An Open Data Portal for Berlin

Submitted on 16 Aug 2011 by

Under the direction of Dr. Wolfgang Both from the Senate of Berlin for Economy, Technology and Women´s Issues, Fraunhofer FOKUS Berlin and the BerlinOnline company work on an implementation of an Open Data Portal for Berlin. The project is headed by Ina Schieferdecker, Fraunhofer FOKUS.

In a preliminary study FOKUS identified the concepts, the architecture and available data sets for a "City Data Cloud". Such a city data cloud would offer public, commercial and private data sets in machine-processable formats, provide tooling to search, filter, transform and visualize them, and selected applications for advanced services. Build on that preliminary study, a more detailed analysis of the situation, driving forces and potentials in Berlin started. Beside other things, the Berlin Open Data Agenda defines six basic principles for Open Data in Berlin. Later on, a realization of a Berlin Open Data Portal was initiated.

For the Berlin Open Data Portal, mainly the processes, technologies and legislation of open data in Berlin are elaborated.

Open Data Processes and Data Sets

The analysis of established processes and data sets in the Senate of Berlin and its institutions has shown that in particular data sets offered by the State Statistical Institute Berlin-Brandenburg fulfill already various open data criteria: they are machine readable, (semi-) structured and use a licensing scheme close to the CC-BY-3.0 license. Therefore, an initial collection of data sets has been chosen for the portal pilot including data sets in Microsoft Excelon e.g. elections, environment, demography, economy, labour, education, etc.

The analysis is deepened now with interviews in various departments to find out and describe official data procedures. This will lead to an Open Data Guide for the administration level.

Technology

The technical implementation of the data portal is developed together with the BerlinOnline company, running the official city website www.berlin.de. The combination of CKAN for maintaining the data catalogue and Drupal for the portal frontend was considered as best practice similar to other Open Data portals launched in Vienna and London. The content management system Drupal is user-friendly for both end-users and editors. Moreover, the data hub CKAN is already established worldwide in the open data community. It guarantees an easy share, use and networking between Open Data Platforms across the world: from a single region like in Berlin up to European platforms like publicdata.eu.

An open point is the definition and commitment of data formats. The standardization of public data is still under discussion.

Legislation

For legislation, adequate licenses and sufficient data privacy protection levels have been considered. For the portal, the licenses for the data sets and for the collection of the data sets on the portal are differentiated. More and more open data platforms are using the CC–BY–3.0 for their data. The CC–BY–3.0 license offers both a maximum of legal certainty and a high applicability for (private, institutional, and commercial) users: they are allowed to copy, share and re-use the data with the

condition of quoting the authority. This satisfies administrative requirements and guarantees the quality of services and applications being built on the data. For the data set collection of the portal, the „Open Data Commons Attribution" license (ODC-BY)is being considered. In order to prevent the publication of personal data, principles and guidelines for the authorities are to be developed too.

The Berlin Open Data Portal will be launched as a draft pilot in September at http://daten.berlin.de. Please stay tuned for further updates.