

Mapping the landscape of Open Geodata

Matthias Hinz
University of Rostock
Justus-v.-Liebig-Weg 6
Rostock, Germany
matthias.hinz@uni-
rostock.de

Ralf Bill
University of Rostock
Justus-v.-Liebig-Weg 6
Rostock, Germany
ralf.bill@uni-rostock.de

Abstract

The search for Open Geodata can be a complex task. Although the amount of data available on the web is overwhelming, users may find it difficult to figure out from where to start. Search engines are no appropriate entry points for systematic searches because they are made for discovering websites of all kind, and they are not specialized in data or data catalogs. There is a need for a roadmap of the Open Geodata landscape that gives directions towards useful data sources. For this purpose, members of the project OpenGeoEdu created a portal of Open Data portals which serves as a central entry point for discovering Open Geodata from the German-speaking area of Europe.

Keywords: Open Geodata, Open Government Data, Spatial Data Infrastructures, data portal, Open Source, web-mapping

1 Introduction

At present, there is a large amount of data available, whether they are Open Data, which are usually cost-free, or whether they are official and commercial data, which are often subject to usage restrictions and not free of charge. All these data are of great interest, as citizens, administrative bodies, science and the economy could benefit from them, but data sources are heterogeneous in their cost and license models as well as their terms of use. While individual data sources are well accessible and business models, as well as usage models, can be developed under clear legal conditions, there are many technical and organizational barriers to using and interlinking datasets on a large scale.

Members of the project OpenGeoEdu (Open Data for education and science in geospatial study programs) develop a comprehensive registry of known data sources from German-speaking countries, which lists relevant websites and platforms and allows their access and discovery using an interactive web map and data table. This portal is online available¹ and will be in focus of this paper.

Although there are existing portals that compile national, regional, and municipal data sources to comprehensive catalogs of Open Government Data (OGD), the manifold possibilities of data discovery go far beyond: The majority of municipalities and regional administrative bodies operate own data- and geoportals, which are not always embedded in the national catalogs but nevertheless should be considered when searching data. Apart from OGD and transparency portals, the websites of statistical offices, environmental agencies, as well as geoportals provide access to public data and services, which, although the license terms not always equal to Open Data, permit cost-free use to a certain extent. Further

resources that must be taken into account are data published by companies and researchers, as well as volunteered data from associations and citizens' initiatives (i.e. Citizen Science projects).

The aim of this work is to foster the efficient use of Open Data, in particular for the domain of Geographic information by providing a one-stop portal for data search. It demonstrates how to register, describe, and visualize all the above-mentioned data sources in a systematic manner and how this approach is used for developing a web portal that serves as a central entry point for the discovery of data portals within the German-speaking area of Europe. This portal of portals offers extensive search, filter, and visualization functions, detailed metadata and editing functionality, by means of which it is to advance to the One-Stop-Portal for Open Geodata in Germany, Austria, and Switzerland.

2 Surveys and catalogs of Open Data Portals

Comparable surveys, lists, and catalogs already exist on an international level: The Company OpenDataSoft published a list of more than 2600 Open Data Portals and displayed it on a Leaflet-based web map². The non-profit organization CTIC runs a similar worldwide platform called the *Public Dataset Catalogs Facets Browser*³, which implements the principles of Linked Open Data (LOD) and Semantic Web Technologies, as it is based on a data catalog in RDF and a SPARQL endpoint.

Although these projects are very ambitious, monitoring worldwide catalogs of Open Data portals require a large community of contributors or extensive administrative support

¹ <https://portal.opengeoedu.de/>

² <https://opendatainception.io/>

³ <http://datos.fundacionctic.org/sandbox/catalog/faceted/>

in order to keep these lists detailed, complete, and up-to-date. They are giving overviews of the major Open Data portals per country, but users may require more comprehensive information when searching for particular data.

International and nation-wide Open Government portals are publically funded since many years: Since 2015, the European Union (EU) finances the European Data Portal⁴ (EPSI platform, 2016), which currently links to more than 800.000 datasets across more than 70 catalogs from the European public sector and Spatial Data Infrastructures (SDI). Linked catalogs include the major national portals from Germany (GovData, GDI-DE), Austria (data.gv.at, Open Data Portal Austria, INSPIRE Portal Austria) and Switzerland (opendata.swiss). Besides of category- and text-based search, a map-based search where datasets can be filtered by selecting a rectangular area (bounding box) is implemented by the European Data Portal and GovData.

For Germany, the first nation-wide Open Data Portal originates from community efforts of citizens in 2010⁵. Around this community, the Open Data Monitor project created the first comprehensive Open Data map of the country (Kubicek & Lippa, 2015a, 2015b)⁶. Contrary to the governmental platforms, where national, regional, and municipal bodies themselves are responsible for registering their Open Data, Kubicek and Lippa favored combined approaches of pro-active searching for datasets to be included (e.g. using web-crawlers). They concluded, however, that this resulted in a great amount of manual work, despite their efforts of automatization.

3 A portal of Open Data portals

3.1 Motivation and concept











The authors of this work saw a need for a web portal that is specific to the Open Geodata landscape of the German-speaking area of Europe, which is more detailed and comprehensive than the aforementioned entry pages, and yet could be linked to larger-scaled overviews. Unlike the other entry pages, the portal herein is not limited to websites titled Open Data or transparency portals, but also related portals, like entry points of Spatial Data Infrastructures, statistics portals, environmental and scientific data portals, and websites of Citizen Science projects. For the countries taken into account, it can be said that there is a tendency of these websites to offer cost-free or affordable data and services, hence they should be taken into account when searching for geographic information on the web. Records of data portals are categorized into these types of portals and the (spatial) range of data offered, as shown in Table 1.

A database entry basically consists of an ID, portal title, range, type, country, the name of the place a portal refers to (e.g. a country, region, or municipality), and the coordinates of the portal, given in longitude/latitude pairs. Data portals can be discovered by two interlinked views: the map view that

is displayed in Figure 1, and the data table view displayed in Figure 3.

The portal is an open resource, made for the community of geospatial professionals, students, and interested non-professionals. Therefore, anyone can obtain the data as well as the source code under Open Data and Open Source license. Likewise, the portal allows any user to correct or to add new entries to the list.

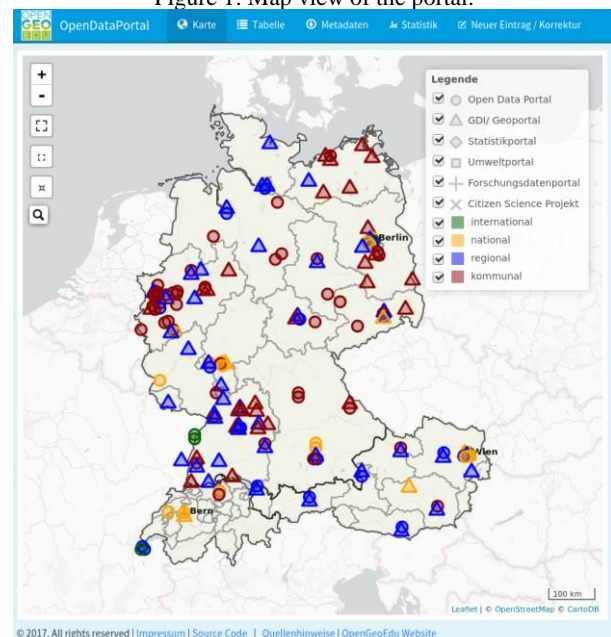
Table 1: Categories and symbols of the portal register.

portal type		range	
	Open Data portal		international
	SDI/ geoportal		national
	statistics portal		regional
	environmental data portal		municipal
	scientific data portal		
	Citizen Science project		

3.2 Features of the portal

Data portals on the map are represented by markers, generally positioned near the given address of the responsible publisher. Color and shape of these markers are determined by how the portal is classified. For instance, a geoportal operated by a municipality or city is displayed as a red triangle. Using the map view, users can easily determine data portals in relation to an area or location of interest. When a marker is selected, the title, URL, and description to the corresponding portal is displayed in a pop-up window.

Figure 1: Map view of the portal.



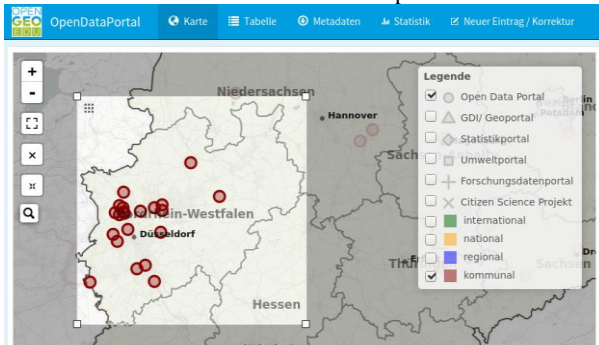
⁴ <https://www.europeandataportal.eu>

⁵ <https://offenedaten.de/>

⁶ <http://open-data-map.de>

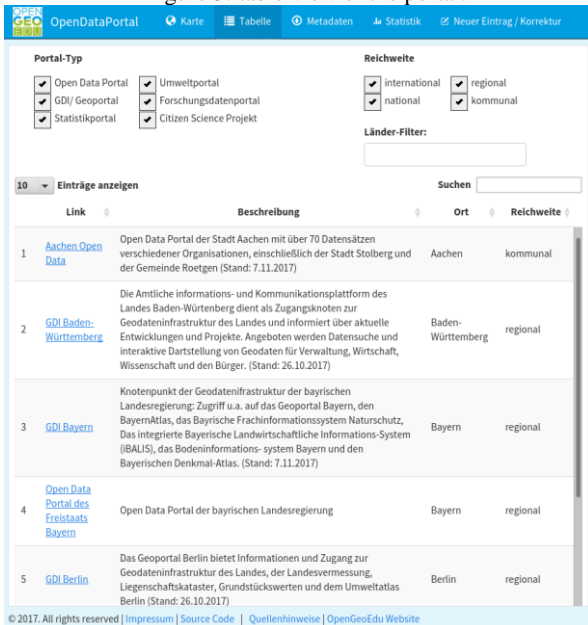
Besides common map-operations (zoom, pan, full-screen, reset view), it is possible to search individual portals by their title and place, and to filter them by attributes. Filters simultaneously take effect on both, the map view and the table view, and they can be applied for portal type, range, and location. The latter filter is applied by drawing a rectangular selection (bounding box). In this way, it is possible to filter, and subsequently list, for instance, only the municipal (range) Open Data portals (portal type) located in the German federal state North Rhine-Westphalia, as shown in Figure 2.

Figure 2: Selection and filtering of regional Open Data portals within North Rhine-Westphalia.



The table view, shown in Figure 3, offers the same category-based filters as the map view, and they are also synchronized with the map. In addition, it is possible to filter the entries per country and to apply a free-text search for filtering table entries. The table can be flexibly ordered by each column (ID, URL/Title, description, place, and range). If users highlight entries inside the table, the corresponding markers are highlighted on the map as well.

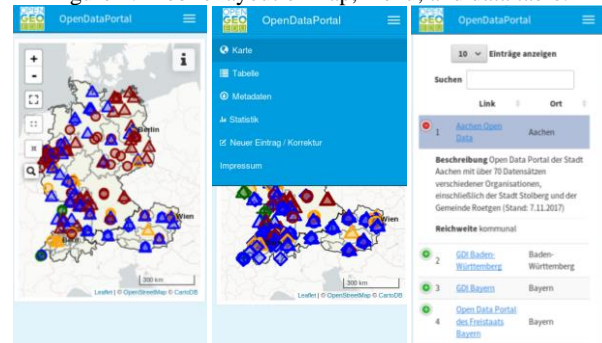
Figure 3: table view of the portal.



Besides the map and table view, the portal is divided into the sections *metadata*, *statistics*, and *new entry/correction*. The metadata section explains the underlying data portal registry with its fields and provides download links for different formats of the data. The *statistics* section gives a summary statistics of the registry data and how the registered portals are distributed in respect to the countries, portal types, and range-category. In April 2018, while it is still enhanced, the registry comprises 262 entries. The last section of the portal provides a web form where users can propose new entries or make corrections on existing ones. Edits are merged to the data after being reviewed by an admin.

The portal is composed of responsive layout elements. As shown in Figure 4, the layout adapts to small screens in a way that its clarity is preserved by hiding elements that take up space (i.e. the menu entries, map legend, and individual table columns), while the complete functionality and metadata remain accessible through hide/show buttons.

Figure 4: Mobile layout of map, menu, and data table.



3.3 Visualisation and map components

The basemap of the application is currently the black and white map *Carto positron* without labels⁷, based on OpenStreetMap-tiles (Openstreetmap Contributors, 2018). This plain map was chosen to give the user just enough context to identify the approximate location of a data portal's scope and spatial belonging while the map remains clearly arranged. On top of the map, administrative boundaries from OpenStreetMap are included as vector data. German place-names of larger cities were included as a separate vector layer based on the GeoNames database (GeoNames, 2018) because there was no suitable basemap service available that uses the German language consistently. Labels and boundaries of lower administrative levels are blended in and out depending on the zoom level.

A particularly challenging visualization problem of marker-based maps can be the occurrence of many markers that are close together or share the same location (e.g. one municipality operates multiple data portals). In the first case, markers are overlapping when zooming out, and in the latter case, markers are inaccessible because they are on top of each other. Existing solutions are zoom-level-dependent clustering of markers, as implemented by the *Marker clustering plugin*

⁷ <https://carto.com/location-data-services/basemaps/>

for Leaflet (Agafonkin, 2018)⁸, statically associating one marker with multiple entries, or spiderfying markers at identical locations so that they are splitting up when the user selects them, which was also implemented by a Leaflet extension⁹. For the current implementation of the web portal, it was chosen to re-arrange markers instead of grouping or clustering them: A simple algorithm detects markers which are close together and calculates new coordinates in a circular radius around the common center using trigonometric equations. This approach could be further enhanced by applying the rearrangement on different zoom-levels and by scaling up or down the size of marker-symbols depending on the zoom level. This visualization technique was applied under the assumption that an accuracy of marker-locations for city- or municipality level is sufficient.

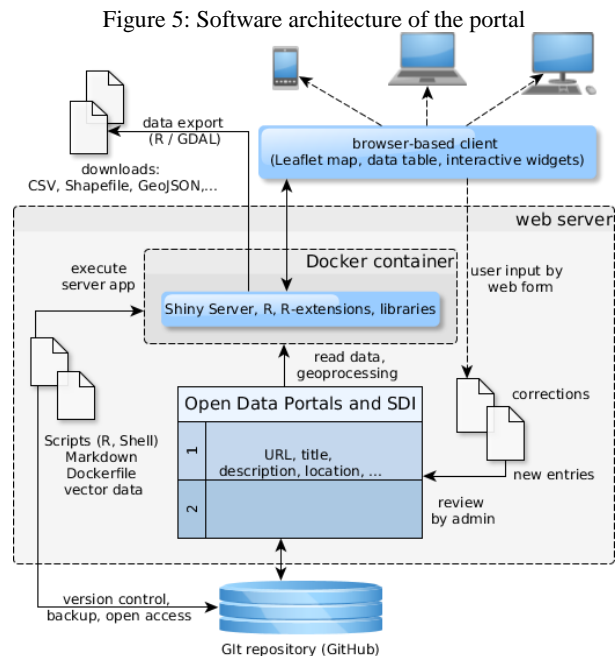
3.4 System components and architecture

The portal is released as a web application based on a client/server architecture and open-source components. Its overall architecture is outlined in Figure 5. The browser-based client incorporates the JavaScript library Leaflet for the interactive map (Agafonkin, 2018). The table view is based on the DataTables (SpryMedia, 2018) library with its extensions for responsiveness and internationalization (i.e. in German language). The R programming language and environment (R Core Team, 2017) was used for implementing the server logic, the data processing from a raw CSV table to the visualized data, and the overall client layout. Thanks to a large developer and user community and many extensions, R became a universal language that can be applied, amongst others, to geospatial analysis (Bivand, Pebesma & Gómez-Rubio, 2013). A number of R extensions used by the web portal facilitate web-mapping – namely the Leaflet extension (Cheng, Karambelkar & Xie, 2017) and Crosstalk (Cheng, 2017). While the former is an interface from Leaflet to R, the latter allows the implementation of cross-widget filters as described in section 3.2, where the map and data view are interlinked. The Flexdashboard extension (Borges & Allaire, 2017) and R Markdown form the basis of the client's layout, whereas a text document with markdown annotations and chunks of R code is transformed to an HTML document with flexible and responsive layout components.

Shiny (Chang *et al.*, 2017) is a web application framework for R that provides a web server and a set of reactive widgets that allow R code to be executed during application runtime. Using Shiny Server and widgets, it was possible to implement the dynamic behavior of the portal and input/output components such as the web form for submitting and saving new entries and corrections.

Using the Geospatial Data Abstraction Library (GDAL) and its corresponding R extension (Bivand, Keitt & Rowlingson, 2017; GDAL Development Team, 2017), the data portals registry is converted to popular geodata formats (GeoJSON, ESRI Shapefile, GeoPackage, GML, and KML) and shared on the portal's download section.

Arranging the server-side application inside a Docker container (Merkel, 2014) simplifies the overall setup. In this way, the software environment can be reproduced on any machine that has the Docker-environment installed. For external backup, version control, and sharing of the source code (in the sense of Open Access), resources of the portal are publically available on a Git-repository¹⁰.



4 Discussion

Crucial the success of the portal is that it will be adopted by a community of geospatial students and professionals in Germany, Austria, and Switzerland. The platform addresses their needs in terms of content, functionality, and German language. The overview of data portals is supported by extensive visualization-, search-, and filter functions as well as a mobile-friendly layout. However, further adjustments in terms of usability and functionality may be necessary depending on user feedback. It is planned that the OpenGeoEdu project will continue beyond the current funding period. Hence it is likely that maintenance will be sustained by follow-up-activities. It is important that the platform content remains up-to-date and comprehensive. This meta-information is largely based on research team's investigations and might be enhanced by voluntary contributors. For minimizing the regular workload, the number of metadata-fields is kept to an expressive, small selection that users can update using a web form.

Descriptions of the portals cannot be retrieved automatically because a majority of the indexed websites does not provide content summaries in a concise and machine-readable form. For instance, extracting the metadata fields from the HTML head elements was not leading to any useful results. The

⁸ <https://github.com/Leaflet/Leaflet.markercluster>

⁹ <https://github.com/jawj/OverlappingMarkerSpiderfier-Leaflet>

¹⁰ <https://github.com/opengeoedu>

quantity and technical heterogeneity of data portals complicate forms of automated monitoring or interlinking of the data portals, e.g. for cross-platform search. Such implementations, aside from simple availability checks, are currently not within the scope of the project but might be achieved in conjunction with further works dedicated to this problem. A possible solution could involve linking the platform with structured meta-information provided as Linked Open Data.

5 Conclusion

The web portal of Open Data portals is a WebGIS-tool for exploring sources of Open Data and Open Geodata that provides transparency and orientation of the current offerings in this field. The portal will continue to evolve as it is used for teaching the usage of Open Geodata in the context of OpenGeoEdu. Both, the web platform and the registry of data portals are open resources, which means they can be re-used and extended by others, e.g. for developing similar portals in other countries or for linking the data in a different context.

It is not sufficient for Open Data just to be published on the web. There is also a need for transparency of how and where they are published. The given tool addresses this need by allowing the systematic discovery of data sources.

References

- Agafonkin, V. (2018) *Leaflet: An open-source JavaScript library for mobilefriendly interactive maps*. Available from: <http://leafletjs.com/> [Accessed 31st January 2018].
- Bivand, R., Keitt, T. & Rowlingson, B. (2017) *rgdal: Bindings for the 'Geospatial' Data Abstraction Library*. Available from: <https://CRAN.R-project.org/package=rgdal>.
- Bivand, R. S., Pebesma, E. & Gómez-Rubio, V. (2013) *Applied Spatial Data Analysis with R*. New York, NY, Springer New York.
- Borges, B. & Allaire, J. J. (2017) *flexdashboard: R Markdown Format for Flexible Dashboards*. Available from: <http://rmarkdown.rstudio.com/flexdashboard> [Accessed 22nd December 2017].
- Chang, W., Cheng, J., Allaire, J. J., Xie, Y. & McPherson, J. (2017) *shiny: Web Application Framework for R*. Available from: <https://CRAN.R-project.org/package=shiny> [Accessed 22nd December 2017].
- Cheng, J. (2017) *crosstalk: Inter-Widget Interactivity for HTML Widgets*. Available from: <https://rstudio.github.io/crosstalk/> [Accessed 22nd December 2017].
- Cheng, J., Karambelkar, B. & Xie, Y. (2017) *leaflet: Create Interactive Web Maps with the JavaScript 'Leaflet' Library*. Available from: <http://rstudio.github.io/leaflet/> [Accessed 22nd December 2017].
- EPSI platform (2016) *Understanding the European Data Portal*. Available from: https://www.europeandataportal.eu/sites/default/files/2016_understanding_the_european_data_portal.pdf [Accessed 22nd December 2017].
- GDAL Development Team (2017) *GDAL - Geospatial Data Abstraction Library, Version 2.2.1*. Available from: <http://www.gdal.org>.
- GeoNames (2018) *The GeoNames geographical database*. Available from: <http://www.geonames.org> [Accessed 1st February 2018].
- Kubicek, H. & Lippa, B. (2015a) Open Data: Appelle reichen nicht. *Kommune 21* (5/2015), pp. 12–13. Available from: http://www.kommune21.de/meldung_21279_Appelle+reichen+nicht.html [Accessed 22nd December 2017].
- Kubicek, H. & Lippa, B. (2015b) Open Data: Kommunale Krux. *Kommune 21* (4/2015), pp. 12–13. Available from: http://www.kommune21.de/meldung_21140_Kommunale+Kruux.html [Accessed 22nd December 2017].
- Merkel, D. (2014) Docker: Lightweight Linux Containers for Consistent Development and Deployment. *Linux J.* 2014 (239). Available from: <http://dl.acm.org/citation.cfm?id=2600239.2600241>.
- Openstreetmap Contributors (2018) *OpenStreetMap*. Available from: <https://www.openstreetmap.org> [Accessed 28th January 2018].
- R Core Team (2017) *R: A Language and Environment for Statistical Computing*. Vienna, Austria. Available from: <https://www.R-project.org/>.
- SpryMedia (2018) *DataTables table plug-in for jquery*. Available from: <https://datatables.net> [Accessed 2nd February 2018].