1. References to the documentation environment

We start with a brief overview on the tools the project team to collaborate. Our main platform for working together on the prototype has been GitHub, a web-based service for software development projects that uses the Git revision control system. Occasionally (like for writing this report) we made use of the hbz's Confluence wiki. We used the tool Huboard that is based on the GitHub API to have an overview over the different tasks and their status. Thus, in the course of this report, several references will thus go to GitHub issues and comments where certain aspects are covered in more detail. The OER World Map itself can be found under http://www.oerworldmap.org.

2. Features of the hbz prototype

Due to the short development time for the prototype the project concentrated on the realisation of an operational service, which allows to:

- input data through web forms
- display data on a world map with basic functionalities
- fulltext search with several filters (e.g. for geographic area, country, resource type)

The hbz prototype consists of two applications (Drupal viewing/editing frontend and a JavaScript map) which are themselves based on an application programming interface (API) that enables programmatic interaction with the data. A detailed look at these three parts will follow. But first, the underlying data, its sources and data model will be described.

2.1. Data

2.1.1. Sources

The prototype is mostly based on data from two different sources:

- OpenCourseware Consortium (OCWC) member data: The people at OCWC were very helpful in providing us with and explaining the OCWC member data. The data was obtained via an API, see GitHub issue #3 for details.
- Global list of OER initiatives from UNESCO's WSIS Knowledge Communities: This data can be downloaded (after registering and logging in) as - rather hard to process - comma-separated values (CSV) at http://www.wsis-community.org/pg/directory/export/672996.

Along with this data collected from pre-existing sources, there is also some manually added data.
2.1.2. Data model

We already noted in our proposal that it wasn't clearly defined what kind of resources an OER world map should cover:

> Mentioned in the request for proposals are projects and initiatives as well as people (OER experts). The report summarizing the international conversation that led to this consultation process also mentions journals, repositories and services.

> Of course the information on projects will also contain information on institutions the project is run by. There are some more resources that would be valuable to also have information about, like events (congresses, workshops, hackdays etc.) as well as publications and (software) tools in the OER world or job offers.

Thus, a first and important work on the prototype - as with any software that creates data - was defining the data model for the OER world map data. Of course, this data model should on the one hand be designed with the use cases for the OER world map in mind and on the other hand the data model should take into account the actual information that is provided in the source data from OCWC and WSIS.

Here is what we came up with:

As we are working with linked data, it was clear that internally and for providing the data we would use the Resource Description Framework (RDF). To represent data based on this data model we could almost entirely resort to the schema.org vocabulary which has the advantage that the OER data will also be indexed by search engines like Google and Bing. Only three RDF properties and one class had to be taken from other vocabularies. RDF is very flexible and can be extended very easily. For example we decided at an early stage of the project to include information on services (e.g. OER repositories, search interfaces) in the data. This seemed to make sense as services (e.g. repositories) play a vital role for the OER community and as the world map should be of help in discovering open educational resources. Further extensions and other adjustments will probably turn out necessary within the next phase of the project.

At the end of the project, the data set the prototype is based on includes information on
398 organisations, 
254 persons, 
321 services, 
31 projects.

2.1.3. Application profile

We put some time into developing an application profile (AP) using RDF for the OER world map. This application profile expresses the application's data model and configures how the data will be viewed in the Drupal editing and presentation environment (see below). In the future, it should also be used as a basis for validating the data input.

The concept of an application profile comes from the Dublin Core Metadata Initiative (DCMI). In short, an application profile is a set of metadata elements, policies, and guidelines defined for a particular application. The elements may be from one or more vocabularies, thus allowing a given application to meet its functional requirements by using metadata from several vocabularies. An application profile is not complete without documentation that defines the policies and best practices appropriate to the application.

The application profile allows us to have configuration of the Drupal editing and presentation environment and future API validation in one central place. In order to configure the API validation and web site, changes have to be included into the application profile - all connected forms and presentation sites will automatically change accordingly. The AP is maintained on GitHub and enables relatively easy maintenance of the data presentation and validation without having to directly interact with the front end or API developer. (The application profile, in other words, is the means of unambiguous communication between a metadata expert and the developers). This feature accelerates and cheapens the further development of the OER world map.

2.2. API

A central element of the hbz prototype is an application programming interface (API), which support the easy development of web applications on top of the data. The prototype's API can be found at http://lobid.org/oer. Currently, there exist two clients (which are both integrated under oerworldmap.org) that are based on the API: a view and editing environment built with Drupal and a JavaScript map to interact with the data in a map environment. Multiple other applications could also be built on top of the API.

2.3. Drupal view and editing environment

The content management system (CMS) Drupal is used to implement views and editing capabilities for the data provided by the API. Thus, we do not use the relational database that comes with Drupal. Instead, a so called Entity Type was implemented to read/write from/to the API. Additionally, a custom Entity Field Query was implemented to query the API. To demonstrate the use-case of linking to external data, the GeoNames Search Webservice is also available via this component.

In order to load the RDF data provided by our API and GeoNames into Drupal, the built-in RDF capabilities of Drupal were extended to not only output, but also read data in RDF. The mappings of RDF-properties and classes to Drupal fields and bundles are parsed from the application profile.

As mentioned above, we link to GeoNames for countries and cities to demonstrate the approach of adding additional context to our data. These links contain data such as the population which could be used for further visualizations. Furthermore, although we do not use Drupal's database for our actual data, all other capabilities of the CMS can be used, e.g. to define users and roles in an editing workflow.

2.4. JavaScript Map

The JavaScript map is a separate read-only front-end based on Leaflet. Although it is currently embedded in the Drupal site, it communicates directly with the API to fetch its data. Most filtering (i.e. by type and country) is done within the map's JavaScript at this point of the prototype. As the dataset grows, this filtering can be moved to the API to increase performance. As a pragmatic approach, the map's pop-ups reuse the views provided by Drupal. These views can be replaced by custom renderings in the future. To demonstrate the advantages of linking to external datasets, country labels from the GeoNames data are provided in the language corresponding to the browser’s settings. These multi-lingual labels are not part of the OER world map dataset, but retrieved from GeoNames using the Linked Open Data approach. Any updates to the GeoNames data will thus be reflected in the UI automatically.
2.5. Out of scope

Out of the scope of the project were:

- aspects of the editorial workflow
- organizational aspects
- the business case
- advanced features of the map design.

3. Course of the project

Our initial planning of the project within our proposal turned out to be quite resilient, although it’s strict linear character is misleading, since in fact we worked in several iterations which are difficult to display visually. The following table gives an overview of the course of the project including links to more detailed documentation on GitHub.

<table>
<thead>
<tr>
<th>Week</th>
<th>Milestones</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week 01: 10.02.14 - 16.02.14</td>
<td>• Start definition of application profile (#1) • serendipity and OCWC data received. (#2, #3) • Set up GitHub repository</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• initial version of application profile created</td>
<td>• human-readable description of AP created (#1) • obtained WSIS-Data (#6) • decision to generate a formalized application profile in RDF (see: see <a href="http://wiki.dublincore.org/index.php/OER-world-map">http://wiki.dublincore.org/index.php/OER-world-map</a>)</td>
</tr>
<tr>
<td></td>
<td>OCWC data has been indexed</td>
<td>• Set up Play module for OER API (#9) • Access initial sample OER data set in Elasticsearch (#10) • Implement query API (#11) • Implement write API (#12) • Create formal application profile (#7)</td>
</tr>
<tr>
<td></td>
<td>• initial version of web formulars available</td>
<td>• OCWC data has been indexed • Convert preliminary initial N-Triples to JSON-LD and index in ES/Indexed initial data (OCWC) (#10) • Implement parser to load application profile into Drupal (#4, #15, #18)</td>
</tr>
<tr>
<td></td>
<td>• Initial version of map interface</td>
<td>• Implement GeoNames query for locality/country field in web form (#26) • Add geospatial filter to API (#34) • Improve transformation of OCWC data (#10) • Support JSONP requests in API (#36) • Support any RDF serialization in read API (#48) • Use remote context in JSON-LD served by the API (#44) • Implement basic JavaScript-based map (#39) • Implement search form in Drupal</td>
</tr>
<tr>
<td></td>
<td>• decision to include WSIS data</td>
<td>• Finish work on OCWC data transformation (#5) • Several adjustments to the application profile (#59, #66, #64) • Implement type and country filters in map (#40, #57) • Support pagination on API level (#85) • Support deletion via API and improve API response codes and content types (#51, #65)</td>
</tr>
<tr>
<td></td>
<td>• decision to include WSIS data</td>
<td>• Add links between consortia and their members to the data (#53) • manual editing/refinement of WSIS data (#95)</td>
</tr>
<tr>
<td></td>
<td>• Transform WSIS data on OER initiatives and add it to API (#95) • Index GeoNames data and allow OER API queries by country names in any language (#96)</td>
<td></td>
</tr>
</tbody>
</table>
4. Lessons Learned

The project differed in many aspects from usual hbz projects:

- it was shorter,
- there was no individual customer that could specify on the requirements if needed,
- the topic of the project differed from usual library projects hbz is typically involved with,
- it was the first time we received funding from a non-German institution.

Considering these circumstances, we learned a lot during the project.

4.1. Project Management

- Since there were few predetermined use cases and since we did not focus on developing use cases our self we neglected to work strict according to defined use cases. This lead to some unprecise formulated tickets and probably cost us some time. For future projects we would take care to define use cases carefully and orient ourselves more strictly on these use cases.
- Communication with external stakeholders was very interesting, but also time consuming. We did not plan much time for external communication, although this would have been helpful for the definition of use cases as well as for community building. For the creation of the production system we recommend to make sure that enough time is spend on external project communication, since overall success of the OER world map will depend heavily on community acceptance.
- One consequence of our data-centric approach was, that large parts of our development time was spent on the backend of the system. As a result there was little time left for the refinement of the frontend. If we would have to do it again, we would try to plan more time for additional development iterations of the front end.
- Not totally unexpected, cooperativeness from the community was very high.
- Although there is a high awareness of the need for content licenses within the OER community, awareness of the need of licenses for data is less developed. Neither the integrated OCWC data nor the WSIS data was licensed and even we ourselves became aware of this question very late. We recommend to use the CC0 Public Domain Dedication by Creative Commons for all data connected with OER projects.
- In the beginning of the project, we followed the our standard workflow for quality control and deploying changes on the production system that was established before by the lobid team. Though this approach is appropriate and needed for a production environment it became clear soon that it isn’t the best approach for developing a prototype in 10 weeks. Thus, we decided to switch to a simpler and quicker workflow. It took us until the end of the project to get accustomed to it. Lesson learned: Next time we should agree on a clearly defined lightweight workflow for quality control and deploying code changes at the beginning of the project.
- The due diligence process of the Hewlett Foundation took longer than we expected, which dampened a bit the positive energy that was generated by the initial acceptance of our proposal. Also this took one week of the planned project time. Although it was no problem for us to keep up with the project plan we would propose to make the assignment process more transparent.

4.2. Design

- The project confirmed our initial assumption that an API-centric approach is preferrable to other approaches. It even led us to the insight that it is advisable to conceptually distinguish between two central parts of the system, the data base and the visualisation in form of a map. In the following we will use the terms "OER data hub" for the backend and "OER World Map" for the frontend. Nevertheless our API development was dominated by the needs of the world map, since there were no other applications naming demands to the design of the API. Therefore our goal of building an API, which provides a more general abstraction level on the data has not yet been fully achieved.
- The project also affirmed our decision in favour of the use of LOD technology. The data included in the OER world map describes basic building blocks the OER community consists of and there should be many useful ways of reusing them (for examples see below, 5.2.2). Thus, the data should be published in a way that assures maximum reusability which we tried to accomplish. Especially with the goal of building an international multilingual service, linked data and the configuration of the data presentation via a central application profile prove to be an ideal approach. The prototype gives a first taste of the advantages. E.g. if you open the http://oerworldmap.org page, country labels to use as filters on the map will appear in the language you have selected in the browser setting as your preferred language. If I chose German, I will see German labels, if I chose English, I will see English ones. This is possible because GeoNames Search Webservice - the service the information on countries and cities is linked to - provides the labels of cities and countries in multiple different languages. This made it also very easy to enable search for persons and organisations in one country by using the language of your choice. E.g. if I am from Poland and want to find out what OER initiatives happen in Poland, it might be the case that I use my primary language - Polish - to do the query. Based on the GeoNames data the API will deliver the same results for ‘Polska’ as if you used ‘Poland’.
- One important function of the OER world map is to give an impressive image of the OER movement at one sight. Such a picture can be used to argue for OER in political and other contexts. In order to do so, questions of aesthetics should not be underestimated. For the development within phase II we would consider cooperating with a specialized design partner to make sure a "high gloss finish" of the world map visualisation is achieved.

4.3. Technical

- For some time now we use at hbz the very powerful open-source software Metafacture for transforming large amounts of library data to linked data. During this project we learned that Metafacture is too heavyweight a software to be used on small data sets that are not that complex as the average library data set. We made some very good experiences with pre-processing the data with open refine, though. In the future we will further explore using open refine for the complete transformation process to RDF. This makes also sense since the to be transformed data will only be initially transformed and so the need to have an automatically transformation (which would be hard to do with Open Refine, if at all) to work on updates does not arise at all. So the use of metafacture is a bit over-engineered.
• A - for us - novel approach for dealing with blank nodes in RDF took a lot of resources. We decided not to have any blank nodes at all, but identify them using a UUID. To reflect this in the data transformation and keep track of the assigned IDs during several consecutive transformation processes we had to invest a lot of time and resources.
• When integrating complete additional datasets there needs to be some editorial quality assurance of the data, which can be quite time intensive. Refinement and transformation of the WSIS Data took us four days. Therefore we would recommend to do at least parts of the tests automatically. In the future, we will investigate ways to automatically validate transformed data against the application profile.
• Transforming the WSIS data which was obtained in a unfavorable form of csv (comma-separated values) was a very time-consuming job. Transforming the OCWC data also took some time but was a lot easier. It confirmed us once more of the necessity to increase the effort in producing data, e.g. by creating linked data, and to minimize the post-processing effort by third if one wants to maximize reuse of a data set.

5. Recommendations for further development of an OER World Map

5.1. Refinement of the hbz prototype

If the productive OER world map service will be developed based on the prototype created in this project, there need to be several refinements done in order to achieve a reliable production environment.

5.1.1. Refine data model and application profile

One point is the data model and application profile. A wider discussion on how it should look like to enable enough people to add and maintain the OER world map data seems necessary. We have identified some points that should be discussed:

• Types of resources described: It might make sense to add more resource types to be described than the existing four ("organisation", "person", "service", "project"). We could think of additional types like "publication" (for adding OER-related publications in general and specifically publications about a registered project or service), "software", "event", "OER policy" etc. As usual, one should keep in mind that adding new resource types results in a more complex data model and, thus, a more sophisticated editing interface.
• Add more options for classifying/subject indexing: Especially for services it makes sense to add information on language of the provided material, target group, subject etc. Regarding persons, one could add information on languages spoken or the type of expertise.
• Mark discontinued services: In order to be a useful resource, it should be possible to only search for services that are actually running. In another context, it might make sense to query for all services that have shut down in 2013. Thus, adding startDate/endDate fields or a operating /discontinued status field for services makes a lot of sense.
• Multilingual data input: Enable storing descriptions etc. in different languages and indicating these. This would make it possible to provide the information in the language a web browser asks for or would make it possible for linguistic communities to register their projects in their own languages in the OER world map.
• Link projects to organisations they are run by: Currently, you can only link projects to persons working on the project but not directly to organisations a project is run within. It probably makes sense to add this possibility especially as projects themselves don't have any geo or address information and thus can't be located on a map.
• Add field "P.O. Box: Currently, you can only add a street address but no P.O. box information (although some P.O. box numbers a stored in the streetAddress field).

5.1.2. Data enrichment

Already for the existing, quite simple data model, there is some information missing in the data because it didn't exist in the core data and/or we didn't have the time to work on getting the information we need out of the data. For example, though we spent some time on the transformation of the WSIS data on OER initiatives, we only pulled out geographic information for organizations but decided not to add the same information to the associated persons. In the future, it would make sense to add the city and country a person is based in either half-automatically or manually.

5.1.3. Automatic data validation based on application profile

In the prototype, data input isn't validated neither on the client nor the API level. To be sure to have a consistent and, thus, maximally useful data set, it will be important to add validation of incoming data on the API level as well as when indexing transformed data from other data sources (like OCWC or WSIS). It is highly desirable to add an automatic method of validating data based on the application profile so that the application profile would serve as a central standard to decide whether data can be added to the data or must be adapted before indexing. As already noted above, the transformation work would very much benefit from such a process because right now transformation is checked against test files which have to be adjusted seperately when the application profile is changed.

5.1.4. Separate application profiles for validation and presentation

Currently, we have the information which could be used for validating the data input (e.g.: “What kind of strings are allowed in the ‘email’ field etc.?“) in the future alongside with information for presenting the data (e.g.: “In which order should fields be shown and with which labels?“) stored in one application profile. It is highly desirable to have separate documents for these use cases.

5.1.5. Add provenance, administrative metadata and versioning

Currently, the API only holds the actual data about the different resources. There is no information about where the data comes from, how it was transformed and who did this. Also, there is no information available about when a resource description was added or when it was last modified. In other words, there is no provenance data or other administrative metadata available. As this data is important to assess whether a data entry is valid and up-to-date, it should be added if a productive service will be developed based on the prototype. Also, it would be really useful to roll back changes made to resource descriptions. Thus, versioning of the records would be desirable.

5.1.6. Improvements in resource presentation & web form
This paragraph deals with the presentation of information on a resource as well as with the web form for editing this information.

Here is a screenshot of an example of how the information about a resource (here: the organisation "Universidad de Granada") is displayed in the prototype:

We are already quite content with the HTML representations of the information about the different resources (organisation, person, service, project). There are some things that should be adjusted, though, in a production service. And it makes sense in general to experiment with different approaches of presenting the data.

One action item would be to replace the extensive information about a linked resource (in the example for instance the information about "UNIVERSIA" where the organisation is member of) that is shown when clicking on them by less information with the possibility to get more on the respective page of the linked resource. The current approach is especially problematic when the linked resource holds more information than the primary resource.

A desirable and easy to implement feature is to show a organization in a small map instead of indicating the geo coordinates (see current view below).

Currently, the web form reflects quite clearly the data model as defined in the application profile which results in nested boxes that might be difficult to understand by editors. A good example ist the box to add or edit an address:
If the production system is developed based on this prototype, one will have to experiment with other ways of presenting the web forms. For a small group of committed editors, the current web form might be sufficient, though.

5.1.7. Improve world map presentation and interaction possibilities

This paragraph deals with the presentation of the actual map and the possibilities for interaction it provides. Right now we see several options to improve the design of the integrated world map:

- Fix the basic layer of the world map: Right now one can switch the map to the left and the right and another world map appears, without any pins. This affects usability and bears the risk that a user gets lost on the screen. Therefore the basic map layer should be fixed.
- Drop down lists for filter options: Right now filtering is done by rather large lists of checkboxes. Instead of this, drop down menus should be used to attain a more elegant and lean look and feel of the site.
- Within the checkbox lists (respectively the drop down menus) items must be listed alphabetically.
- The search functionality should also be included in the map, so that it is possible to perform a query without leaving the map.
- Right now information about resources are displayed in Drupal default pop-up fields, which are neither very good looking nor easy to use. One solution to improve the display would be to open a field at the left or right margin of the map by clicking on a pin, which shows the relevant information. Within the remaining space the map should be displayed with the activated pin in its center.
- Experiment with different pin forms: A future problem that might occur is that the map will look too crowded. Therefore it might be reasonable to use less dominant pin forms, e.g. in the form of needles.

5.2. General recommendations for collaborative advancement of the project

5.2.1. Discuss editorial process

One important question, which was not included in the scope of our project is the design of the editorial processes for the OER world map. Generally we would argue that as much effort as possible should be carried by the community in order to save costs. Nevertheless it has to be understood that more detailed and sophisticated data models inevitably require more time and understanding on the side of the responsible editor. Without experience and the necessary understanding it can be difficult to distinguish between an organisation, its services and its projects. For example we found that WSIS classified some entries as "communities". We transformed them into services which are run by an organisation. Sometimes these differentiations can be hard to decide.

Counteractions to keep community participation high could be:

1. the development of intuitive smart templates, including intelligent autocompletion and validation mechanism.
2. a short and easy to understand explanation of the structure of the used data model, preferably in the form of a short video
3. generation of high additional value of the map for the user. For example we assume that an OER provider will be willing to spend more time on data submission, if this will lead to proven increase of usage of his materials.

But even if motivation of the community to participate is high we would recommend to integrate some kind of editorial quality control in order to avoid data inconsistencies and to make sure that data is updated regularly. Since we consider it rather unlikely that there will be one organization which commits itself to editing the complete world map data (unless it is paid for this), a decentralized solution probably will be favourable. It would be very helpful to convince organizations which use parts of the data for their own purposes (like the UNESCO/WSIS Knowledge Community or the OCWC) to use the OER World Map platform to collect and edit their data. Additionally it might be suitable to appoint national responsibilities in the way that a group of volunteers takes over the responsibility for editing the data of a country.
In each case the platform should be developed to support these processes by defining different editorial rights for different users. It could also be helpful to include an alert service which makes it possible for users to report outdated data to the editorial team with one click.

5.2.2. Maximize usage and ensure sustainability

How to attain sustainability seems to be one of the most important questions of the upcoming phase II of the World Map project. Since it will be difficult to define a business model, especially in the short run, it would be very helpful if the funding of phase II would be extended, so that there will be possibilities to run and refine the system for 2-3 years after its initial development.

Avoiding redundant collection and editing of the data in different projects would also be very helpful in order to reduce overall costs for the OER community. Therefore potentially existing possibilities of cooperation should be investigated carefully. Ideally initiatives like the OCWC, UNESCO/WSIS Knowledge Community, or the OER Research Hub should bundle their resources and use the OER World Map respectively the OER data hub as a common platform. However our experience shows that such cooperations are difficult to achieve, since there are often slight differences within the needed data models. Since the problems that result from different data models could be largely compensated by the LOD approach and since the OER community emphasises cooperation it might be nevertheless possible in this case.

Apart from these provisions we would recommend to maximise the additional value of the data to the OER community. Once there is value, it is easier to generate revenue. One simple way to do so would be to link the data to other relevant data sets, as has been approved by Tim Berner Lee

6. Conclusion

The delivered prototype demonstrates that hbz offers a platform for developing a scalable production system of the OER World Map which provides maximum data connectivity and reusability. In doing so, it combines elements of open source, open data and open educational resources. It is important to distinguish between the OER World Map as the front end of the system and the OER data hub as its backend. Although the actual focus lies on the frontend of the system, we expect that in the future there could be many other applications which could be developed using the data of the OER data hub. The OER World Map data, especially the institutions model an important backbone of the OER ecosystem. Extending this model by linking it to other resources will maximize the added value of the data, which increases the chance that a OER World Map production system will be sustainable.

Endnotes

RefNotes


5 Cf. the DCMI Glossary. Recently, a “RDF Application Profile” activity within DCMI was started. The experiences made in the development of the OER world map prototype may serve as valuable input in developing a standard for representing an application profile in RDF.