Austrian Environmental Data Consumption – A Mashup-based Approach

Peter Wetz, Tuan Dat Trinh, Ba Lam Do, Amin Anjomshoaa and A Min Tjoa

Vienna University of Technology,
Favoritenstraße 9-11, 1040 Vienna, Austria
peter.wetz@tuwien.ac.at,
{dat,lam,anjomshoaa,amin}@ifs.tuwien.ac.at,
http://www.ifs.tuwien.ac.at

Abstract. Researching Environmental Open Data consumption needs to be a key interest when dealing with environmental issues from an information technology point of view. Therefore, on the one hand, this paper aims at presenting the status quo of suitable Open Data sources in Austria. On the other hand, based on the available datasets, a novel data consumption approach relying on Linked Data Widgets and combining Linked Open Data and Environmental Open Data is presented. First, the field of research is introduced and related work is described. Then, the evaluation of datasets is illustrated and relevant results are presented. Finally, a new Mashup-based approach for consuming, combining and presenting environmental data via Linked Data principles is proposed. This working example demonstrates an easy and flexible approach for the creation of innovative services, helping, e.g., in environmental decision making. Above all, the creation of such Mashups does not require any specific knowledge in the Semantic Web or environmental domain, hence enlarging the target group significantly.

Keywords: Environmental Informatics, Linked Open Data, Linked Mashups, Linked Widgets, Data Consumption

1 Introduction

In recent years, problems concerning our environment rose to an extent, which cannot be ignored anymore. The annual report of the United Nations Environment Programme (UNEP), called Global Environment Outlook (GEO), points out that the "World Remains on Unsustainable Track Despite Hundreds of Internationally Agreed Goals and Objectives" [1].

The concept of Open Data and Linked Open Data (LOD) has the potential to partially deal with these global problems by providing a means of making certain data available to the public, and therefore enabling bigger awareness about our environment. In addition, based on such data, new use cases and applications can be built to support cognitive activities, e.g. decision making, when dealing with environmental

issues. In order to prove these ideas, this paper presents a Linked Data Mashup-based architecture, which enables exploitation of such data.

Open Data stands for data being freely available for everyone, so that people can look at the data and also reuse and republish it. The Open Knowledge Foundation [2] defines Open Data as follows: "A piece of data or content is open if anyone is free to use, reuse, and redistribute it". *Mashups* are defined as "web applications generated by combining content, presentation, or application functionality from disparate Web sources" [3]. Therefore, Mashups allow for building effective and light-weight information processing solutions based on exposed web services [10].

We aim at bringing the concept of LOD into the environmental domain by enabling the creation of LD Mashup services, combining LOD and environmental data. After presenting relevant Related Work (see Sect. 2), the paper presents insights into Austria's Environmental Open Data landscape (see Sect. 3). These serve as a foundation for showcasing an environmental data consumption approach based on a Mashup architecture (see Sect. 4). This section shall be seen as a working use case of how local datasets can be integrated and utilised in a practical manner. A Mashup-based solution to consume Linked Data (LD) and Open Data enables flexible creation of unique use cases. That is accomplished by combining the advantages of LOD and Mashups. The resulting Mashups rely on a semantic data structure, which allows for querying and linking amongst them. Finally, this enables novel ways of data consumption with data coming from heterogeneous sources in the environmental domain.

2 Related Work

In the field of LOD combined with Mashups some research has already been done with a similar target as presented and discussed in this work, i.e. the utilisation of Linked Open Environmental Data and Mashups to create novel services.

The RDF book Mashup demonstrates how Web 2.0 data sources can be integrated into the Semantic Web in order to close the gap between non-RDF resources and RDF client applications [4]. Another example is the Bio2RDF tool, which addresses the problem of knowledge integration in bioinformatics [5]. DERI's Semantic Web Pipes focus on creating solutions for dynamic data sources [6]. Their solution acts as a workflow, which processes a set of input data, such as RDF, XML, Microformats, JSON and binary streams, in order to deal with the data integration process.

All of these approaches look promising, however, either they require very good knowledge of Semantic Web technologies so they can be used effectively, or they are not extendible. This paper presents an approach, which enables Mashup-based data consumption of LOD and environmental data in a flexible manner without the need to have expert knowledge. Additionally, the semantification of widgets and Mashups enables further improvements.

3 Datasets

In previous work we conducted a survey on available Open Data dataset sources from Austria [7]. Available Open Data portals of Austria were looked up and 139 environmental datasets from eleven different sources were identified and investigated on different Open Data related parameters. Subsequently, this study provides the basis for the proposed environmental use cases built on our Mashup-based approach.

Table 1 demonstrates a summary of the survey's results and shows the absolute counts of datasets according to data types and content types. Data Types identify the nature of the stored data, e.g. geographical data, image data, statistical data, etc. Content Types declare which real world entity is described in the dataset, e.g. territories, air, water, etc.

Table 1. Analysis of Datasets according to Data Types and Content Types

Data type / Content type	*geo*	*image*	*measured*	*basic*	*statistical*	*Total*
territory	81	12	-	-	-	93
air	-	-	23	-	-	23
water	-	-	1	-	2	3
waste	12	-	-	3	2	17
energy	-	-	-	-	2	2
geology	1	-	-	1	-	2
Total	94	12	24	4	6	139

It can be observed, that most datasets offer geographical data as data type (94). Measured data represents sensor data, which is updated very frequently. Four out of eleven data sources provide this type of data. The remainder of the dataset consists of images (12), statistical data (6) and basic data (4). Regarding content types, Table 1 also shows, that most datasets contain territorial data (93), meaning, that these datasets describe the landscape. For instance, this can be park areas in urban regions. There is also much data containing air quality measures (23).

Additionally, a very important indicator of Open Data datasets quality is the compliance to the five star scheme [8]. Most of the datasets are three star datasets (84%), a minority of them being only two star compliant (16%). Not a single dataset is four star compliant or above, implying that no dataset can be categorized as LOD.

Based on the survey, we choose a dataset containing hourly updated air quality values of the city of Vienna for showcasing the Mashup-based approach described in the next Section. We choose this dataset, because it gets updated frequently and contains important environmental information, therefore yielding more interesting results in a real world scenario than static data.

4 Mashup-based Data Consumption

4.1 Approach

The results of the conducted survey (see Sect. 3) show, that there is no LOD dataset in the Austrian Environmental Open Data domain. Still, in this research, location data from the Open Government Data of the City of Vienna[1] were converted to Linked Data and made accessible via a SPARQL Endpoint[2]. Based on this, we present an approach, which combines an environmental dataset from the survey, i.e. near real-time air quality measure[3], with location data available as Linked Data.

We propose an architecture based on the idea of Mashups. The lightweight resource composition of Mashups makes them appropriate for environmental applications, where different data sources can be combined together in a "quick and dirty" manner. The main components of the Mashup stack are APIs (serving as data sources), widgets (serving as modules, which are interconnected to build a Mashup) and the Mashups themselves. This approach introduces special widgets called Linked Data widgets, which are described semantically and based on the W3C widget specifications [9]. The semantification of widget specifications enables query and retrieval of required widgets and also makes them accessible via the LOD cloud in future. Additionally, they are expected to improve the Mashup composition process, e.g. in terms of error prohibition. For these numerous reasons, we aim at designing a system, which benefits from the Mashup stack to integrate heterogeneous environmental data sources with LOD sources to, ultimately, enable innovative ways of data consumption.

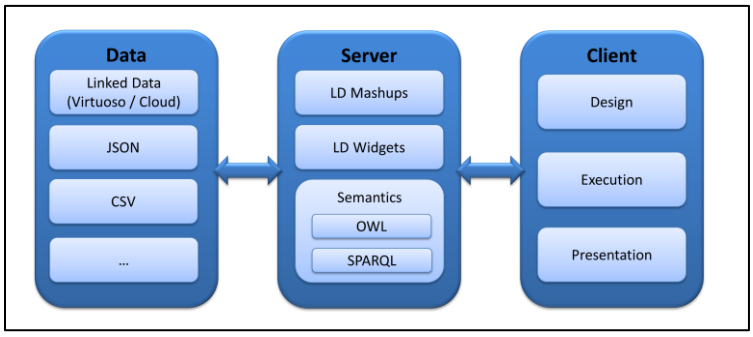

Fig. 1. Architecture of the Mashup-based Environmental Data Consumption Approach.

By leveraging the proposed architecture (see Fig. 1), we show how Linked Data can be integrated with external environmental data to flexibly compose a new service as a Mashup (see Fig. 2). The architecture depicts the following building blocks:

[1] https://open.wien.at/
[2] http://ogd.ifs.tuwien.ac.at/sparql
[3] http://www.data.gv.at/datensatz/?id=8b3b3cdf-2be6-4f0b-8c86-f6be67e5b002

- **Data:** this component shows, that for designing Mashups, different kinds of data sources can be used. They can vary in terms of format or location. Furthermore, Linked Data can be leveraged. Currently, we use our own LD repository – stored in a Virtuoso[4] environment – as well as external Linked Data from the LOD cloud.
- **Server:** the server side element provides the storage of LD widgets and Mashups. For this purpose, an ontology was developed, which structurally describes these objects. This enables unique identification of widgets and Mashups, allowing semantic storage, retrieval, reuse, exploration and referencing via URIs.
- **Client:** this component implements a user interface to enable the design of Mashups, as can be seen in Fig. 2. In addition, created Mashups can be executed and corresponding results are presented.

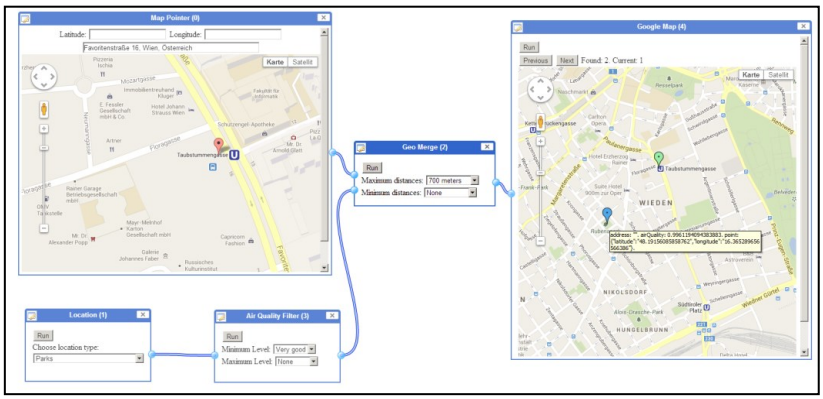

Fig. 2. Design of a new Mashup combining location and air quality data.

4.2 Use Case

To present an actual use case of the introduced architecture, Fig. 2 depicts the design of a new Mashup, which can answer the following query: "Find all *parks* located within 700m to my *office*, which have very good *air quality*".

The Map Pointer-module is used to define an arbitrary location, i.e. my office, as the first input. The second input is defined via the Location-module's dropdown menu as Parks. As a matter of fact, the location data for the parks is obtained from the aforementioned Linked Data repository, but this complexity is encapsulated and hidden from the user. Since we want to find only parks with very good air quality, the Air Quality Filter-module is used and configured accordingly. In the next step, the Geo Merge-module is needed to combine both input variables, based on the maximum and minimum distance parameters, into one result set. To visualise the results, the Google Map-module is used. As can be seen, two

[4] http://virtuoso.openlinksw.com/dataspace/doc/dav/wiki/Main/

5 Results and Future Work

We present parts of the results of a survey on available Environmental Open Data datasets in Austria. We have identified, that most datasets contain geographical data. Frequently updated datasets, which are of high importance for environmental decision making, are rare. No dataset could be identified as being four or five star Open Data according to LOD's data quality schema. Especially for Open Data in the environmental sector, five star data would be of high importance, since novel applications, built on the advantages of Linked Data and the Semantic, would then be enabled.

Based on the survey results and on the transformation of Open Data to LOD, two datasets are used to showcase a working example of a novel data consumption approach. We demonstrate the integration of Environmental Open Data with LOD by the help of a LD Mashup-based architecture. As a result, our prototype shows the simple and flexible creation of LD Mashups, which help non-expert users to draw new information out of raw data. The presented use case combines air quality data with location data. Future work will focus on improving the semantification of widgets and Mashups, thus enabling, e.g., special support in the Mashup-creation process.

Acknowledgement. This research was partially funded by the Vienna University of Technology through the Doctoral College Environmental Informatics.

References

1. UNEP: GEO-5 Assessment Full Report. http://www.unep.org/geo/pdfs/geo5/GEO5_report_full_en.pdf (2012). Accessed 30 August 2013
2. Open Knowledge Foundation. http://opendefinition.org/ (2013). Accessed 30 August 2013
3. Jin Yu, Benatallah, B., Casati, F., Daniel, F.: Understanding Mashup Development. Internet Computing, IEEE 12(5), 44–52 (2008). doi: 10.1109/MIC.2008.114
4. Bizer, C., Cyganiak, R., Gauß, T., Freie Universität Berlin: The RDF Book Mashup: From Web APIs to a Web of Data, 3rd Workshop on Scripting for the Semantic Web. In: 6, 2007. Available online at http://CEUR-WS.org/Vol248/paper4.pdf, (2007)
5. Belleau, F., et al.: Bio2RDF: Towards a mashup to build bioinformatics knowledge systems. J. of Biomedical Informatics 41(5), 706–716 (2008). doi: 10.1016/j.jbi.2008.03.004
6. Le-Phuoc, D., et al.: Rapid prototyping of semantic mash-ups through semantic web pipes. In: Proceedings of the 18th international conference on World wide web, pp. 581–590. ACM, New York, NY, USA (2009). doi: 10.1145/1526709.1526788
7. Wetz, P., Anjomshoaa, A., Tjoa, A M.: A Survey on Environmental Open Data in Austria. In: Proceedings of the IEEE International Conference on Systems, Man and Cybernetics (SMC) (2013) (in press)
8. Tim Berners-Lee: Linked Data - Design Issues (2006)
9. "Requirement For Standardizing Widgets". http://www.w3.org/TR/widgets-reqs/ (2011). Accessed 30 August 2013
10. Anjomshoaa, A., Bader, G., & Tjoa, A. (2009, August). Exploiting mashup architecture in business use cases. In International Conference on Network-Based Information Systems, Indianapolis, USA, pp. xx–xxvii. IEEE.