

The Open Data Workflow: Towards an Open Data Middleware

Position Paper “USING OPEN DATA: policy modeling, citizen empowerment, data journalism”

Olaf Görlitz, Gerd Gröner, Christian Hachenberg, Stefan Scheglmann, Matthias Thimm
Institute for Web Science and Technologies (WeST), University of Koblenz-Landau, Germany
{goerlitz,groener,hachenberg,schegi,thimm}@uni-koblenz.de

Abstract

This position paper gives a brief overview on current and future work of the Institute for Web Science and Technologies (WeST) in the field of using Open Data for eGovernment, eParticipation, and similar topics. With this document, we would like to stress our interest in the workshop “USING OPEN DATA: policy modeling, citizen empowerment, data journalism”.

Introduction

Political decision making is often criticised for its lack of transparency or for referencing ambiguous figures which cannot be verified easily. However, usually large amounts of data, which are of interest to the general public, are collected by governmental institutions but often only a small amount of datasets is made publicly available. The Open Data and eGovernment movement wants to address these issues by promoting open access to these kinds of datasets. This is to improve transparency and legitimacy of complex processes instantiated by policy decision-makers in the government or local authorities. Furthermore, these processes should be accompanied and supported by an active citizenship, i.e. ensure participation and the possibility for feedback at any time during the process in order to provide for better decision making in the future.

The Open Data domain comprises mainly three actors: data providers, policy decision-makers and the public. A *Data Provider* publishes data, preferably in a raw, unprocessed format, that is accessible to everyone. Note that data providers include private parties as well as public parties, though public or governmental agencies are the most common ones. *Decision Makers* need to push Open Data to increase transparency and to legitimate their proposals and decision. *Active citizens* or the public are furthermore encouraged to participate in the decision processes. They are asked for feedback and thus require a transparent processing and presentation of the data in a way that is commonly comprehensible.

Although, every actor comes with its own perspective and its own requirements or expectations we would like to further detail the public's or citizen's part in the following.

From a citizen's perspective it is often hard to figure out if and where governmental data regarding the topic of interest is publicly available. Consider a person, let's call him Tom, who is moving to a new city and would like to find the best place for renting an apartment. Tom is 30 years old. Since he does not own a car, he wants to live close to the city center to have everything within walking distance. Therefore, he also needs a good connection with the public transportation to get to work. Moreover, he also sets value on living in an area with

mostly younger people and recreational areas. All data, which is needed for such an evaluation, can basically be found on the Internet. However, the data is spread over various data sources and offered in different (proprietary) formats. Raw statistics and customized charts about demographics, infrastructure, public transportation, and geographic points of interest need to be aggregated manually. Hence, an automatic integration and aggregation of Open Data is necessary in order to make all the data more accessible to the public.

The Open Data Workflow

The process of using Open Data resembles a traditional data processing workflow, cf. Figure 1. First, data has to be imported from heterogeneous sources. This may include proprietary file formats as well as already machine-readable data sources. Thereafter, the data has to be integrated and filtered so that further processing is possible with rather generic approaches. Second, the so-called business logic will take place. This includes normalizations, aggregations and further data enrichment. The final goal then is to present the data and its clues to the user or audience in an appropriate way. For example, this can be achieved with powerful visualizations or the like.

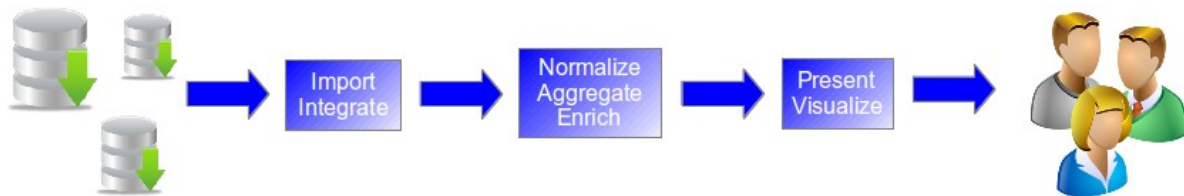


Figure 1: The Open Data Workflow

LISA: Aggregating and Presenting Open Data

LISA (Local information, search and aggregation) is an award-winning web application from the WeST institute (see <http://lisa.west.uni-koblenz.de> for further information) providing local search that can help policy decision-makers and citizens to find relevant areas in a city for a particular purpose. In the case of decision-makers this might be used for urban development whereas for citizens it might be useful for moving or real estate purchase.

LISA extracts data from several data sources like statistics of population, households and real estate prices, information about public transportation, and points of interests. All data are freely available as Open Data. As the data is provided in several data formats, the data import also includes a data integration to obtain data in a common format. Depending on the individual search and information request, the corresponding data is extracted from the data sources and aggregated according to the user preferences. The aggregation includes a normalization since some information is given as relative numbers, while other is given as absolute numbers. The aggregated data indicate the relevance of particular areas and points of a city, e.g., the higher the aggregated value of an area the more relevant is this area for the user. Users of the application determine the relevance by individual and weighted selection criteria for various features. Finally, this relevance is illustratively presented to the user, e.g., by heat-maps or diagrams.

Towards an Open Data Middleware

Current research in the field of using Open Data in eGovernment and eParticipation at WeST is concerned with the development of an Open Data middleware that will allow a simpler development of Open Data applications. Most of the data published by governmental bodies is given in propriety formats such as Excel sheets, text files, or even PDF files. In order to make use of this data a cumbersome pre-processing step is necessary. Our Open Data middleware aims at alleviating the task of integrating sources of different origin and format by providing a common methodology for data import, data aggregation, and data visualization.

Related Projects

The Institute for Web Science and Technologies (WeST) actively performs research in several fields related to the World Wide Web such as Semantic Web, Social Web, Information Retrieval, and, in particular, the use of Linked Open Data and public governmental data for eGovernment and eParticipation. In the following we present two examples of ongoing projects with topics related to the call of the workshop. See <http://west.uni-koblenz.de> for more information.

EU Project “WeGov”

The EU project WeGov (started in January 2010) aims at developing a toolset that allows full advantage to be taken of a wide range of existing and well established social networking sites (Facebook, Twitter, Bebo, WordPress etc.) to engage citizens in two-way dialogs as part of governance and policy-making processes. The tools will make it possible to detect, track and mine opinions and discussions on policy oriented topics. For more information we refer the interested reader to the project website at <http://www.wegov-project.eu/>.

EU Project “Live+Gov”

The EU project “Live+Gov” (started in February 2012) aims at developing mobile solutions that allow citizens to better interact with local administration and government institutions. By using modern mobile sensing technologies citizens can passively and actively express their needs to public bodies. For example, one of the use cases of “Live+Gov” is concerned with urban maintenance where citizen actively report issues (road damages, etc.) to the administration, see <http://liveandgov.eu> for more information.

Summary

WeST is an active research institute with a special focus on using Open Data for eGovernment and eParticipation. We are engaged in several projects on topics related to the call of the workshop and we are confident that we will make a valuable contribution to its success.