Evaluating Semantic Interoperability of Government Open Data Portals

Cathy S. Lin [0000-0002-7976-3523] and Hsin-Chang Yang [0000-0001-5851-2760]

National University of Kaohsiung, Kaohsiung 811, Taiwan
{cathy,yanghc}@nuk.edu.tw

Abstract. Open data emerged in the last decade as governments worldwide endeavor to publicize their owned data for government transparency, effectiveness, and social goodness. Many applications and use cases of open government data have been reported and revealed the potentials of them. Most of the users accessed these data via portals established by governments in different scales. Aspects such as usability, accessibility, and openness should be considered in building such portals to provide users data and services in high quality. In this study, we applied a set of measures covering the semantic aspects of open data portal design and maintenance to evaluate the quality of government open data portals. Results on several major national-level portals were reported.

Keywords: Government Open Data, Open Data Portal, Open Data Quality, Portal Quality Evaluation, Semantic Interoperability.

1 Introduction

Drive by the trend of open data, authorities of the public sector release their data actively or proactively and play the roles of data providers, producing enormous amount of *government open data* (GOD). In this aspect, the governments can improve many aspects of governance, such as transparency, accountability, innovative and intelligent services, and government-citizen interactivity, through the use of open data. It is essential and common to provide a portal, which is generally implemented as a Web service, for accessing a large quantity of data maintained by their owners. Many governments have established their open data portals as the volume of GOD increases drastically recently. As a Web service, an open data portal should meet requirements in various aspects to provide quality services. The arguments on the architecture and functions of open data portals are still ongoing since we lack standard and good practices on these subjects. Schemes for evaluating the quality of portals are thus required for data providers in building their portals. The quality assessment should take account of several aspects such as functionality, usability, accessibility, openness, and so forth. Several schemes have been devised to tackle this problem (Máchová & Lnénicka, 2017).

Data semantics is important information that can be applied to many domains such as information retrieval, e-commerce, knowledge acquisition, etc. An open data portal that can provide datasets together with much semantic information will be helpful for

service providers and application developers. Thus, it is crucial to access the semantic interoperability, which is defined here as the completeness and easiness of implementation and access of the data semantics, of an open data portals. However, current practices on open data evaluation did not pay much attention to this issue. In this work, we will try to give an initial attempt on evaluating the semantic interoperability of open data portals.

2 Quality Assessment Measures for Open Data Portals

Open data portals are essential infrastructure for disseminating open government data. Governments publish their data on these portals and allow individuals or organizations to use them freely to create values. Currently, several measures and metrics have been proposed to ensure the effectiveness of such portals (Máchová & Lnénicka, 2017). According to Máchová & Lnénicka's survey, the United Kingdoms ranks first, followed by India and the United States. However, Taiwan was ranked 21 out of 89, which contrasts with the result of Global Open Data Index[1] and Open Data Barometer[2] (Brown, 2017).

Unfortunately, the semantic interoperability of open data portals were rarely discussed. To evaluate the semantic interoperability of the open data portals, we observed that two major aspects should be addressed. First, the descriptive semantics of datasets should be provided. These descriptive information could be stored as their metadata, which could be easily conformed to existing standards. Second, the portals should provide a way to allow users utilizing such data semantics. In this work, we identified 22 indices which could be used to evaluate the semantic interoperability of open data portals. Table 1 lists the proposed indices. These indices were categorized into 2 classes, namely 'semantic description' and 'semantic operation', which reflect the two aspects mentioned above. We believe that these indices could reflect semantic information in aspects of content and access to some extent. Evaluations on such indices may give an insight into the semantic interoperability of the open data portals.

Table 1. List of indices for evaluating semantic interoperability of GOD portals.

Number	Category	Description
1		Each dataset includes at least a title and a description
2		A creation date is given for each dataset
3	Semantic description	The most recent date on which the dataset was changed, updated, or modified is given for each dataset
4		The identity and role of the agency and person responsible for each dataset is specified
5		The datasets are categorized

[1] https://index.okfn.org/
[2] https://opendatabarometer.org/

Number	Category	Description
6		The datasets are tagged with keywords
7		Each dataset is accompanied by a reference to the language used
8		If a dataset refers to a specific range of time, its temporal coverage is specified
9		If a dataset refers to a specific geographical area, its spatial coverage is specified
10		The descriptive record contains a direct link to the URL of the data
11		The name and email address of the publisher of a dataset are provided
12		The frequency with which dataset is updated is provided
13		Each dataset is accompanied by a descriptive record
14		Metadata associated with each dataset is available in a standard format
15		Metadata describing the datasets is structured in a standard way
16		Any vocabularies used within the dataset are identified and documented
17	Semantic operation	Topics are linked to published vocabularies and taxonomies
18		Each dataset is given a unique identifier
19		A checksum and/or signature is available to verify the validity of each file
20		Data adheres to the defined syntax of any specified vocabularies
21		It is possible to query data and metadata in accordance with standards of the web of data (Linked Open Data)
22		Data sources linked from a dataset are reported

3   Evaluations on Semantic Interoperability of GOD Portals

For preliminary justification of the proposed metrics, we adopted the indices in Table 1 to evaluate some open data portals. In this study, six examiners were assembled to assess the portals using the checklist during June 2015. These examiners are all graduate students taking the open data related courses for at least two semesters. They were also trained in how to apply the indices. We translated the indices into Chinese and gave two-hour instructions on the meaning of each index. Examiners then assessed every portal independently. For each index, the five-level Likert scale was adopted to score the index where a score of 5 means that the function or requirement is fully implemented or applicable and 1 means the function or requirement is missed entirely

or inapplicable. The scores for each index were then aggregated over all examiners as its final score. We evaluated three GOD portals, namely US[3], UK[4], and Taiwan[5], in the preliminary experiments. The result is shown in Table 2. We only show the overall result of the two categories due to space limitation. The result of the evaluation shows that the US portal outperformed the other two in both categories. However, all portals received lower ratings in 'semantic operation' category, reflecting the fact that implementation of functionality and support for standards such as linked data (Berners-Lee, 2006) for access and usage of semantic information were still immature for these portals.

Table 2. The result of the semantic interoperability assessment. The TW, US, and the UK stand for the official GOD portals of Taiwan, US and UK governments, respectively. We only show the average scores of each theme due to space limitation.

Category	TW	US	UK
Semantic description	3.11	**3.73**	3.48
Semantic operation	2.07	**3.47**	2.09

4　Conclusions

In this study, we devised a set of criteria to assess the quality of three major government open data portals on their semantic interoperability. Two aspects of semantic interoperability were defined to measure the descriptive and operational nature of semantics provided in the portals. A total of 22 indices were suggested. We believe that this is the first attempt to assess the semantic interoperability of open data portals. Preliminary experiments on three major government open data portals, namely US, UK, and Taiwan, were conducted. The US portal topped the other two according to the result.

References

1. Berners-Lee, T. (2006). Linked Data. Retrieved from http://www.w3.org/DesignIssues/LinkedData.html
2. Brown, C. (2017). Open Source Data: Just How Open Is Taiwan's Government? Retrieved from https://www.ketagalanmedia.com/2017/07/27/open-source-data-just-open-taiwans-government/
3. Máchová, R., & Lnénicka, M. (2017). Evaluating the Quality of Open Data Portals on the National Level. Journal of Theoretical and Applied Electronic Commerce Research, 12(1), pp. 21-41.

---

[3] https://www.data.gov/

[4] https://data.gov.uk/

[5] https://data.gov.tw/