The Open Government Data Stakeholder Survey

Michael Martin+, Martin Kaltenböck*, Helmut Nagy*, and Sören Auer+

+ Universität Leipzig, Institut für Informatik, AKSW,
Postfach 100920, D-04009 Leipzig, Germany,
{martin|auer}@informatik.uni-leipzig.de – http://www.aksw.org/
* Semantic Web Company GmbH,
Lerchenfelder Gürtel 43, A - 1160 Wien,
{m.kaltenboeck|h.nagy}@semantic-web.at – http://www.semantic-web.at/

Abstract. This paper describes the results of the LOD2 Open Government Data Stakeholder Survey 2010 (OGD Stakeholder Survey). The objective of the survey was to involve as many relevant stakeholders as possible in the 27 European Union countries in an online questionnaire and ask them about their needs and requirements in the area of open data as well as for the publicdata.eu portal. The main areas of the survey have been questions about Open Government Data itself, questions about data, about the usage of data, questions about the requirements for a centralised data catalogue as well as questions about the participants themselves.
The goal of the OGD Stakeholder Survey has been to reach a broad audience of the main stakeholders of open data: citizens, public administration, politics and industry. In the course of the survey that was open for 5 weeks from November 2010 to December 2010 in total 329 participants completed the survey. The results have been published in April 2011 in the form of HTML and PDF, the raw data in CSV. In addition to these publication formats (HTML, PDF, CSV) we published the data also as Linked Data using various vocabularies and tools.

1 Introduction

The idea for the LOD2 Open Government Data Stakeholder Survey 2010 (OGD Stakeholder Survey) appeared in the course of the requirements elicitation and specification phase of the EU funded project LOD2 - Creating Knowledge out of Interlinked Data[1]. As one of the three use cases of the LOD2 project is publicdata.eu[2] – the design and implementation of a single point of access, a centralised data portal / data catalog for open data in EU 27 using linked data principles and technologies – the objective was to involve as many relevant stakeholders as possible in the 27 European Union countries in an online questionnaire and ask them about their needs and requirements in the area of open data as well as for the publicdata.eu portal.

[1] LOD2 Website: http://www.lod2.eu
[2] http://publicdata.eu

The main areas of the survey have been questions about Open Government Data itself, questions about data, about the usage of data, questions about the requirements for a centralised data catalogue as well as questions about the participants themselves.

The goal of the OGD Stakeholder Survey has been to reach a broad audience of the main stakeholders of open data: citizens, public administration, politics and industry. The survey has been designed and set up by the LOD2 project partners Open Knowledge Foundation (UK)[3] and the Semantic Web Company (Austria)[4]. Support has been given by the LOD2 partners DERI Galway (Ireland)[5], Wolters Kluwer Germany[6] and the University of Leipzig (Germany)[7]. The survey has been realized using a web based survey tool and has been promoted via blogs, mailings, mailing lists and additional viral marketing channels as well as at related events in Europe. In the course of the survey that was open for 5 weeks from November 2010 to December 2010 in total 329 participants completed the survey. The results have been published in April 2011 in the form of HTML and PDF, the raw data in CSV.

In addition to these publication formats (HTML, PDF, CSV) we published the data also in the *Resource Description Framework* (RDF) using various vocabularies and tools as described in section 5. Publishing data in that format enables interested users to aggregate information on the basis of own constructed queries using SPARQL [7]. Due to the fact that the data is interlinked with further data sets in the *Linked Open Data Cloud* (LOD) [4], users are able to construct queries not only about resources in the local information space, which enhances information retrieval enormously.

2 The OGD-Stakeholder Questionnaire

The structure of the OGD stakeholder survey has been split into the following major sections:

1. questions about 'Open Government Data' (OGD)[8] in general
2. data related questions,
3. questions related to the usage of Open Government Data,
4. requirements for a centralised data catalogue,
5. questions about the area of activity of the participants,
6. and questions about the participants.

The questionnaire in total had about 20 questions – single and multiple choice questions, as well as matrix of choice questions and also many open question for

[3] `http://www.okfn.org`
[4] `http://www.semantic-web.at`
[5] `http://www.deri.ie`
[6] `http://www.wolterskluwer.de`
[7] `http://aksw.org`
[8] `http://opengovernmentdata.org/`

text input to receive as much feedback and input of the participants ideas and expectations as possible.

The first section of the questionnaire started with an open questions asking for general remarks and ideas regarding 'Open Government Data' (OGD) and the motivations and expectations the participants do have thinking of Open Government Data. In the second question the participants where asked to select an OGD user type for themselves. The available user types where developed along the value creation chain: "producer and publisher" (mainly producing and publishing OGD), "use and produce" (using and producing OGD) and "user and consumer" (mainly consuming OGD).

The second section of the survey contained several questions about the expected data types, formats quality etc. In the first question the participants could show their interest in different domains of data by ranking them. The following questions where multiple choice questions to identify the formats of data used by the participants at the moment and the formats the participants wanted to see / use in the future including an open question for formats not included in the multiple choice questions. The second section concluded with two questions regarding the importance of the regional provenance of data (regional, national, EU-wide, worldwide) and the quality of data (e.g. format, completeness etc.).

In the third section the participants could state what they are actually doing with OGD, what they would want to do and they could state the importance of OGD for their everyday work. The section concluded with an open question asking the participants to state the importance of OGD in more detail and to give examples on how they are using OGD at the moment.

In the next section the participants could express their expectations regarding an OGD data catalogue by stating what features and information they were expecting to have in a data catalogue. Again the section closed with a open question asking the participants to give some more details on their expectations and also asking them to state their opinion on several special issues like licensing, the demand for a 'European Data Market Place' and what actions / activities could bring such a market place in position.

Finally the questionnaire ended with some questions about the professional background of the participants (e.g. workplace, workplace location etc.) and about the participants themselves (e.g. age, educational background).

2.1 Target groups and structure of the survey

The target groups of the OGD Stakeholder Survey has been the main stakeholder groups that are involved in open data: citizens, politicians, public administration and industry as well as the 2 additional target groups: media and science.

The structure of the questionnaire has been well discussed by the survey team members and has been chosen to provide a well structured, nicely arranged, easy and quick to fill questionnaire as well as to receive important input and ideas for the Open Government Data use case in the LOD2 project by the wisdom of the crowd.

3 The OGD-Stakeholder survey

The process of the survey creation started in October 2010 and the OGD Survey has been launched on 08 November 2011. The duration of the survey has been 5 weeks until the final date of 15 December 2011. In total 329 participants filled the survey: 185 participants completed the survey filling all given questions (including 'about you' questions) and 144 participants filled the survey only partial (mainly questions about the participants themselves where not filled here).

The questionnaire has been set up using a web based tool named SurveyGizmo[9]. It allows to use several types of questions; for instance multiple choice questions allowing single or multiple answers, open questions (to fill in free text) and matrix of choice questions providing answer rankings etc. SurveyGizmo also provided useful reporting tools that enabled – besides limited HTML and PDF reporting – the export of a CSV file as an open machine readable format for further re-use of the raw results data that was the basis for the survey analysis and the publishing of the survey results data as linked open data.

The survey has been open to the interested public and has been promoted via existing mailing lists (LOD2, OKFN, W3C etc), direct mailings to experts (via the European Commission), via several blog posts as well as using additional viral marketing channels as for instance Twitter and Facebook.

Furthermore, the survey has been pro actively promoted at events where the LOD2 team participated as e.g. the 'Open Government Data Camp'[10] in London in November 2011, the 'EuroVoc conference'[11] and the Open Data Workshop in Luxembourg [6] in November 2011 and many more events in several EU countries using presentations and flyers to reach the targeted audience.

3.1 Analysing the survey – report and result generation

The main target of the OGD Stakeholder Survey was to include the input and ideas of as many relevant stakeholders as possible in European Union 27 countries in the requirement elicitation process for the 'Open (Government) Data Portal' (publicdata.eu) that will be developed in the OGD use case of the LOD2 project. Since the survey was publicly available it has reached an audience not only in Europe but also in America, South America and Asia. Still most participants come from the European Union countries.

The analysis of the results of the survey has been done by the members of the LOD2 consortium. It is presented in this paper mainly focused on interpreting the results in respect to the requirements for the OGD portal. Since the sample of 185 complete results showed that no significant results could be gained from the personal questions at the end of the questionnaire also the 144 partial results have been included in the sample for the analysis since it was mostly the personal information that has been missing in those partial results. So the complete sample of 329 questionnaires has been used for the analysis.

[9] http://www.surveygizmo.com

[10] http://opengovernmentdata.org/camp2010/

[11] http://eurovoc.europa.eu/drupal/?q=node/936

All data has been exported to a CSV file preserving the information which questionnaires have been filled in completely and which ones partially. The analyses has been based on this CSV file and the result of the whole sample have been correlated to the following factors: user type, age, company size and region.

As mentioned before the personal questions have not been answered by the partial results so only for the user type there where answers for most of the sample (92%) for the other factors the sample was reduced to 50-60% of the questionnaires. Still we expected that correlating the complete results to those factors could show some interesting tendencies. Since there was no question which region the participants come from but which country they come from we also decided to assign the countries to two regions: Already involved in OGD and starting with OGD.

We thought that we would find differences in the results especially for the questions regarding the requirements for the OGD data portal in those two regions. The table 1 shows which countries were assigned to which region.

Already involved in OGD		**Starting with OGD**	
GB	27	DE	44
ES	11	AT	21
NL	10	IT	12
US	9	FR	6
CA	3	BE, FI, IN	4
AU, IE, IM, NZ	1	HU, NO, PT	3
		AD, BY, BR, HR, GR, LU, RO, RU, SI, SE	2
		AL, AR, BT, BG, CL, DK, IL, MX, PL, SM, CH, TW, TR	1
SUM (%)	61 (18,54%)	**SUM (%)**	93 (28,27%)

Table 1. Number of participants per country

Finally, the open questions where partly included in the analysis by just providing quotes (e.g. additional formats) but also relevant / often mentioned aspects / categories where derived from the open questions (e.g. expectation regarding an OGD portal) and the tendencies (positive to negative) where calculated based on the answers (e.g. motivation / expectation on OGD).

4 Results of the survey

The following section shows the most important results from the analysis of the OGD Stakeholder Survey in context of the requirements we developed for the OGD data portal. The survey showed that regarding the preference for the "format of data" going for RDF/XML and APIs is the right direction. At the

moment formats like HTML, PDF or CSV are most widely used but the participants show that they expect to use APIs, XML and RDF in future (see figure 1). The results per user type show that those formats are already more important for the "user and producer" type. There are also some more formats like e.g. JSON mentioned that should be taken in consideration.

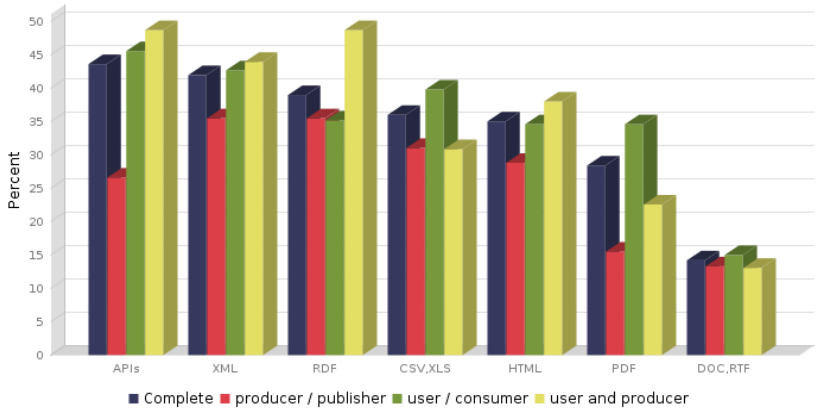

Fig. 1. Formats of data expected to use in future

Regarding the provenance and quality of data the results of the survey show that 'national' data is ranked highest followed by 'regional', 'EU-wide' and 'worldwide' and that there is a focus on 'provenance/source of data', 'format of data' and 'completeness of meta data'. Compared to the results per age we can see that provenance and source of data is more relevant to the older groups while in the younger groups completeness of meta data is valued higher. In relation to the format preferences mentioned before the integration of data conversion mechanisms is an issue that should be considered to be integrated in a data portal and especially a harmonised meta data structure seems to be already a step in the right direction for an EU27 data catalogue.

The top ranked topics regarding what users want to do with Open Government Data are 'research / analysis', 'visualisation', 'simply consuming the data' (see figure 2). This can also be related to the 'expected to have' and the 'like to have' features of a open data catalogue. Expected to have features are: "providing raw datasets", "information about versions of data sets" and "searching exploring, grouping and clustering of data sets". Like to have features are: "crowd sourcing mechanisms", "alerts on (regional) information" and "analysis of visualisation tools".

Again these matches the format preferences mentioned above and shows that for the OGD data portal there should be a strong focus on search mechanism, the visualisation of search results and data, and a focus on features for data curation. Finally, the results show that users are still looking for basic information on OGD and the use of OGD since 'white papers & best practice', 'news on Open

Government Data' and 'and use cases & success stories' are ranked highest when asked for the information that should be provided by an OGD data portal.

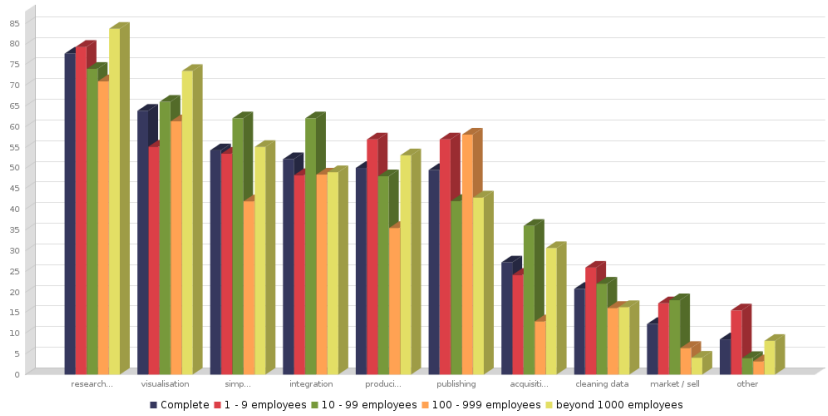

Fig. 2. Use of data.

The results of the LOD2 Open Government Data Stakeholder Survey has been published in several ways and formats as follows: Via the main entry page of the survey results under: http://survey.lod2.eu an introduction about the survey itself can be found. Giving the idea, the objectives as well as the questionnaire itself and finally a short summary of the most important results. From this entry page there are links to the following survey results:

1. the survey results in the form of HTML pages (including several charts as well as cross-depending analysis) structured in areas as well as giving the results along the questionnaire,
2. in the form of a PDF for printing and download,
3. the raw data of the survey results in open and machine readable CSV format for unrestricted re-use as for own analysis and / or visualisations etc.,
4. in the form of linked open data for re-use, browsing as well as querying via a SPARQL endpoint (using the open source tools Virtuoso [3] and OntoWiki[1]) put in context by establishing links to DBpedia [5] to allow more complex queries and richer results.

The survey results have been promoted via several communication channels of the LOD2 project as mailing lists, blog posts, tweets, presentations and via direct mailing to participants. The HTML survey result area provides commenting functionality to enable feedback mechanisms to include this feedback into future work.

5 Publishing the collected data in LOD

In this section we describe how we modelled the survey as RDF, how we published and integrated it into the LOD cloud.

5.1 Creating the RDF representation of the survey data

In order to represent the survey results as RDF we created the survey RDF schema [12] (SRS) depicted in figure 3. At first we created RDF representations of the survey elements such as *survey sections* (5 resources), *survey questions* (60 resources) and *survey options* (221 resources) depending on the answer type (freetext or multiple choice) of the specific question. Compared to the original

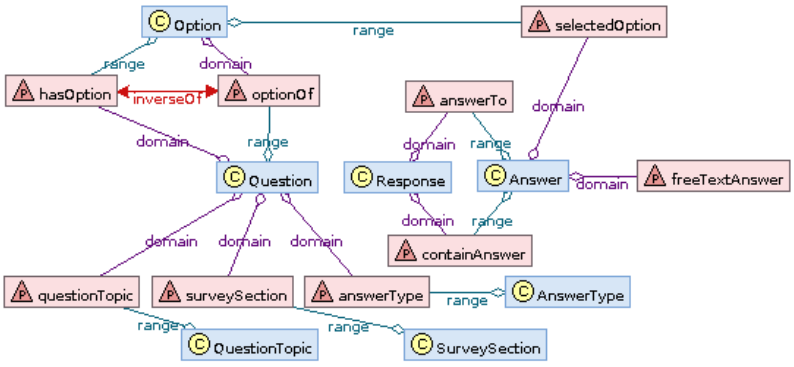

Fig. 3. Schema of the questionnaire.

version of the survey we encoded not only 20 questions but overall 60. Due to the fact that 6 of the multiple choice question are similar to the following example, we had to encode every line of such tables as question that aggregates the column headers as options and the leading question as the question topic.

```
8.) Which features do you find most important for an Open
    Government Data Catalogue? (min 1 answer required)

                      expect to have    like to have   no opinion  don't want
providing raw data          [ ]             [ ]            [ ]         [ ]
registering data sets       [ ]             [ ]            [ ]         [ ]
...
```

After encoding the questionnaire, we transformed the survey results into RDF. The survey results were represented originally in CSV, wherein every row contained the data of one survey response. The central class in SRS is `srs:Response` which is used to type all survey response resources. All information about a particular survey participant is attached to such a resource, like *submission date*, *spoken language*, *provenance* etc. Furthermore, all answers encoded as resources of type `srs:Answer` given by the participant are attached to the respective survey response resource. The resulting dataset contains at the end 329 survey response resources, 12,891 answer resources and overall more than 70,000 triples.

Some of the used properties and resources are not part of the survey namespace. In addition to RDF/S and OWL we used also the Dublin Core Vocabu-

[12] Survey RDF Schema: `http://ns.aksw.org/survey/`

lary[13] and the *Friend-Of-A-Friend* Vocabulary (FOAF)[2] to represent information about the maintainer of the data. Furthermore, some of the properties such as `srs:city` and `srs:country` are interlinked with the DBpedia Ontology[14]. Supplementing data for questions regarding the geospatial context were pulled from DBpedia in order to retrieve further information about that resources.

5.2 Publication of the data

We selected the Virtuoso-backed OntoWiki deployment accessible at `http://data.lod2.eu/` to provide the LOD integration. This setup allows the publication of our survey data in a both human and machine friendly way.

Browsing the data with OntoWiki OntoWiki itself is a web application providing support for agile, distributed knowledge engineering scenarios. It facilitates the visual presentation of knowledge bases as information maps, with different views on instance data[15]. These views are for example the list view providing sets of resource links according to selected filter criteria and the resource view presenting information about the selected resource. The selection of the central class of that dataset `srs:Response` leads to the list of all survey responses. To obtain information about a particular survey response such as depicted in figure 4 an element of the resulting list has to be selected.

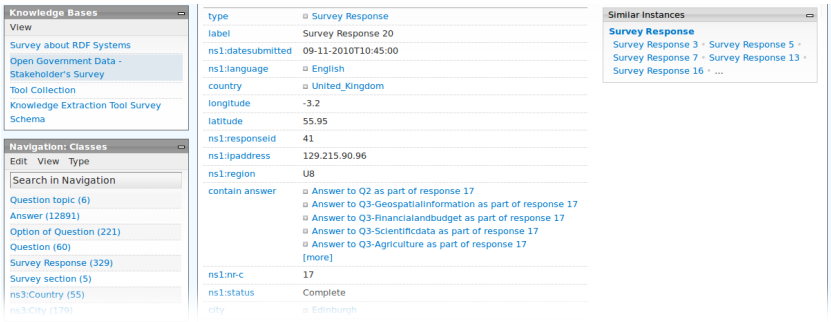

Fig. 4. Screenshot of the OntoWiki GUI displaying a survey response resource.

OntoWiki as SPARQL Query Editor and SPARQL Endpoint In addition to the human centric data exploration via OntoWiki, we also provide data access via SPARQL interface. OntoWiki's SPARQL endpoint is accompanied by a graphical and easy-to-use query editor[16].

[13] Dublin Core Terms: `http://purl.org/dc/terms/`
[14] DBpedia ontology namespace: `http://dbpedia.org/ontology/`
[15] OntoWiki Projectpage: `http://code.google.com/p/ontowiki/`
[16] Shortcut to the Editor: `http://data.lod2.eu/OGD/sparql/editor`

OntoWiki as Linked Data Server and Client While transforming the data into RDF we designed the URIs in accordance with LODs principle of de-referenceability in mind. This type of publication is also supported by OntoWiki built-in functionality in order to create accessibility of resources for Linked Data clients.

6 Conclusion

We presented an overview on the creation and main results of the Open Government Data Stakeholder Survey, which was performed by the LOD2 project in the end of 2010. The analysis of 329 survey results showed that facilitating Open Government Data is of crucial importance. It is interesting to see that national data is deemed the most important resource, followed by regional, EU and world-wide data. Also, most of the stakeholders seem to still slightly prefer APIs and XML to RDF for data access. In order to make the results of the survey available as Linked Data we developed a survey vocabulary and represented the survey results adhering to this vocabulary. The results are published at `http://survey.lod2.eu` for humans and at `http://data.lod2.eu` as Linked Data. We plan to perform a similar survey later in 2011 or 2012 in order to observe how the stakeholder opinions evolve.

Acknowledgments

This work was supported by a grant from the European Union's 7th Framework Programme provided for the project LOD2 (GA no. 257943).

References

1. S. Auer, S. Dietzold, and T. Riechert. OntoWiki - A Tool for Social, Semantic Collaboration. In *ISWC 2006*, volume 4273 of *LNCS*. Springer, 2006.
2. D. Brickley and L. Miller. FOAF vocabulary specification. Technical report, FOAF project, 05 2007. `http://xmlns.com/foaf/spec/20070524.html`.
3. O. Erling and I. Mikhailov. RDF support in the virtuoso DBMS. In *CSSW*, volume 113 of *LNI*, pages 59–68. GI, 2007.
4. T. Heath and C. Bizer. *Linked Data: Evolving the Web into a Global Data Space*. Morgan & Claypool, 1st edition, 2011.
5. J. Lehmann, C. Bizer, G. Kobilarov, S. Auer, C. Becker, R. Cyganiak, and S. Hellmann. DBpedia - a Crystallization Point for the Web of Data. *JWS*, 7(3):154–165, 2009.
6. N.N. Report - technical workshop on the goals and requirements for a pan-european data portal, November 2010.
7. E. Prud'hommeaux and A. Seaborne. SPARQL Query Language for RDF. Technical report, W3C, 2008. `http://www.w3.org/TR/rdf-sparql-query/`.