OR2018 Presentation Submission
13th International Open Repositories Conference, June 4th-7th, Bozeman, Montana, USA

Qvain – a Generic Research Dataset Metadata Editor

Esa-Pekka Keskitalo, The National Library of Finland, esa-pekka.keskitalo@helsinki.fi

Wouter van Hemel, The National Library of Finalnd, wouter.van.hemel@helsinki.fi

Session Type

- Presentation

Abstract

Qvain is a generic research dataset metadata editor. It takes a JSON data model and turns it into a form where metadata can be added or edited. It provides tools for the administrator to manage the look and feel: the order and grouping of fields, the help texts not included in the data model, etc. Also, it makes it possible to add a widget to enhance the way data is entered. Qvain can support more than one data model, provided that the administrator makes the necessary adjustments to how the model is displayed.

Qvain works as a standalone tool. But, first and foremost, we are developing it as a component in a cluster of IT services for researchers. The metadata records will be transferred to a shared Metadata Repository, where other systems can access them.

Being generic always comes with drawbacks. But there is more need for describing datasets than there are discipline-specific tools. We hope that we will be able to better integrate metadata creation to the daily work of research groups.

.

Conference Themes

List the conference theme(s) your proposal best addresses (remove the others):

- Content - research data, digital preservation, persistent urls, archiving
- Infrastructure/Integrations - integrations between systems, changing technical environments

Keywords

metadata, cataloguing, research data

Audience

Librarians, data managers and curators, research support services

Background

Quality metadata is an important part of making research outputs genuinely open – not merely online but findable and reusable. This is especially true in the long run when researchers are no more around to answer questions about their data. Creating metadata requires effort but we can make it easier by providing right tools.

Content

Qvain is a metadata editor for describing research datasets.

Good metadata is integral for research data management. Libraries and other support services in research organizations do their best to assist researchers in making their data more visible, and easier to re-use. Quality of metadata is also vital for preservation, as it must answer questions about the data after people involved cannot be reached anymore.

But support service only can help those who help themselves. Researchers' input on their own data is very important. They know what it is all about and how to communicate about the data to their colleagues. We need their contributions but we know that they have other things to do, too.

Some disciplines are more advanced in data sharing, and have best practices as well as data management and archiving infrastructures in existence. However, many researchers lack that kind of community. Many of the researchers and certainly many of the support functions are still learning about the proper life cycle management of data.

In Finland, research data management services are being built on the national level. The Open Science and Research (www.openscience.fi) was an initiative launched run from 2015 to 2017. Its aim was to turn Finland into one of the leading countries in openness of science and research. Among other things, it improved existing services to Finnish researchers and research organizations. These include research data storage services, and the open research data interfaces. A research data search portal provides access to dataset and also tools to apply for a permit to use restricted datasets.

Into these news services are being added.

- The preservation services is based on an existing preservation solution for cultural heritage data (http://www.kdk.fi/en/digital-preservation).
- There is a new Metadata Repository of research data metadata, utilized by the several IT services.
- In the focus of this presentation is Qvain, the metadata creation and editing tool.

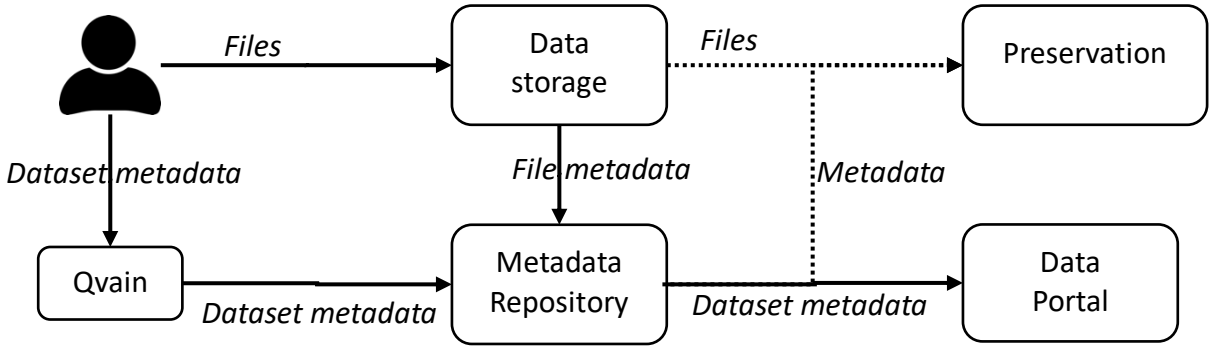

Figure 1. IT Service Components of the Open Science and Research Initiative.

Qvain takes a data model expressed as JSON Schema and turns it into a web form for filling in. Of course, a direct transformation could be very hard to use and difficult to understand. Therefore, Qvain provides an

administration tool that helps to re-order and re-group the fields, to add help texts etc. It also provides a way to employ different widgets in entering different kinds of data. For example, in case of spatial data a map widget might be used.

Qvain has a user management module that uses OpenID. It has an internal Postgresql database for storing the records.

Qvain is designed to work as a standalone system. At the same time, it is being integrated to the larger aggregation of IT services described above (Figure 1):

- The user may work with an unfinished description within the tool. When the user considers the metadata record ready for publication, it is transferred to the shared metadata repository and from there it will be publicly available.
- Metadata records in the metadata repository may be fetched to Qvain for further editing.
- There will be a shared identity management for data storage and Qvain. That means that when describing a dataset in Qvain, the user may view files (that is, file metadata) in the metadata repository, and mark them as belonging to the dataset being described (always provided that they have right to view the files in the storage system).
- The data model that Qvain works on is external. In the first stage we are working with a data model developed in the Open Science and Research Initiative (available at http://iow.csc.fi/model/att/) but of course the tool can work with different models.

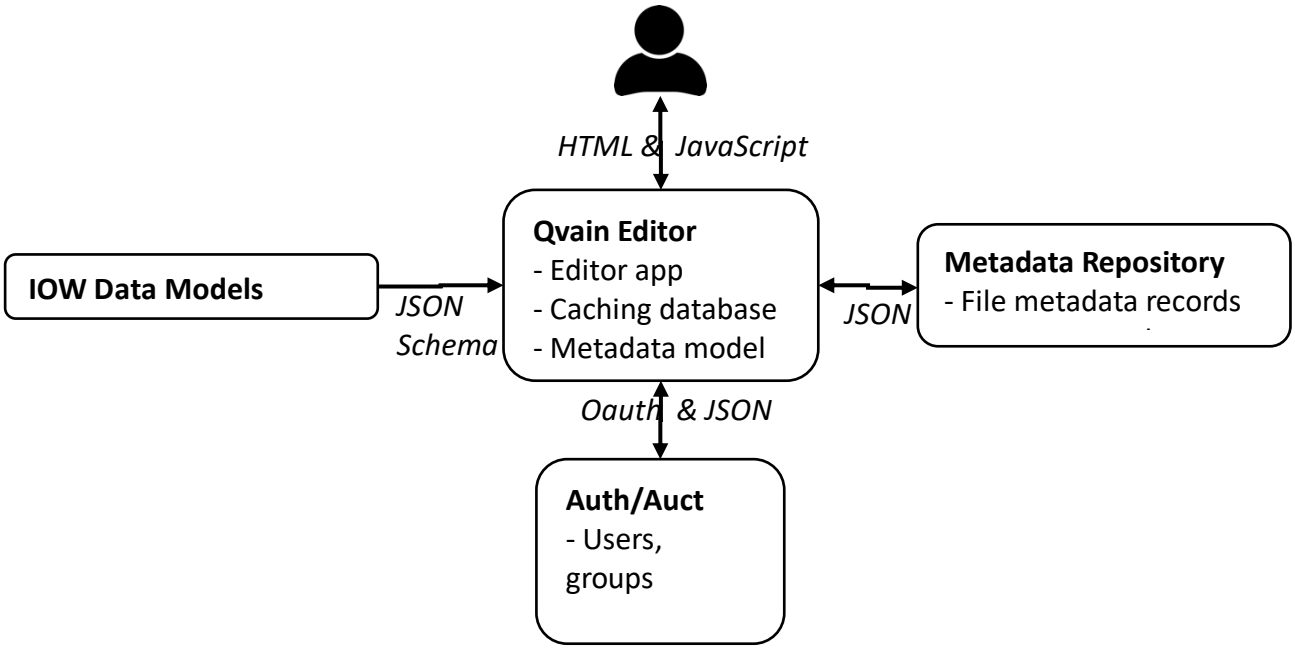

Figure 2. Interactions of Qvain.

Qvain is intended for high-level, manual description. Each file in a dataset can be accompanied by metadata of its own. Qvain does not support such hierarchies but it is always possible to consider a file as an independent dataset, give it a description, and then describe an aggregation of such datasets. In many cases vast amounts of metadata of a more technical nature already exist. We are considering the best way to upload existing metadata into the Metadata Repository. – In addition, the Data Storage will create basic technical metadata of the files and store it in the Metadata Repository.

There were design reasons why Qvain does not manage files at all, in other words, it does not upload or download data files. It is always presumed that there is a PID that can be used as a reference to the files. This is perhaps a decision that will require reconsideration.

Qvain is being developed in the National Library of Finland in a close cooperation with CSC – IT Center for Science, Ltd. The National Library is part of the University of Helsinki. CSC is a state-owned non-profit company that hat provides ICT expert services for research, education, culture, public administration, and enterprises. The work is funded by the Ministry of Education and Culture.

Conclusion

As we all know, being open is about more than putting stuff online. What is not visible and findable will not be used – and really not open at all. We need to engage the researchers in making their output stand out but we must not burden them. We are aware of the pitfalls of generic tools. On the other hand, not all disciplines will have standard metadata practices in place any time soon, not to speak about infrastructures supporting them. We believe that Qvain fills a gap.

Repository System

- Not Applicable