

REVISITING OPEN DATA RESEARCH THROUGH THE LENS OF THE DATA VALUE CHAIN

Csaba Csáki¹ and Andrea Kő²

DOI: 10.24989/ocg.v331.17

Abstract

While the idea of open government data is not new, there appears to be a constant shift in leading research objectives guiding the field. This is because the reasons behind the increased research interest keep changing. The latest motivation stems from economic arguments, namely the reuse of public sector information) – which in turn creates a market for open-data based value added services. Although over the last decade many research topics have been identified and various research agendas have been proposed, most of them either focus on specific areas or are rooted in the popular approach of the time. Using the idea of the data value chain, this paper provides an integrated view of open government data research – which then allows a systematic and consistent identification of research topics and clarification of corresponding open questions in the area of open data quality. Research areas proposed are context/environment (policy and regulations), supply side (government organizations), consumer side (service providers and users), (societal or economic) impact, and technology (supporting the DSC, including the data). In addition – and what is regularly missing from most reviews – theory (providing definitions and a sound base) is considered as the sixth area.

1. Introduction

While the idea of open government and making public sector data available is not new [41], the idea keeps returning to the forefront of academic interest. While it is originated in the (typically constitutional) right of access to information and is often requested under the goal of transparency and accountability [22], over the last two decades it has gotten new fuel from the technology backed e-Government initiative [18; 28; 5]. The latest push to access even more public sector information (PSI) comes from commercial interest, namely the trend to apply open government data (OGD) in innovative value added services [10; 26; 55; 54]. “One of the key purposes of open data platforms is to promote access to government data and encourage development of creative tools and applications to engage and serve the wider community through the visualisation of patterns and relationships” [52, p. 287]. “Governments have started to share and open up their own data, yet the real value of open data often comes from integrating government data with non-government data sources” [46, p 1]. Given the increased interest in open data reuse, the quality of open data has become even more important [49]. There are many research frameworks and models [7] – including open data quality (DQ) measurement approaches [49; 53; 42] – but most of them do not provide avenues to investigate OGD quality in the context of integrated utilization (where data from several sources are combined to create services customers willing to pay for [6]). Furthermore, reuse is not the final goal, instead the real value is in the impact of those utilization efforts [9]. Reuse related open data quality is often addressed under the umbrella of linked open data (LOD), where one of the main concern is provenance of the data sets used [40]. Even though quality is often judged by knowing something’s origins and how it was produced, LOD and provenance only addresses some

¹ Corvinus University of Budapest, csaki.csaba@uni-corvinus.hu

² Corvinus University of Budapest, andrea.ko@uni-corvinus.hu

of the quality issues, such as difficulties with identifying the origin of data or inability to connect separate datasets [25]. One of the most advanced views of open data is built around the concept of ecosystems [19; 45]. This approach, however, provides a complex analytical model setting focusing on relationships and flows and less readily applicable to the concern of quality. Even specific frameworks dedicated to the issue of quality in the context of open data may only focus on providing dimensions and characteristics of quality or may offer mathematical formulas to provide measurement of quality parameters (upon which OD quality assessment may be based) [49] – but most of them fell short when it comes to offering guidelines what to do about quality lapses identified.

To address the root causes of OD quality issues and to identify when and how quality defects are introduced into datasets (eventually published as open data) a new view is required to allow for new research focus to be established and new research questions to be posed. To this extent this research proposes the application of the ‘data value chain’ metaphor [20] to establish a research framework within which deeper research questions may be asked in the open data quality area leading to practical considerations for both issuers and users of open data. This paper is organised as follows: after a review of key terms and various frameworks proposed in the open data quality field, the concepts of ‘data supply chain’ and ‘data value chain’ are discussed. This is followed by the application of the value chain model to open data – leading to new research topics and questions in relation to OD quality control. The paper is completed by reviewing the key recommendations and proposing further research directions.

2. Value creation and quality in Open Government Data

While the term ‘open data’ (OD) may cover a lot of different data from differing sources – including scientific and private datasets (Link) – open government data is “*non-privacy-restricted and non-confidential data, which is produced with public money and is made available without any restrictions on its usage and distribution*” ([23], p. 258). Originally, the idea of publishing public sector data (PSD) was the result of promoting accountability and transparency [22]. Later the e-Government idea [18], then the push for open government led to increased demand for Open Government Data (OGD). The latest trend is based on economic interest, namely the idea of innovative, commercial reuse of public sector information (PSI). However, with the advent of commercial reuse [4] – including integration with other, existing datasets (forming mashups) [4] – the focus becomes how value is created and what role quality plays in these processes [10]. While in technical terms data differs from information – the former being a term related to the storage and preservation of symbols (in itself having no meaning), while the latter referring to data interpreted by an actor in a given context [33] –, reusing open data typically means contextual matching, which is thus interpreted as information by the end user. From this point of view there appears to be little differentiation between data and information in the OD literature, especially so when it comes to the question of quality.

It is self-evident that low information quality (IQ) is one of the most difficult and pressing problems for consumers of information, especially given the explosion in the number of informational outlets. But ‘quality’ can be an elusive concept. There are different frameworks from which information (or data) quality issues can be assessed. For example, adopting a technical view mandates associating information quality with the accuracy of the information in products such as databases. This may be viewed as ‘data system quality’, looking at issues surrounding timeliness of update, system reliability, system accessibility, system usability and system security [14]. Another, the machine readability approach [13] is concerned with linking, finding, relating and reading data typically

using automated processes [42], and characteristics typically considered include number of formats, traceability, automated tracking, use of standards, trustworthiness, authenticity or provenance. Perhaps the most commonly used simple definition of user side IQ interprets the term as ‘fit-for-use’ [51]. However, IQ defined this way remains a relative construct whereby data considered appropriate for a given use may not display acceptable attributes in another setting [47]. Furthermore, fit-for-use does not immediately allow for ready measurability and it requires additional detail in order to be operationalized [16]. Moreover, the literature’s appreciation of specific characteristics of information quality reveal that the number, definition, and measurability of recommended features or dimensions varies widely. This motivates [44] to state (p. 2): *“Generally speaking, data quality can be related to a set of “dimensions” that are usually defined as quality properties or characteristics”*.

Potential IQ criteria may be classified based on whether they are related to the user, to the information itself or to the manipulation of the underlying data [36]. Even though the use and application of quality frameworks expand, they remain focused on the underlying technical and data aspects of IQ. Ultimately, however, it is the user who must decide between qualitatively good and poor information and whether there is an acceptable level of quality required to achieve certain goals and if data is usable to generate value. Indeed, [29] distinguishes accuracy, authority, currency and novelty as quality dimensions. In a similar perspective, [37] differentiated information quality based on accessibility, actual value, completeness, credibility, flexibility, form, meaning over time, relevance, reliability, selectivity and validity. Considering a user-centric perspective of the Internet, information quality would identify the degree to which information is suitable for doing a specific task by a specific user in a certain context [12]. This should hold for OGD as well, yet it is also easy to imagine that given the context, quality expectations for open data might dramatically diverge from data and informational quality issues associated with private data and there may be additional considerations that are special to open data. It is safe to assert that users of open data seek information which may or may not be readily available in the published data set. Indeed, that is the basis for value added services.

One open data quality (ODQ) approach considers technical abilities and concerns raised deviate little from those associated with general DQ investigations. They are concerned about the processes and outcomes of producing and managing datasets as well as about corresponding technical standards [48; 16]. Another stream is centred on the availability and accessibility of various types of data. As an example, the Open Data Barometer [8] raises awareness about the gap between data haves and have-nots on several different availability measures of open datasets around the world. Other related concerns cover whether intended audiences are aware of the availability of relevant datasets and even if they are, whether data is easy to find. Yet another set of frameworks is concerned about specific sectors and take into account the content of the datasets. Finally, it is customary to ask about the value of open data, which, in general terms considers the needs of end users [16]. The current disposition of ODQ characteristics is aptly demonstrated by the work of [53] who, in pursuit of the measurability of ODQ, define and operationalize 68 metrics along 6 dimensions. The Linked Open Data (LOD) ‘movement’ concentrates on provenance that may enrich the context of open data [40]. Key principles concern the traceability and informational links about the source, the structure of data provenance, linkages to individual elements and linking provenance records. Therefore, dimensions such as origin, attribution, traceability, accessibility and presentation can provide evidence for supporting the assessment of quality, including reliability and trustworthiness. In addition, LOD discussions often centre on potential privacy issues.

However, the core of the literature on IQ/DQ focuses on assessment only. There is a dearth of literature actually providing open data quality control guidelines or even offering best practices. For a data provider organization to be able to provide quality control [49] it would be necessary to know where and why quality issues occur or are introduced. Even quality guidelines fall short of an answer as they only focus on the organizational process and tend to forget about important factors of the public sector setting. Limits of these views with respect to reuse requires a more sophisticated understanding of the relationship between the processes producing OGD and the quality requirements of the various users creating and utilizing value added services. Even though the proposed frameworks and models explaining various aspects of the OGD phenomena regularly refer to “supply” and “demand” most of them do not provide a definition or at least a consistent view of these sides and their related relevant processes potentially introducing errors into the datasets as published and used. Works analysing roles and offering some recommendations tend to focus on after-the-fact quality control by checking datasets prior to release and only for technical issues [49, 21]. This paper considers *the data value chain model* to identify areas of research where insights about quality root causes and corresponding remedies may be uncovered.

3. The concept of a “data value chain”

Viewing data as a result of production-like activities originated in the 1990s. It was [2] who first used the notion of ‘information manufacturing’. Then [50] advanced a data quality framework based on the similarities between the two manufacturing processes: in the ‘data supply chain’ (DSC) “*an information system can be viewed as a data manufacturing system acting on raw data input*” (p. 623). Supply chain is a system of organizations, people, processes, technology, information and resources in moving a product or service from the supplier to the customer (consumer). Data management often discussed as production, storage and application of data in creating value for some end user. Indeed, [50] had already proposed this to be extended into the concept of the ‘data value chain’ (DVC). “*Use of the term “data product” emphasizes the fact that the data output has value that is transferred to customers, be they internal or external to the organization*” (p. 624). This metaphor bodes well with the idea of OGD reuse, as the final goal is to generate some (societal or economic) value. Later [38] pointed out the importance of a formal ‘quality assurance’ in the DSC and warned about the importance preventing errors happening.

The data supply chain has resurfaced again when [17] investigated transparency and reliability of linked open data using the DSC. He raised important questions such as who is responsible for an error, or whether the information comes from a curated source. In essence he raised issues of provenance through the supply chain metaphor. Here the supply chain begins when data is created, then it is imported or combined with other data (creating new data), then data moves through the supply chain often being further transformed [32] (and eventually reused). [35] even extended the DVC approach to discuss issues of Big Data. Although the focus is on organizational decision making, they investigate how to bring disparate data together, which is relevant for the mashup-like expectations in OGD reuse. Their model consists of three key data processes with various steps in each: discovery (collect and annotate, prepare, organize), integration (integrate), and exploitation (analyse, visualize, make decisions). Indeed, the Big Data industry relies on constructing data supply chains [32], where exchange and integration of data across different platforms is at the heart of creating value. One issue with the above view of a data (supply and) value chain (summarized in Figure 1) is its ‘linear’ nature. When discussing the use of raw data, authors often use the ‘data lifecycle’ metaphor [39] instead, which focuses the attention on the fact that producers of data are also consumers from a different perspective. In summary, while the supply chain and lifecycle approaches focus on ‘producers’ and ‘consumers’ (or ‘supply’ and ‘demand’), the value chain

approach considers the context and the processes of transforming components of data into valuable information to be used for some end.



Figure 1: Data Value Chain (based on [20])

As for the supply or producer side, open data is usually (and originally) published in static platforms such as government web portals [43]. Reuse introduces several roles on the ‘demand’ or consumer side: access providers, cleaners, integrators, service developers and so on, all contributing to increase the value of data and provide benefits – resulting in, hopefully, social and economic impact. While [1] considered the possibility of discussing open data as a value chain, his conference poster presentation has not been explored further. Consequently, this research now turns to the application of the value chain model to the open government data research field.

4. Methodological considerations

To address the issue of preventative data quality guidelines in the context of open data this research applied a systematic literature review followed by theoretical arguments through the application of the data value chain model to the public sector. The research first looked at what relevant research areas had been discussed in the literature, followed by a short investigation considering the special characteristics of the public sector pertaining to opening up their data. Then the data value chain approach is applied to distil research topics with OD quality related questions. For the first part an extensive literature review is utilized [30] searching on specific phrases appearing in either the title, the abstract or the keywords section of papers, namely “open data” and either of the following: “taxonomy”, “research areas”, “research agenda”. Instead of focusing on a basket of journals or a set of leading conferences as is typically done [34], research is based on Google Scholar with the two authors separately sorting out the results leading to 9 papers with meaningful OGD research areas, taxonomies or agendas. Papers that fit the search criteria, but were too narrowly focused on one area (such as innovation only), or covered only specific domains (such as health) were excluded. Based on the research topics such identified, further systematic analysis was executed to see how they relate to the concepts of the data value chain and what meaningful research directions may emerge. The intent was to identify research areas where insights into the root cause of OGD quality issues may be uncovered.

5. Applying the data value chain approach to open government data research

5.1. Open data research: areas and agendas in the literature

Over the last fifteen years or so there appeared regular efforts to review the progress of the field and provide an organization of the various areas within it and propose research agendas. One of the firsts to address the issues of opening up data for reuse was [3], who – while studying these questions on behalf of OECD – proposed 5 domains of a data access regime: technological,

institutional and managerial, financial and budgetary, legal and policy, and cultural and behavioural. While discussing the concept of open government, [19] put forward four domains of OG: policies and practices, users, technology innovation, and context (considering the legal, policy and economic environment), then within this, concluded 6 themes related to open data, namely the workflow of defining data of interest, prioritizing data collection, conducting data collection, publishing the data, using data, and generating value. [9] were interested in the impact of open data and considered 4 'fronts' of scientific interest: history of OD, readiness assessments, implementation studies, and impact studies. [33] focused on research of OD services and their agenda included 7 categories of challenges: Information, Technologies, Processes and Activities, Products and Services, Participants, Customers, and Environment. Reviewing the state-of-the-art of open data related innovation, [55] argue that 7 perspectives should be investigated: legislative, political, social, economical, institutional, operational, and technical - and after reviewing their current state puts forward three research directions: theory and development; policies, use, and innovation; and infrastructures and technologies. [7] has reviewed several research programs (four from those mentioned above) and constructed 35 research topics into four major research areas management and policies, infrastructures, interoperability and usage and value. According to [45] there are 3 points of focus in OGD research: policy and practices, data management, and stakeholder engagement. One of the latest OGD research agendas is [27] with again a focus on innovation and with 3 proposed directions: conducting domain-specific studies, examining the use of tools, and expanding the existing set of research methods and theoretical foundations. [46] constructed an OGD taxonomy from the point of view of cross-sector partnerships consisting of 2 categories (with 6+8 dimensions): data sharing (type, content, admin level, provider diversity, facilitation and access degree of data) and data use (target, selection, policy problems, incentives, continuity, outcome, collaboration, and purpose).

5.2. Special considerations of the public sector in the context of the data value chain approach

There are a few major issues here compared to the traditional data value chain (and lifecycle) models. Data in the public sector has an original role, i.e. it is collected for a specific administrative purpose [31]. Consequently, the way data is collected, stored and manipulated depends on that purpose, which in turn requires transformation to make it reusable – in an unknown context and for goals not related to the original use. Both the original use as well as the process to open up the data is strictly regulated by policies and laws. The producer side thus operates in a strictly regulated context – regarding both data management (collection and storage) and allowing for reuse itself. Freedom of information laws regulated information availability (what must or should not be published, what may be accessible and who can access it); and there are related, but often separate set of rules to govern the issue of reuse (how data such published may be used, what is free and what requires fees to be paid). The producer is strongly separated from the consumer [11]. This is not unlike the inter-organizational information ecology model discussed by [15], where there is a clear split between producers and suppliers when sharing information resulting in disjoint processes. The argument thus can be made that in the context of OGD there are two (asymmetric) data lifecycles: one on the producer side (public entities) and one in the consumer side (commercial re-use of open data in value added services) [11; 39]. This setting is further emphasized by the infomediaries put forward in the ecosystem approach [24], who clean, sort, interpret, reformat, link, and improve the data. Since a “*key quality principle is that customers define quality*” ([21], p. 560), the above specialties result in new challenges of quality going beyond the traditional views of data quality. Consequently, OGD requires improvement both in usability and technical qualities (with the latter including IT-DB qualities of the data as well as provenance and linkages to other data sets). Therefore, to create more value and to improve the quality of the end product (i.e. the services

based on integrated data sets) a close partnership between the supply and demand of data is necessary [46]. To close the loop special feedback channels need to be created [19]. Any research addressing issues in the field of OGD quality should be able to deal with this context.

5.3. Key research areas based on the data value chain approach

Considering the arguments above, the following research areas proposed based on the application of the DVC model to OGD quality: 1) context/environment (policy and regulations – legal background and governance), 2) supply side (government organizations), 3) service providers (on the consumer side), 4) end users and societal or economic impact (also on the consumer side), and 5) enabling technology (supporting the DVC, including the data). In addition – and what is regularly missing from most reviews – theory (providing definitions and a sound base for research) is considered as the sixth area (see Figure 2).

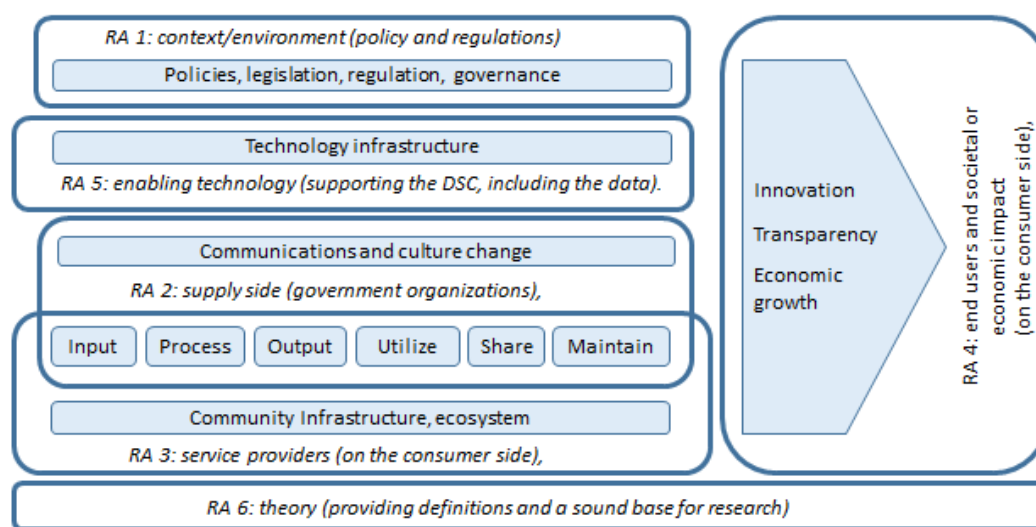


Figure 2: Open Government Data Quality Research Areas (RA) in the Data Value Chain

Context/environment – Legal background and governance: Contextual elements include policy making and legislative actions forming the legal background as well as the processes implementing the resulting regulatory setup. Policy clarifies strategic directions under which regulations should be drawn up and institutions should be organized, while the legal framework creates the context for the rest of the perspective (through publishing and reuse rules), however, a governance element is required for the operationalization of the relevant policy and legal expectations. Research question thus should focus on 1) Are information quality issues considered during the legislative process (what is the level of awareness)? 2) Do resulting regulations reflect quality expectations of end users and if yes, how? 3) How regulatory requirements may introduce restrictions on quality control capabilities?

Supply side – Governmental organizations: Agencies at all levels of governing are involved in implementing OGD policies, or following data publication rules. Research questions of the organizational aspect consider decisions, management processes, and roles: 1) What governance processes may support quality control of OD? 2) What type of errors are associated with the original data management and use process? 3) At which points of the data management process such errors occur? 4) What other errors may be introduced during the generation (publication) of open data?

Consumer side – Service providers: Once open data is ‘out there’, it is difficult for the originator to control what happens to it. One typical way of dealing with such issues is to restrict the actual use or even charge money – which could defeat the purpose. Therefore, questions of interest here are: 1) What are the quality issues committed by the value added service providers? 2) What quality control mechanisms could re-user install? 3) How to provide feedback: what channels exists to indicate errors or clarify questions?

Consumer side – End users and (societal or economic) impact: According to the re-use model, end users are often even further removed from the ‘source’. What is most important here is similar to the service creator role: 1) How to provide feedback: what channels exists to indicate errors, issues or clarify questions?

Technological enablers: Technology plays a crucial role, obviously, in managing the open data value chain. Public entities manage data portals but often provide a platform to engage with their data. There are existing and emerging standards supporting various aspects of the DVC. Data in itself has not been considered as a separate area of interest here, because while data is important, it is not the source of any potential quality issues. Questions: 1) What is the impact of various architecture on information quality? 2) How technology and standards can be used to restrict the likelihood of introducing errors during the data management process (e.g. digitization and various transformations)?

Theoretical foundations: While there are numerous frameworks dealing with OD quality, their main shortcoming is the lack of quality control guidelines. Theoretical treaties of OD quality should provide definitions and a sound base for research – also including critical views as well as historical overviews, reviews of trends and research agendas. Such frameworks should go beyond enumerating quality dimensions and offering measurement addressing research questions related to quality control recommendations: 1) Is there special typology of OGD quality errors and issues? 2) How to overcome the limitations of existing frameworks regarding DQ control? 3) What are the preventative actions offered or discussed in existing case studies? 4) How open data quality maturity models may be used to improve pre-emptive quality control mechanisms?

6. Conclusions and Future Research

Open data quality related research got remarkable interest in the literature. Several papers investigated OD from many aspects, but in the majority of the cases the holistic approach of the discussion is missing, as we detailed in our literature review. We analysed various frameworks proposed in the field of open data quality and highlighted research gaps. Detailed study of these frameworks led us to select data value chain, as a framework for our discussion. The paper has provided an integrated view of OGD research using the data value chain approach – which then has led to a systematic and consistent identification of research topics. We identified six main research areas: 1) environment; 2) government organizations; 3) service providers; 4) end users and impact; 5) technology; and 6) theory. The main advantage of the approach provided here is that it allows for the clarification of open questions corresponding to the given areas – with special focus on the quality of open government data that can be used in value added services. Using the questions posed above, the next step in this research is to focus on the root causes of OGD quality issues and to create a model allowing for the analysis of those causes with the intent to find remedies, such as best practices in quality issue prevention. One model that could also be investigated in relation to the value creation from open data is the ecosystem approach to see if that offers new avenues to insights. The ecosystem model deals with roles and their dynamic relationships in creating value in

ecologies. It would be then possible to investigate the root causes of ODQ issues in relation to those roles and their activities including the nature of communication among them.

7. Acknowledgement

This work was created in commission of the National University of Public Service under the priority project KÖFOP-2.1.2-VEKOP-15-2016-00001 titled „Public Service Development Establishing Good Governance” and of the Corvinus University of Budapest.

8. References

- [1] ALBANO, C.S., Open government data: A value chain model proposal, Poster presented at the 14th Annual International Conference on Digital Government Research, Quebec City, Canada, June 17–20, 2013.
- [2] ARNOLD, S.E., Information manufacturing: the road to database quality, in: *Database* 15(5), 32-39, 1992.
- [3] ARZBERGER, P., Schroeder, P., Beaulieu, A., Bowker, G., Casey, K., Laaksonen, L., and Wouters, P., An international framework to promote access to data, in: *Science*, 303(5665), 1777–1778, 2004.
- [4] ATTARD, J., Orlandi, F., Scerri, S., Auer, S., A systematic review of open government data initiatives, in: *Government Information Quarterly*, 32(4), 399-418, 2015.
- [5] BERTOT, J.C., Gorham, U., Jaeger, P.T., Sarin, L.C., and Choi, H., Big data, open government and e-government: Issues, policies and recommendations, in: *Information Polity*, 19(1, 2), 5-16, 2014.
- [6] CHAN, C.M., From open data to open innovation strategies: Creating e-services using open government data, in: *Procs. of the 46th Hawaii International Conference on System Sciences*, 1890-1899, 2013.
- [7] CHARALABIDIS, Y., Alexopoulos, C., and Loukis, E., A taxonomy of open government data research areas and topics, in: *Journal of Organizational Computing and Electronic Commerce*, 26(1-2), 41-63, 2016.
- [8] DAVIES, T., Open Data Barometer. 2013 Global Report, online at: <http://www.cococonnect.org/publication/open-data-barometer-2013-global-report>, last downloaded January 21, 2017.
- [9] DAVIES, T. and Perini, F., Researching the emerging impacts of open data: revisiting the ODDC conceptual framework, in: *The Journal of Community Informatics*, 12(2), 2016.
- [10] DAWES, S.S. and Helbig, N., Information Strategies for Open Government: Challenges and Prospects for Deriving Public Value from Government Transparency, in: Wimmer, Chappellet, Janssen, and Scholl (Eds) 9th IFIP 8.5 Conference on Electronic Government, Springer LNCS-6228, 50-60, 2010.

-
- [11] DE KEYZER, M., Loutas, N., and Goedertier, S., The Linked Open Government Data & Metadata Lifecycle, https://joinup.ec.europa.eu/sites/default/files/d2.1.2_training_module_2.1_the_linked_open_government_data_lifecycle_v1.00_en.pdf, last accessed November 11, 2017.
- [12] EMAMJOME, F.F., Rabaa'i, A.A., Gable, G.G. and Bandara, W., Information quality in social media: a conceptual model, in: Proceedings of the Pacific Asia Conference on Information Systems, Seoul, 2013.
- [13] ERICKSON, J.S., Viswanathan, A., Shinavier, J., Shi, Y. and Hendler, J.A., Open Government Data: A Data Analytics Approach, in: IEEE Intelligent Systems, 28(5), 19-23, 2013.
- [14] FOX, C., Levitin, A., and Redman, T. C., Data and data quality: Total Data Quality Management Research Program, Sloan School of Management, MIT, Boston, 1995.
- [15] FEDOROWICZ, J., Gogan, J.L., and Ray, A.W., The ecology of interorganizational information sharing, in: Journal of International Information Management, 13(2), 1, 2004.
- [16] FRANK, M. and Walker, J., User centred methods for measuring the quality of open data, in: The Journal of Community Informatics, 12(2), 47-68, 2016.
- [17] GROTH, P., Transparency and reliability in the data supply chain, in: IEEE Internet Comp., 17(2), 69-71, 2013.
- [18] GRÖNLUND, Å. and Horan, T. A., Introducing e-gov: history, definitions, and issues, in: Communications of the Association for Information Systems, 15(1), 713-729, 2004.
- [19] HARRISON, T. M., Pardo, T. A., and Cook, M., Creating open government ecosystems: A research and development agenda, in: Future Internet, 4(4), 900-928, 2012.
- [20] HUGHES, J., An Open Data Value Chain: Making Data Flowers Bloom, London Data Store 31, 2011, online at <https://www.slideshare.net/janet-hughes/open-data-value-chain>, last downloaded December 15, 2017.
- [21] HUH, Y.U., Keller, F.R., Redman, T.C., Watkins, A.R., Data quality, in: Inf. & SW Tech., 32(8), 559-565, 1990.
- [22] JANSSEN, K., The influence of the PSI directive on open government data: An overview of recent developments, in: Government Information Quarterly, 28 (4), 446-456, 2011.
- [23] JANSSEN, M., Charalabidis, Y. and Zuiderwijk, A., Benefits, adoption barriers and myths of open data and open government, in: Information Systems Management, 29(4), 258-268, 2012.
- [24] JANSSEN, M., and Zuiderwijk, A., Infomediary business models for connecting open data providers and users, in: Social Science Computer Review, 32(5), 694-711, 2014.

-
- [25] JETZEK, T., Managing complexity across multiple dimensions of liquid open data: The case of the Danish basic data program, in: *Government Information Quarterly*, 33(1), 89-104, 2016.
- [26] JETZEK, T., Avital, M., and Bjorn-Andersen. N., Data-driven innovation through open government data, in: *Journal of theoretical and applied electronic commerce research*, 9(2), 100-120, 2014.
- [27] KANKANHALLI, A., Zuiderwijk, A., and Tayi, G. K., Open innovation in the public sector: A research agenda, in: *Government Information Quarterly*, 1(34), 84-89, 2017.
- [28] KASSEN, M., Globalization of E-government: Open Government as a Global Agenda, Benefits, Limitations and Ways Forward, in: *Information Development*, 30(1), 51-58, 2013.
- [29] KLOBAS, J.E., Beyond information quality: fitness for purpose and electronic information resource use, in: *Journal of Information Science*, 21(2), 95-114, 1995.
- [30] KRIPPENDORFF, K. H., *Content Analysis - An Introduction to Its Methodology*, Sage Publications, 2013.
- [31] KUK, G., and Davies, T., The roles of agency and artifacts in assembling open data complementarities, in: *32nd Int. Conf. on Inf. Systems*, 2011, online at: <http://eprints.soton.ac.uk/273064/>, last downloaded Dec. 15, 2107.
- [32] LI, P., Wu, T.Y., Li, X.M., Constructing data supply chain based on layered PROV, in: *The Journal of Supercomputing*, 73(4), 1509–1531, 2017.
- [33] LINDMAN J., Rossi, M., and Tuunainen, V., Open Data Services: Research Agenda, in: *Proceedings of HICSS-46*, 1239-1246, 2013.
- [34] LINK, G.J.P., Lumbard, K., Conboy, K., Feldman, M., Feller, J., George, J., Germonprez, M., Goggins, S., Jeske, D., Kiely, G., Schuster, K., and Willis, M., Contemporary Issues of Open Data in Information Systems Research: Considerations and Recommendations, *Communications of the AIS*, Vol. 41 , Article 25, 2016.
- [35] MILLER, H.G. and Mork, P., From data to decisions: a value chain for big data, in: *IT Profes.*, 15(1), 57-59, 2013.
- [36] NAUMANN, F. and Rolker, C., Assessment methods for information quality criteria. in: *Proceedings of the 5th International Conference on Information Quality*, Humboldt-Universität zu Berlin, 148-162, 2000.
- [37] OLAISEN, J., Information quality factors and the cognitive authority of electronic information, in: Wormell I., *Information quality: Definitions and dimensions*, Taylor Graham, London, 91-121, 1990.
- [38] OLSON, J. E., *Data quality: the accuracy dimension*, Morgan Kaufmann, 2003.

-
- [39] PARSONS, M.A., Godøy, Ø., LeDrew, E., De Bruin, T.F., Danis, B., Tomlinson, S., and Carlson, D., A conceptual framework for managing very diverse data for complex, interdisciplinary science., in: *Journal of Information Science*, 37(6), 555-569, 2011.
- [40] PIGNOTTI, E., Corsar, D. and Edwards, P., Provenance Principles for Open Data, in: *Proceedings of DE2011*.
- [41] PARKS, W., The open government principle: applying the right to know under the constitution, in: *The George Washington Law Review*, 26(1), 1-22, 1957.
- [42] RULA, A. and Zaveri, A., Methodology for assessment of linked data quality, in: *Proceedings of the 1st Workshop on Linked Data Quality at the 10th International Conference on Semantic Systems*, 2014.
- [43] SÁEZ MARTÍN, A., Rosario, A.H.D., and Pérez, M.D.C.C., An international analysis of the quality of open government data portals, in: *Social science computer review*, 34(3), 298-311, 2016.
- [44] SCANNAPIECO, M. and Catarci, T., Data quality under a computer science perspective, in: *Archivi & Computer*, 2, 1-15, 2002.
- [45] STYRIN, E., Luna-Reyes, L.F., and Harrison, T.M., Open data ecosystems: an international comparison, in: *Transforming Government: People, Process and Policy*, 11(1), 132-156, 2017.
- [46] SUSHA, I., Janssen, M. and Verhulst, S., Data collaboratives as a new frontier of cross-sector partnerships in the age of open data: taxonomy development, in: *Proc.s of the 50th Hawaii Int. Conf. on System Sciences*. 2017.
- [47] TAYI, G.K. and Ballou, D.P., Examining data quality. *Communications of the ACM*, 41(2), 54-57, 1998.
- [48] UBALDI, B., Open Government Data: Towards Empirical Analysis of Open Government Data Initiatives, OECD Working Papers, No. 22, 2013, online at <http://dx.doi.org/10.1787/5k46bj4f03s7-en>, downloaded Dec. 10, 2017.
- [49] VETRO, A., Canova, L., Torchiano, M., Minotas, C.O., Iemma, R., and Morando, F., Open data quality measurement framework: Definition and application to open gov. data, in: *Gov. Inf. Quart.*, 33(2), 325-337, 2016.
- [50] WANG, R.Y., Storey, V.C., and Firth, C.P., A framework for analysis of data quality research, in: *IEEE transactions on knowledge and data engineering*, 7(4), 623-640, 1995.
- [51] WANG, R.Y., and Strong, D.M., Beyond accuracy: What data quality means to data consumers, in: *Journal of Management Information Systems*, 12(4), 5-33, 1996.
- [52] WEERAKKODY, V., Irani, Z., Kapoor, K., Sivarajah, U., and Dwivedi, Y.K., Open data and its usability: an empirical view from the Citizen's perspective, in: *Information Systems Frontiers*, 19 (2), 285-300, 2017.

-
- [53] ZAVERI, A., Rula, A., Maurino, A., Pietrobon, R., Lehmann, J. and Auer, S., Quality assessment methodologies for linked open data, in: *Semantic Web Journal*, 1(5), 1-31, 2012.
- [54] ZELETI, F.A., Ojo, A., and Curry, E., Exploring the economic value of open government data, in: *Government Information Quarterly*, 33(3), 535-551, 2016.
- [55] ZUIDERWIJK, A., Helbig, B., Gil-García, J.R., and Janssen, M., A Review of the State-of-the-Art and an Emerging Research Agenda, in: *Journal of Theoretical and Applied Electronic Commerce*, 9 (2), 1-8. 2014.