Experience: Open Fiscal Datasets, Common Issues, and Recommendations

FATHONI A. MUSYAFFA, University of Bonn, Germany

CHRISTIANE ENGELS, University of Bonn, Germany

MARIA-ESTHER VIDAL, German National Library of Science and Technology, Germany & Universidad Simón Bolívar, Venezuela

FABRIZIO ORLANDI, University of Bonn & Fraunhofer IAIS, Germany

SÖREN AUER, TIB Leibniz Information Center Science and Technology & L3S Research Center, University of Hannover, Germany

Public administrations are continuously publishing Open data, increasing the amount of government open data over time. The published data includes budgets and spending as part of fiscal data; publishing these data is an important part of transparent and accountable governance. However, open fiscal data should also meet open data publication guidelines. When requirements in data guidelines are not met, effective data analysis over published datasets cannot be performed effectively. In this paper, we present Open Fiscal Data Publication (OFDP), a framework to assess the quality of Open fiscal datasets. We also present an extensive Open fiscal data assessment and common data quality issues found; additionally, Open fiscal data publishing guidelines are presented. We studied and surveyed main quality factors for Open fiscal datasets. Moreover, the collected quality factors have been scored according to the results of a questionnaire to score quality factors within the OFDP assessment framework. We gather and comprehensively analyze a representative set of 77 fiscal datasets from several public administrations across different regions at different levels (e.g., supranational, national, municipality). We characterize quality issues commonly arising in these datasets. Our assessment shows that there are many quality factors in fiscal data publication that still need to be taken care of so that the data can be analyzed effectively. Our proposed guidelines allow for publishing Open fiscal data where these quality issues are avoided.

CCS Concepts: • **Information systems** → *Data exchange*; • **Social and professional topics** → *Governmental regulations*;

Additional Key Words and Phrases: open fiscal data, open data assessment

1 INTRODUCTION

Many public administrations have started publishing Open data. Open Knowledge International (OKI) indexed 122 countries in Global Open Data Index (GODI) [OKI 2015d] that have published Open data in various domains. Open Government Partnership (OGP)[1] establishes an Open data working group to develop Open data plans and actions across

---

[1] http://www.opengovpartnership.org/

Authors' addresses: Fathoni A. Musyaffa, University of Bonn, Regina-Pacis-Weg 3, Bonn, NW, 53113, Germany, musyaffa@cs.uni-bonn.de; Christiane Engels, University of Bonn, Regina-Pacis-Weg 3, Bonn, NW, 53113, Germany, engelsc@cs.uni-bonn.de; Maria-Esther Vidal, German National Library of Science and Technology, Germany & Universidad Simón Bolívar, Venezuela, Welfengarten 1B, Hannover, NI, 30167, maria.vidal@tib.eu; Fabrizio Orlandi, University of Bonn & Fraunhofer IAIS, Fraunhofer IAIS, Schloss, Sankt Augustin, NW, 53757, Germany, orlandi@cs.uni-bonn.de; Sören Auer, TIB Leibniz Information Center Science and Technology & L3S Research Center, University of Hannover, Welfengarten 1B, Hannover, NI, 30167, Germany, auer@l3s.de.

71 OGP member countries. Beyond the country level, publication of Open data is also done on supra-national levels, such as UN[2] and World Bank[3].

One of the high-value domains in Open data is public finance [UK Cabinet Office 2013] or fiscal data. The term *fiscal* refers to any activities related to government expenditures, revenues and debt [American Heritage Dictionaries Editors 2016]. Open fiscal datasets include, but are not limited to, budget, spending, contract, and procurement data. In this paper, our main focus is on budget and spending datasets, which are the most frequent types of released Open fiscal data. Budget datasets determine the various income and expenditure allocations within a certain period. Spending datasets provide details regarding the amount of money paid for specific items.

The open publishing of such datasets has a number of motivating benefits, such as: increasing transparency and compliance [Shadbolt et al. 2012; Tygel et al. 2016], preventing corruption [Graft et al. 2016], raising democratic control and participation in politics [Huijboom and Van den Broek 2011; Tygel et al. 2016], encouraging innovation in services and products [Huijboom and Van den Broek 2011; Shadbolt et al. 2012; The World Bank 2015], performing a comparative analysis [Tygel et al. 2016], enhancing law enforcement [Huijboom and Van den Broek 2011], adding business value [Tygel et al. 2016], improving efficiency and effectiveness [Tygel et al. 2016], as well as generally reducing the barrier between government and citizens [Huijboom and Van den Broek 2011; Tygel et al. 2016]. Governance *transparency* is concerned with the capability of finding information about what happened in the government [Piotrowski and Van Ryzin 2007]. *Accountability* exists when tasks done by a particular individual or an organization can be requested, overseen, and directed by others [Stapenhurst and O'Brien 2008]. *Transparent public administration* improves public trust which engages more political participation from their citizens. Open data implementation in Brazil has successfully uncovered corruption scandals, as reported by [Graft et al. 2016]. A summary of Open data values and impacts (projected market value, number of Open data jobs created, economic benefit, etc.) is provided by [Srinivasan 2016].

An analysis can be done on different fiscal datasets that have similar properties. For example, comparing budget allocations from different municipalities with a similar population, area size and/or GDP. This analysis requires the datasets to be consistent and to contain common classifications, which make the datasets comparable. When such detailed fiscal practices can be publicly scrutinized, the chance of public officials conducting fiscal malpractice is lower, as illustrated in the report by [Graft et al. 2016]. Open fiscal data allows for measuring effectiveness and efficiency of an executed particular funding program, which can be assessed by the outcome of the funding.

Many Open datasets have quality issues; an analysis by Computer Weekly on UK's Cabinet Office open spending data showed that the released data have inconsistent computer encoding and therefore an advanced programming skill is required to scrutinize the data systematically [Ballard 2014]. In addition, Open datasets generally have not been designed to be interoperable [Hendler 2014], for example, datasets are provided in various formats, structure, classification schemes and languages which prevent the datasets from being effectively integrated and analyzed [Hendler 2014]. Open datasets should follow data publishing guidelines to ensure the quality of the datasets. If the guidelines are not followed, effective data analysis cannot be performed. To see the effect of data quality problem, a fiscal data assessment framework is substantial which in turn act as a guideline to improve the quality of Open fiscal datasets.

There are three main contributions in this paper: *first*, we propose a comprehensive assessment framework for Open fiscal data; *second*, we assess current Open fiscal data and present a number of quality issues that were found; *third*, we provide guidelines for publishing Open fiscal data based on the assessment.

---

[2]http://data.un.org//
[3]http://data.worldbank.org/

2 RELATED WORK

The challenge of data interoperability across different dataset publishers is not particularly new. From the company and business perspective, XBRL (eXtensible Business Reporting Language) format has been used as a standard for business information exchange. XBRL allows the representation of standardized accounting processes. Authorities play an important role in the XBRL adoption, as it became mandatory in 2009 for top 500 U.S. companies to report in XBRL [Ahrendt 2009]. Within the Open fiscal data domain, such matured standard for datasets publishing is not developed due to lack of a binding order from authorizing bodies, since the standards of Open data publication normally come from the grassroots communities and NGOs. Moreover, differences in Open fiscal data accounting processes and classification hierarchies across different public administrations complicate datasets standardization. Such differences limit the usefulness of financial disclosures [Lucas 2012, 2014].

There are few works that assess Open datasets based on common quality factors that should be present in Open data. The Open Data Monitor project[4] reports Open data implementation across Europe. An assessment in the general domain was reported by GODI [OKI 2015a] and ODB [Davies 2013]. GODI provides country rankings based on nine GODI factors and the availability of 13 different Open data domains in each country, including budget and spending. ODB provides Open data analysis and ranking based on Open data initiative readiness, program implementation, and impact on business, civil society, and politics. Peters et al. [Peters et al. 2017] asses Open fiscal data and portal quality specifically for ESIF funding in EU countries. Currently, several Open data publishing guidelines for the general Open data domain exist, including [OKI 2015b], [Tauberer 2012], and [Wonderlich 2010]. The Open Data Handbook [Dietrich et al. 2009] provides a guide for the legal, social, and technical aspect of Open data. The 5-star data schema [Berners-Lee 2010] is well-known among Linked Open Data communities. It is also worth mentioning a survey by [Zaveri et al. 2015] for Linked Data quality assessment. A specific guide to publishing Open data as Linked Data is provided by [Bauer and Kaltenböck 2011]. [Shadbolt et al. 2012] mention lessons learned from data.gov.uk implementation. Data Management Maturity (DMM) Model provides *Capability and Maturity Levels* [CMMI Institute 2014], which has six different data management process areas: data management strategy, data governance, data quality, platform and architecture, data operations, and supporting processes. Data quality process area is directly affecting data management strategy, data governance, platform and architecture, and data operations, which is why data quality process area is important. Data quality process area is composed further of data quality strategy, data profiling, data quality assessment and data cleansing. The OFDP Framework provides a more elaborate framework of *data quality* for fiscal data within the sub-process area of data quality strategy, data profiling, and data assessment.

Fung et al. [Fung et al. 2007] mention that a sustainable transparency system improves on three important dimensions over time: expanding information scope, increasing information accuracy and quality, and increasing use of information. Compared to other works in data quality assessment, we propose a framework, OFDP, which aims to improve these dimensions within fiscal data. Our assessment of surveyed fiscal datasets exemplify the heterogeneity issues as mentioned by [Lucas 2014]. In comparison with related work regarding datasets assessment, we aim to assess fiscal datasets on multiple public administration levels and provide guidelines accordingly. We believe that the dimensions mentioned by [Fung et al. 2007] will be improved if the dataset publishers comply with our guidelines, as can be seen in Section 6.

---

[4]http://project.opendatamonitor.eu/

Fig. 1. Methodology for obtaining the proposed OFDP framework and guidelines.

3 METHODOLOGY

Our methodology to analyze Open fiscal data using the OFDP Framework is summarized in Figure 1. We gathered links to the fiscal datasets from the OpenSpending community[5], which is an active community which aims to track and analyze global public financial information. The community members submitted extensive links using Github[6]. These links were then used to obtain the actual datasets for our assessment. In addition, we also explored and added additional links outside of the submitted links. As for the main framework, we studied the literature to acquire common motivations for publishing Open data. Later, we gathered quality factors that support these motivations and then measured the weights of each quality factor. To achieve a more objective weighing of the factors, we collected fiscal communities' views through a questionnaire[7], which was distributed in several fiscal communities (OpenSpending, Follow the Money, OpenBudgets.eu, IODC 2016) and government officials. We collected 24 responses from this questionnaire and used the median as the weight for our identified quality factors. The collected and assessed data are then ranked using three methods: OFDP, ODB, and GODI. We evaluated the ranking results using Spearman's coefficient. Finally, we highlighted the deviation between the OFDP framework and assessment result in the form of a guideline for fiscal data publishers. The overall result is profiling gathered fiscal datasets.

4 THE OFDP FRAMEWORK

We identified a set of comprehensive quality factors which are presented in previous works [Berners-Lee 2010; Caplan et al. 2014; Davies 2013; OKI 2015a; Vetrò et al. 2014,?; Wonderlich 2010; Zaveri et al. 2015]. We also present additional quality factors from our experience in processing Open fiscal data.

The readers are referred to the documents by GODI [OKI 2015d] and ODB [Davies 2013] for the explanation of factors that are originating from the respective documents (see Figure 2), as well as open the data guide [Dietrich et al. 2009; Sunlight Foundation 2014]. The definition of uncommon or non self-explanatory quality factors is provided in this section. The *availability of semantics*, i.e., data availability in a semantic format such as RDF, facilitates data integration and concept linking between different datasets. A dataset publication is *sustainable* whenever it is hosted on a government Open data portal, an official website, or a preservable public platform (e.g., Github). *Timeliness* relates to how soon the data are published by government officials after the data have been collected. This is especially relevant for time-sensitive data. *Permanence* is concerned with getting information over time, which ideally provides an archival feature and *version tracking*. *Open Format* refers to any file format that is published publicly, free of charge, and without reuse limitations, so anyone can read and implement the format without intellectual property constraints [OKI 2015c] (e.g., CSV format). A *code list* is a classification or a set of enumerated concepts that restricts the possible values of a field, e.g., currency or country code.

Some studied quality factors are excluded from OFDP because it is non-trivial to measure these factors in a dataset. We are constrained by several quality factors that are not included in our assessment, such as *granularity*, *accuracy*, and

---

Table 1. Our OFDP quality factors and their weights according to the survey result.

FACTOR	WEIGHT	FACTOR	WEIGHT	FACTOR	WEIGHT
Data Existence*	5	Authoritative	4	Regular Update	4
Easily Available*	5	Complete Code List	4	Search Mechanism	4
Documentation	5	Contact Point	4	Sustainable Publication	4
Free of Charge	5	Dataset Filtering	4	Up to Date	4
In Digital Form	5	English Info Available	4	Version Tracking	4
Mentioned License	5	In Bulk	4	Mentioned Contributors	3
Online	5	Metadata	4	RDF Availability	3
Public	5	Open Format	4	Visualization	3
Structured Data	5	Open License	4	Dereferenceable LD URI*	3
API Availability	4	Persistent URI	4		

completeness. Each of those factors requires a fine-grained definition regarding the level of granular/accurate/complete to make the assessment of the datasets within these factors objective. In practice, the granularity levels of these factors are very diverse across different budget and spending datasets publishers. The quality factor *primacy* [Wonderlich 2010] is implicitly provided by decomposition into the three quality factors: *authoritative*, *mentioned contributors*, and *version tracking*.

Three of the quality factors were excluded from the questionnaire (*easily available* and *data existence* due to their obvious importance, and *dereferenceable linked data URI* since it is overly technical for people outside the linked data community). For these factors, we assigned the weight manually. The term *easily available* refers to how easy it is to obtain the full datasets that contain all of the complementary information without investing a significant amount of time. Being *easily available* and having the *data exist* are very important. Easily available determines how the data can be found for further consumption by interested parties.

Overall, we collected 29 quality factors (Figure 2). Quality factors in OFDP subsume all quality factors in GODI (up to May 2017) and ODB (2013). Subsequently, we weight the quality factors for OFDP framework according to the questionnaire result (described in section 3). The quality factors and their weights are provided in Table 1.

5 EVALUATION

The detailed analysis of the datasets is available publicly in an online spreadsheet[8]. This spreadsheet includes links, full assessment, total assessment score for each dataset and additional contexts (e.g., geographical area, data model, coverage, domain, granularity, and comments). We outlined the analysis in Figure 3, 4, and 5.

Figure 3 provides an overview of the star-data categorization rating [Berners-Lee 2010] of the datasets. Due to restricted license or license unclarity in many of the datasets, a major percentage (72.7%) of the dataset is listed as zero-star. This means that at least one of the necessary permissions required in Open License (access, use, modify, and redistribute) is not clearly mentioned. There are 2.6% of assessed datasets that were categorized as one-star data, none as two-star data, 20.8% as three-star data, 3.9% as four-star data, and none as five-star data. As a side note, two-star data requires the data to be published on the web with an open license, in a structured but proprietary format. In our analysis,

---

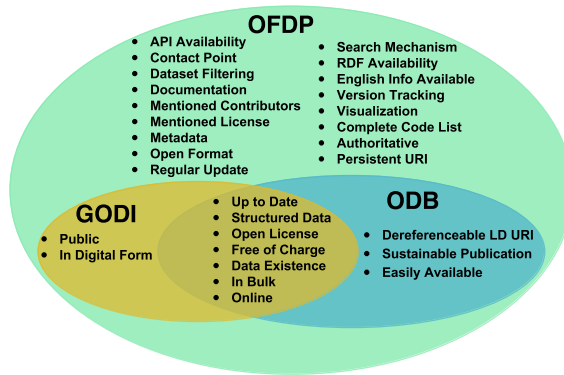[8]http://bit.ly/jdiq-datasheet-view

Fig. 2. Quality factors considered in our OFDP framework which subsume factors from GODI and ODB.
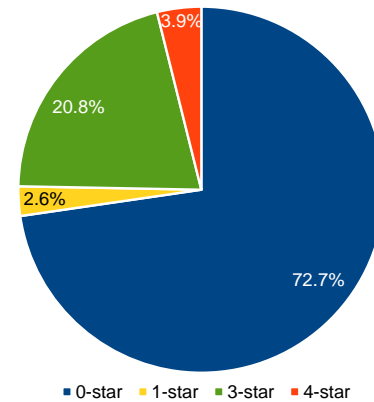

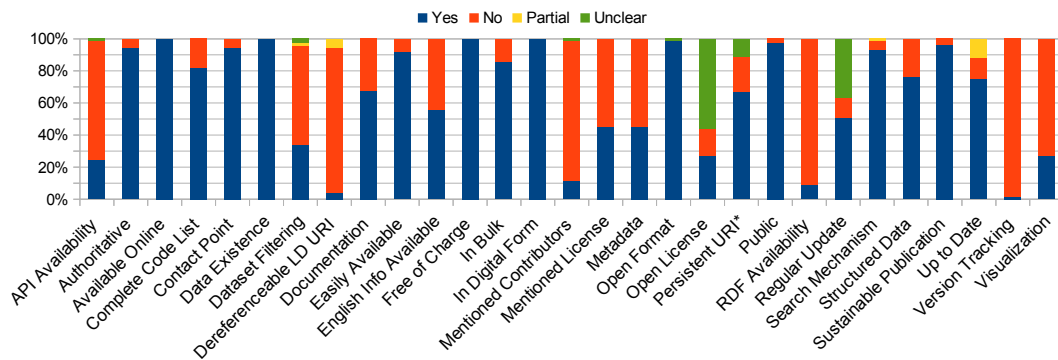Fig. 3. Datasets categorization according to the 5-stars data schema by Sir Tim Berners-Lee.

Fig. 4. Open fiscal data quality factor presence in our sample of 77 Open fiscal datasets.

there are several datasets that are published in Excel format, which previously was a proprietary format. However, Microsoft has published the Excel file format specification openly so that this format can be implemented by anyone.

The percentage of each quality factor's presence in the datasets is shown in Figure 4. Figure 5 plots the resulting score for each dataset using GODI, ODB and OFDP methodologies. The scores of these methodologies are normalized and therefore range from 1-100. We rank the datasets according to these scores. Based on the ranking results, Spearman correlation values are computed. Value of 0.86 between ODB-GODI shows that both rankings are correlated. The values between ODB-OFDP (0.78) and GODI-OFDP (0.75) show a lower correlation as our newly developed OFDP takes more comprehensive quality factors into consideration (see Figure 2). The detailed correlation calculation is available in the dataset analysis spreadsheet.

6   OFDP GUIDELINES TO PUBLISH FISCAL DATA

As a result of our assessment, we recommend that Open fiscal data publishers follow quality factors listed in Table 1. A higher weight indicates a higher priority for the quality factor. We found that most analyzed datasets have performed well for being available online, free of charge, public, in digital form, easily available, in an open format, published

Fig. 5. The proposed OFDP score compared to ODB and GODI.

in a sustainable manner, published with the contact point, authoritative, and also published with search mechanism provided. However, there are many quality factors that need more attention from fiscal data publishers, which we provide in the following section in alphabetical order.

API Availability. Publishing datasets as an API is useful if the data are large and frequently change, especially when only a small portion of the data is needed [Jaquith 2015]. Only 24.7% of the datasets publishers from our analysis provide an API endpoint for their dataset. Some data publishing Content Management Systems (e.g., CKAN, DKAN) provide an API endpoint feature. Publishing datasets via an API endpoint should ideally be accompanied by publishing datasets for bulk download, too.

Complete Code list. Code lists can be used to link concepts among different datasets and enable comparative analysis between different fiscal dataset sources. However, 18.2% of the datasets are not published with complete code lists. We encourage Open fiscal dataset publishers to provide full code lists in a structured format (instead of legal, textual documents). The most popular way is to publish the code lists within the main dataset itself. Based on our experience, the most efficient way for providing code lists is by including a list of the code and the description of each code in a separate file. Therefore, codes and information (e.g., label, descriptions) are covered without redundancy on the main fiscal data itself, and no manual code list extraction effort is necessary.

Dataset filtering. The dataset filtering feature is recommended as it eases the users if a particular selection over the dataset is required. This feature is essential for understanding and analyzing the data by giving a specific selection criterion. Among the analyzed datasets, 62.3% data publishers do not provide this feature.

Documentation. For 32.5% of the analyzed datasets, no documentation could be found, which hinders the understanding of the datasets. Meaningful documentation should be provided and shall consist of at least the datasets content, datasets context, available classifications, and the definition of fields present in the datasets. The OpenCoesione[9] initiative provides a good example for documentation.

English Info Availability. In assessed datasets, 44.2% are published without English documentation. Machine translation is prone to errors especially for classifications and specific terms. We recommend the dataset maintainer to provide at least English documentation, especially for international communities who analyze the data.

In Bulk. Publishing Open data in bulk (e.g., CSV, instead of only as API endpoint) is important because it is a familiar format for non-programmers, easy to mirror, produce, host, and distribute [Jaquith 2015]. Datasets not published in

---

[9]http://www.opencoesione.gov.it/opendata/#fs0713-title

bulk are not *open* according to the open definition[10] as they are not provided as a whole. For technical practicality and openness reasons, the bulk availability should always be considered while publishing Open fiscal datasets. Around 14.3% of the analyzed datasets lack this feature.

Mentioned license. The dataset's license should be mentioned explicitly, stating all basic permissions for public access, usage, modification, and sharing. *Restricted* or *open* license clarity is counted only on 45.4% of the surveyed dataset (Figure 4). We recommend the datasets publications with a commonly known license type.

Metadata. Although metadata helps data acquisition and identification, 54.5% of the datasets are not published with metadata. There are three kinds of metadata: descriptive, structural and administrative metadata [Guenther and Radebaugh 2004]. Descriptive metadata explain datasets discovery and identification, such as title, abstract, author and keywords. Structural metadata describes the arrangement of objects within the data, e.g., table of contents and chapters. Administrative metadata indicates resource management, e.g., technical information, how and when the data was created, intellectual property rights and archival information. Whenever possible, we recommend to provide all these types of metadata following, for example, the W3C recommendation DCAT [Maali et al. 2014], or it's adaption DCAT-AP by the EU Committee.

Open License. An open license allows data users to access, use, modify, and redistribute the data. This is essential to enable and foster data reuse for analysis purposes. Open definition enlists open-conformant licenses which we recommend. In assessed datasets, 27.3% are openly-licensed, 16.9% are restricted, and 55.8% are unclear.

Persistent URI. Maintaining permanent links to datasets is recommended, and at least a redirection mechanism should be provided from the original link once the link has changed. Persistent URI is a relevant concern, as 21.8% of the analyzed datasets are no longer accessible under the previously-valid URI (Figure 4). In addition, a human-readable URI is preferred for the datasets to improve search engine optimization.

Regular Update. The regular update provides an expectation of when interested stakeholders can find the latest dataset. Most of our questionnaire respondents agree that regular update is an important quality factor in publishing Open fiscal data. We also recommend that the dataset's publisher publish their datasets regularly. Unfortunately, only 50.6% of the datasets provide regular updates, while 13.0% of the datasets do not provide regular updates and the other 36.4% are unclear.

Up to Date. We encourage the dataset publishers to provide the latest information so that the dataset's analysis process can be more interesting to do for the stakeholders and journalists. During our analysis (which was done in 2016), we categorized any budget datasets up to 2016 and spending datasets up to 2015 as up to date datasets. From the analyzed datasets, 11.7% are partially up to date, and 13% are not up to date.

Structured Data. Even though publishing structured data allows users to analyze the data easily and maximizes the technical access, 23.4% of the assessed datasets are published in a non-structured format (e.g., PDF). The importance of publishing structured data has been highlighted in previous works [OKI 2015b; Sunlight Foundation 2014; Tauberer 2012; UK HM Government 2012]. Publishing the dataset in a non-structured format makes the transformation process more difficult as no specific pattern can be followed by tools performing datasets transformation. Hence, we highly recommend publishing datasets in a structured data format.

Version Tracking. Version tracking or version control for data supports distributed data contribution, collaboration, broader participation, provenance tracking and incremental development [Pollock 2010]. We encourage the publishers to provide version tracking to see the changes made and the user who changed the datasets. In our analysis, 98.7% of

---

[10]http://opendefinition.org/od/2.1/en/

dataset publishers do not provide a version tracking feature on their datasets web page. CKAN features basic activity monitoring on published datasets, specifying modified data and the user involved.

7 CONCLUSIONS AND FUTURE WORK

In this paper, we described our experience with Open fiscal datasets and analyzed their quality under different aspects. In particular, we achieved the following goals: *(i)* Identify several important factors impacting quality and reuse of Open fiscal datasets, *(ii)* evaluate these factors' relevance, and *(iii)* assess the presence of these factors in recent Open fiscal datasets. The assessment was performed on a representative number of datasets from different public administrations: 77 datasets from different public administrations. Moreover, we compared our assessment results with previous existing assessment frameworks. Our OFDP assessment framework considers a larger and more fine-grained set of quality factors, specifically targeted at fiscal datasets. Several qualitative issues of Open fiscal datasets have been raised within our analysis. Hence, we highlighted these issues and provided guidelines for publishers of Open fiscal data.

In the future, we will enhance the OFDP framework with both datasets and Open fiscal data portals. A semi-automatic quality assessment of Open fiscal datasets might be developed for public administrators to evaluate their fiscal data themselves, which may include a file format development to represent the assessment result. Finally, further studies will focus on whether the adherence or not to the proposed publication guidelines actually influences Open fiscal datasets' consumption.

ACKNOWLEDGMENTS

REFERENCES

Bastian Ahrendt. 2009. *What are the costs and benefits of XBRL in the financial services industry?* Erasmus University.

American Heritage Dictionaries Editors. 2016. *The American heritage dictionary of the English language, Fifth Edition.* Houghton Mifflin Harcourt. Fiscal.

Mark Ballard. 2014. Poor data quality hindering government open data programme. (2014). http://www.computerweekly.com/news/2240227682/Poor-data-quality-hindering-government-open-data-transparency-programme

Florian Bauer and Martin Kaltenböck. 2011. Linked open data: The essentials. *Edition mono/monochrom, Vienna* (2011).

Tim Berners-Lee. 2010. Linked Data - Design Issues. (2010). https://www.w3.org/DesignIssues/LinkedData.html

R Caplan, T Davies, A Wadud, S Verhulst, JM Alonso, and H Farhan. 2014. Towards common methods for assessing open data: workshop report & draft framework. *World Wide Web Foundation* (2014).

CMMI Institute. 2014. Data Management Maturity Model V1.0. (2014). http://cmmiinstitute.com/data-management-maturity

Tim Davies. 2013. Open data barometer: 2013 global report. *World Wide Web Foundation and Open Data Institute* (2013).

Daniel Dietrich, Jonathan Gray, Tim McNamara, Antti Poikola, P Pollock, Julian Tait, and Ton Zijlstra. 2009. Open data handbook. (2009).

Archon Fung, Mary Graham, and David Weil. 2007. *Full disclosure: The perils and promise of transparency.* Cambridge University Press.

Auraloce Graft, Stefaan Verhulst, and Andrew Young. 2016. *Open Data's Impact - Brazil's Open Budget Transparency Portal: Making Public How Public Money Is Spent.* Technical Report. Govlab. http://odimpact.org/files/case-study-brazil.pdf

Rebecca Guenther and Jacqueline Radebaugh. 2004. Understanding metadata. *National Information Standard Organization (NISO) Press, Bethesda, USA* (2004).

James Hendler. 2014. Data integration for heterogenous datasets. *Big data* 2, 4 (2014), 205–215.

Noor Huijboom and Tijs Van den Broek. 2011. Open data: an international comparison of strategies. *European journal of ePractice* 12, 1 (2011), 4–16.

Waldo Jaquith. 2015. API versus Bulk Data. (2015). https://how-to.usopendata.org/en/latest/The-Basics-of-Open-Data/API-vs-Bulk-Data/

Deborah Lucas. 2012. Valuation of government policies and projects. *Annu. Rev. Financ. Econ.* 4, 1 (2012), 39–58.

Deborah Lucas. 2014. Evaluating the government as a source of systemic risk. *Journal of Financial Perspectives* 2, 3 (2014), 45–58. https://EconPapers.repec.org/RePEc:ris:jofipe:0048

Fadi Maali, John Erickson, and Phil Archer. 2014. Data catalog vocabulary (DCAT). *W3C Recommendation* (2014).

OKI. 2015a. Global Open Data Index - Methodology. (2015). http://index.okfn.org/methodology/

OKI. 2015b. How to Open up Data. (2015). http://opendatahandbook.org/guide/en/how-to-open-up-data/

OKI. 2015c. Open Data Handbook: File Formats. (2015). http://opendatahandbook.org/guide/en/appendices/file-formats/

OKI. 2015d. Place Overview: Global Open Data Index. (2015). http://index.okfn.org/place/

Michael Peters, Anna Alberts, and Bela Seeger. 2017. European Structural Funds - A Data Quality Index. (2017). http://openbudgets.eu/post/2017/04/04/esif-data-quality/

Suzanne J Piotrowski and Gregg G Van Ryzin. 2007. Citizen attitudes toward transparency in local government. *The American Review of Public Administration* 37, 3 (2007), 306–323.

Rufus Pollock. 2010. We need distributed revision/version control for data. (2010). https://blog.okfn.org/2010/07/12/we-need-distributed-revisionversion-control-for-data/

Nigel Shadbolt, Kieron O'Hara, Tim Berners-Lee, Nicholas Gibbins, Hugh Glaser, Wendy Hall, and M. C. Schraefel. 2012. Linked Open Government Data: Lessons from Data.gov.uk. *IEEE Intelligent Systems* 27, 3 (2012).

Prianka Srinivasan. 2016. GovLab Index on Open Data – 2016 Edition. (2016). http://thegovlab.org/govlab-index-on-open-data-2016-edition/

Rick Stapenhurst and Mitchell O'Brien. 2008. Accountability in governance. *The World Bank, Washington DC* (2008).

Sunlight Foundation. 2014. Open Data Policy Guidelines. (2014). http://sunlightfoundation.com/opendataguidelines/

Joshua Tauberer. 2012. *Open government data.* https://opengovdata.io/

The World Bank. 2015. Starting an Open Data Initiative. (2015). http://opendatatoolkit.worldbank.org/en/

Alan Freihof Tygel, Judie Attard, Fabrizio Orlandi, Maria Luiza Machado Campos, and Sören Auer. 2016. How Much? is not Enough: an Analysis of Open Budget Initiatives. In *Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance.* ACM, 276–286.

UK Cabinet Office. 2013. Policy paper: G8 Open Data Charter and Technical Annex. (2013). https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex#technical-annex

UK HM Government. 2012. Open Data White Paper Unleashing the Potential. (2012). https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/78946/CM8353_acc.pdf

A Vetrò, Marco Torchiano, C Minotas Orozco, Giuseppe Procaccianti, Raimondo Iemma, and Federico Morando. 2014. An Exploratory Empirical Assessment of Italian Open Government Data Quality With an eye to enabling linked open data. (2014).

John Wonderlich. 2010. Ten principles for opening up government information. *Washington, DC: Sunlight Foundation. August* 11 (2010), 2010.

Amrapali Zaveri, Anisa Rula, Andrea Maurino, Ricardo Pietrobon, Jens Lehmann, and Sören Auer. 2015. Quality assessment for linked data: A survey. *Semantic Web* 7, 1 (2015), 63–93.