ICT units have taken up open data initiatives. With outside assistance from consultants – who are in some cases open data advocates themselves – they develop an infrastructure and design target processes. The latter are thereupon negotiated with all other administrative units and a joint decision – often also involving the political level – is made, about how to set up an open data platform.

This outlined ignition and concept formation, often involves general decisions about licensing and pricing policy. "It was decided [by a Government Resolution 2011] that it is free [of cost], that means the data as a raw material is free." (interview participant, national level Finland) In addition, the more technical issues are decided, e.g. data access, hosting and data formats. (interview participant, national level Spain). These latter too are highly political, but rarely appear to be discussed or perceived as such [see also 4]. Instead, the discussions mainly focus on which data is published, under which license and whether fees can be charged. This latter point proves especially relevant for organisations who have shared data on a regular basis with third-parties before the concept of open data had even been formed, but which are used to charging fees for the data they share (e.g. meteorological and geographical data). For them, third-party usage of "their" data does not seem as alien as it appears to be for most other public sector organisations. These are to a large extent built around a notion of secrecy, depending on administrative culture and traditions [23, 24, 29]. Thus, involved actors are often reluctant to share data, citing data protection, potential liabilities and resource-scarcity as some of various reasons [see 30]. As a consequence, these organisations that "own" the data play a major role in decisions about which data to publish and ensure a prominent role for themselves in the data publication process.

To comply with an open data program at least on the face of it, most organisations do provide some data. These data can often be published with little effort, is not politically sensitive and does not require frequent updating. "In general, it is the availability of potential data. This is still a point that data are selected based on

how easy it can be made available and less based on its usefulness" (interview participant, municipal level Germany). This strategy of least effort can be termed the "availability approach", in contrast to a strategy that starts out from actual user interests in high-value data sets. Furthermore, data owners also play a vital role in the data publication process. They regularly extract the data, clean it of non-publishable (i.e. personal data) parts and provide meta data. Since they regularly have no specific interest in the re-usage of the data or lack understanding its significance, they do not pay particular attention to conditions of usability. This leads to flawed or lacking meta data or arbitrary vocabulary in the data.

5. DISCUSSION

In the observed cases, the more general questions embedded in process models, e.g. which data to publish and under which license, are not part of the operative data publication process. Instead, most of these decisions are agreed upon in general for all open data or at least restricted to a few available options. Furthermore, these decisions are mostly made by different actors than those actually extracting, cleaning and publishing data. Rather, decisions about data, licenses etc. are made at higher levels of the administrative hierarchy. These decisions are rarely left to the discretion of the administrative clerk handling the data before it is published.

Processes of shaping technology are not restricted to its initial adoption, but are better understood as an ongoing process of negotiation during all phases of technology production [16] and usage [21, 22]. In the context of technology usage by and in government institutions Dovifat et al. modelled the shaping of technology according to its framing and perception (ignition), the design of its adoption (concept formation), its employment (implementation) and its usage (routinisation) [6, 7]. These arenas are not necessarily opened sequentially, but can stay open or be re-opened even after they have closed.

In the context of open data a process model approach appears adequate only for the actual data publication. However, considering the whole institutional situation, particularly for the non-operational and more strategic aspects of open data, arenas for ignition, concept formation, implementation and routinisation seem to provide a far more comprehensive theoretical approach.

Ignition

The concept of Open Data is built around the idea of technology activists to re-use government data primarily in order to raise transparency and create economic value [27]. This concept is heavily value-laden. Political interests play a major role in whether or not political actors advocate the idea and do so not only as an opposition party, but follow through when they come to power [19]. Furthermore, the specific democratic tradition, in particular the history of government transparency and freedom of information (FOI) legislation play a role in how the concept is received [14], for example if there are powerful institutional advocates of government transparency. Especially its relationship to fundamental principles of the public sector has an impact, like security, confidentiality, the administration's obsession with planning on the one hand and on the other hand the ascription of modernity and the air of transparency that surround open data. These principles are not homogeneous even within a country, but can differ among various areas of administration depending on actor interests and the habit of data sharing. In addition, the way influential actors frame the topic has a significant impact on how the latter is perceived and who gains responsibility for open data. For example, if open data is framed as an economic issue that is about jobs and growth, chances are that the ministry of economics becomes in charge of open data. This thereupon has an influence on the concept formation of open data. Thus, the institutional structure and power distribution influence whether and how open data is taken up by various other actors.

Concept Formation

The assigned administrative unit often forms the concept independently, thereafter negotiating it with other departments that are foreseen to later-on publish open data. Mainly technical and legal issues are negotiated in this arena and planned in significant detail. Depending on its legislative mandate, the administrative unit responsible decrees general rules for open data. These may comprise a general license or license scheme under which the data will be published, data format and meta data standards and even the target processes for publishing open data. Depending on the power structure, i.e. the level of autonomy in general and the data and ICT governance in particular, future data providers – often considered "data owners" – reserve considerable leverage for what are seen the major political issues around open data. These are mainly decisions about which data will be published and what data use is permitted. Furthermore, the unit responsible for open data rarely gains power to enforce the rules. The fundamental rules of the game are rarely altered.

While political actors and the higher echelons of the executive mostly dominate the ignition arena, the concept formation arena is largely characterized by negotiations among administrative actors, especially IT units, IT providers and consultants. These actors have a strikingly technical view on open data. They are thus more focused on the technical issues and the target processes of data publication, largely blanking out the fundamental political issues around open data. This said, it is important to understand that the results of these sheer technical discussions themselves have major political impacts that often are only understood by the majority of "non-technical" actors at a later stage.

Implementation

In the implementation arena, the open data platform is developed, tested and implemented. Furthermore, departments adopt the target processes, implement the actual data publication processes and assign responsibilities internally for open data, for example by appointing open data officers in each department. However, even creating such decentralised agents of open data, does not necessarily alter the power structure, if these appointees cannot decide which and how data will be published, but depend on the complicity of the data owners.

The open data unit rarely has leverage on data providing units. Quite on the opposite it depends on their cooperation and willingness to contribute data, since their open data initiative's success hinges thereupon. Thus, during implementation the open data unit can hardly enforce any standards regarding data and meta data quality or data harmonisation. Efforts to harmonise and share data across public sector organisations have proven pretentious in the past, even regarding internal use [25]. Different vocabulary are not simply arbitrary technical differences, but reflect differences in construct formation. Epistemology is often influenced by perceptions, preferences and interests. Subsequently to the implementation, the role of the responsible open data unit

often further degenerates to an assisting and promoting role, without much actual influence.

Routinisation

In the routinisation arena, actors struggle about whether and how to integrate open data into their daily work. Depending on the incentive structure, data owners rarely have a significant interest in publishing data. For the overall success of an open data initiative, they uncommonly share the praise. Instead, they risk blame by exposing flawed data, battling alternative interpretations of their data by users and generating workload to update and correct published data. Thus, they are prompted to comply with the task by name only, publish data that is easily available and poses little political risk. However, this data can rarely be used for sophisticated, illuminating and widely used applications. As a consequence, there rarely emerge eye-catching applications that draw attention and generate a push to provide further data. In addition, the operating departments that own, extract and provide the data have sparse contact with actual users and no specialised understanding of data use. Thus, scant attention is paid to aspects that influence data usability, like vocabulary, structure, meta data, further inhibiting data use.

By hiding their data from outside view and use, data owners try to protect the zone of strategic uncertainty they control, from which they draw their power. They play "routine games", because they do not see any benefit in playing "innovation games" [7]. "[S]ources of power can lose their importance during change processes" [7] and data is one of their sources of power. The innovation game on the other hand would be to make the data available so that third parties can make beneficial use of it. Therefore, actor strategies needed to be changed in the sense that making data available becomes the winning strategy. For example, departments may have an incentive to provide data, if some kind of benchmark measures the publication of data sets, thereby creating competition among departments, provided the benchmark receives political attention. Alternatively, budget is sometimes used to pay-off "losers" who previously earned money for selling data or to compensate for additional workload, thus buying complicity. Depending on the concept formation, decision structures might give the open data unit the power to mandate the publication of certain data sets, thereby redistributing power. Optionally, an institutionalised open data imperative, i.e. that data is open by default, could have significant influence, since it includes a mechanism to constantly make more data available. In the face of meagre results, sufficient support for open data among influential actors at the political level could for example initiate renegotiations in the concept formation arena to redistribute powers.

Apart from the push to publish data, the pull by users has received scant attention in open data literature [19]. However, structural aspects of civil society and technology activists also do play a role, since the individual citizen making use of open data appears to be rather a myth than reality [5, 18]. Therefore, intermediary organizations will continue to play a vital role in assisting individuals to overcome the difficulties associated with interpreting and acting on the information. Their readiness to use data and their ability to create a significant pull-effect is imperative, because of resource-, capacity- and competence-requirements [19].

6. CONCLUSION

The phenomenon of open data is still fairly young. Only in the last couple of years a heterogeneous techno-political movement consisting of transparency activists, civil society organisations, professional politicians and midrange to senior-level public administrators has began to lobby for and foster the publication open data. Half a decade after the launch of the much vaunted US government website data.gov the results of the open data movement are promising, but their impacts still appear impalpable. While in principle there is little to be said against the concept and implementation of open data in order to advance political transparency, economic growth and efficiency gains in the public sector, the everyday practise of open data in Europe, at least, faces problems to provide and update interesting data that can be used to achieve the desired effects.

Many models have been provided in the last years that aim at describing an optimal open data publication process. They rather normatively try to establish a standard approach open data initiatives should follow in order to succeed in their efforts to provide publicly accessible and machine readable government information. However, ignoring institutional factors and political realities in most public administrations on all governmental levels these models fail to explain the existing difficulties of the open data phenomenon.

This paper is an attempt to conceptualise the open data phenomenon differently. Rather than providing a practical template to open data initiatives it aims at delineating the social and institutional context in which the latter operate. Using a theoretical approach rather than a descriptive model as an analytic frame it sheds light on the different developments of the worldwide open data movement. By understanding the open data phenomenon as a techno-political arena in which different interests of a variety of actors potentially and actually collide this paper alludes to the toe-holds of the phenomenon that can both explain the successes, as well as the failures of individual open data projects.

7. REFERENCES

[1] Auer, S. et al. 2012. Managing the life-cycle of linked data with the LOD2 stack. *International Semantic Web Conference 2* (2012), 1–16.

[2] Brady, H.E. et al. 2010. Refocusing the Discussion of Methodology. *Rethinking Social Inquiry*. H.E. Brady and D. Collier, eds. 15–31.

[3] Van den Broek, T. et al. 2011. Walking the extra byte: A lifecycle model for linked open data. *Linked Open Data – Pilot Linked Open Data Nederland*. E. Folmer et al., eds. Remwerk. 95–111.

[4] Courmont, A. 2012. How to Govern Open Data? The politics of open data portals. *Paper presented at the IPSA Conference, Madrid, July 8-12, 2012* (2012).

[5] Davies, T. 2010. *Open data, democracy and public sector reform: A look at open government data use from data.gov.uk*. University of Oxford.

[6] Dovifat, A. et al. 2004. Explaining Successes and Failures of e-Government Implementation with Micropolitics. *Electronic Government, LNCS 3183* (Berlin, Heidelberg, 2004), 308–312.

[7] Dovifat, A. et al. 2007. The "model of micropolitical arenas" – A framework to understand the innovation process of e-government-projects. *Information Polity*. 12, 3 (2007), 127–138.

[8] Geiger, C.P. and von Lucke, J. 2012. Open Government and (Linked) (Open) (Government) (Data). *eJournal of eDemocracy & Open Government*. 4, 2 (2012), 265–278.

[9] George, A.L. and Bennett, A. 2005. Case Studies and Theory Development. *Case Studies and Theory Development in the Social Sciences*.

[10] Gerring, J. 2007. *Case Study Research: Principles and Practices*. Cambridge Univ Pr.

[11] Hausenblas, M. and Karnstedt, M. 2010. Understanding Linked Open Data as a Web-Scale Database. *2010 Second International Conference on Advances in Databases, Knowledge, and Data Applications*. (2010), 56–61.

[12] Huijboom, N. and Broek, T. Van Den 2011. Open data: an international comparison of strategies. *European Journal of ePractice*. 12, March/April (2011), 1–13.

[13] Hyland, B. and Wood, D. 2011. The Joy of Data - A Cookbook for Publishing Linked Government Data on the Web. *Linking Government Data*. D. Wood, ed. Springer. 3–26.

[14] Janssen, K. 2011. The influence of the PSI directive on open government data: An overview of recent developments. *Government Information Quarterly*. 28, 4 (Oct. 2011), 446–456.

[15] Julien, N. 2012. *Business Opportunities Arising from Open Data Policies*. Imperial College London.

[16] Latour, B. 2005. *Reassembling the Social - An Introduction to Actor-Network-Theory*. Oxford University Press.

[17] Mahoney, J. and Goertz, G. 2006. A Tale of Two Cultures: Contrasting Quantitative and Qualitative Research. *Political Analysis*. 14, 3 (Jun. 2006), 227–249.

[18] Mayer-Schönberger, V. and Zappia, Z. 2011. Participation and Power: Intermediaries of Open Data. *1st Berlin Symposium on Internet and Society October 26th – 28th 2011* (2011).

[19] McClean, T. 2011. Not with a Bang but a Whimper. The Politics of Accountability and Open Data in the UK. *Draft prepared for the American Political Science Association Annual Meeting Seattle, Washington, 1-4 September 2011* (2011).

[20] McKinsey Global Institute 2013. *Open data: Unlocking innovation and performance with liquid information*.

[21] Miller, D. 1995. Consumption studies as the transformation of Anthropology. *Acknowledging Consumption. A review of new studies*. D. Miller, ed. Routledge. 264–295.

[22] Miller, D. and Slater, D. 2000. *The Internet: an ethnographic approach*. Berg.

[23] Moynihan, D.P. 1998. Secrecy. The American Experience. Yale University Press.

[24] Rourke, F.E. 1957. Secrecy in American Bureaucracy. *Political Science Quarterly*. 72, 4 (1957), 540–564.

[25] Scholl, H.J. et al. 2012. Process integration, information sharing, and system interoperation in government: A comparative case analysis. *Government Information Quarterly*. 29, 3 (Jul. 2012), 313–323.

[26] Shadbolt, N. and O'Hara, K. 2013. Linked data in government. *IEEE Internet Computing*. July/August 2013 (2013).

[27] Ubaldi, B. 2013. *Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives*. Technical Report #No. 22.

[28] Villazón-Terrazas, B. et al. 2011. Methodological guidelines for publishing government linked data. *Linking Government Data*. Springer. 27–49.

[29] Weber, M. 1946. *Essays in Sociology*. Oxford University Press.

[30] Zuiderwijk, A. et al. 2012. Socio-technical Impediments of Open Data. *Electronic Journal of e-Government*. 10, 2 (2012), 156–172.