EUDAT as a long-term repository for the University of Porto

The DataPublication@ U.Porto pilot gathers experiments where Dendro, a prototype Research Data Management platform, is used as a gateway to EUDAT. Dendro provides an ontology-based environment for dataset description and publication for the long tail of research. It is built as a multi-disciplinary platform and its preliminary evaluation was carried out with a panel of research groups from the University of Porto. In the scope of the pilot, researchers from several domains within the University of Porto will be asked to follow the steps of a prescribed workflow and organize, describe and deposit datasets created in the scope of their projects.

The Scientific Challenge

The main scientific challenge in the RDM research line where DataPublication@ U.Porto fits is the definition of diverse metadata models and their joint use in the Dendro platform. This has led to the use of recommendation techniques in Dendro, to help users in each domain pick the appropriate descriptors for their data. The second challenge concerns the data management workflows. We intend to build on previous small-scale experiments covering the definition of metadata models, and their use in Dendro to describe datasets, expanding the pilot to a larger multi-domain community.

The main technical challenge in the DataPublication@ U.Porto pilot is the use of EUDAT as a long-term repository for the University of Porto. Besides this, the pilot will also consider the data staging services of EUDAT and assess their features, in order to compare them with those already available in Dendro. Given the diversity of research domains in the pilot, we expect that this will result in some solutions being more appropriate for some research groups than others. The extension of the panel will provide more evidence of the effectiveness of the Dendro platform, while in other cases an all-EUDAT solution may prove more effective.

Another possibility to be considered is a hybrid approach, where Dendro is used in the early stages of RDM, providing a data storage, description and deposit environment to researchers, similarly to what B2Drop and B2Share already do, but long-term deposit and retrieval will be handled by the EUDAT platform. Our platform has so far been tested with a panel of 11 research groups, which we expect to extend to 50 groups during the pilot.

Who benefits and how?

The community served by the [DataPublication@UPorto](#) pilot is the research community of U.Porto, which is the second-largest Portuguese university, covering a broad spectrum of disciplines. It includes large schools of engineering, humanities, science, medicine, architecture, psychology and education sciences, and a business school, among others. We have tested tools to support researchers in the daily RDM tasks and gave emphasis to the design of metadata models to associate to datasets. These models help researchers build metadata records with a good tradeoff between creation effort and exhaustiveness. We expect that benefits from the data pilot will be delivered along three main lines.

The first is the commitment to data publication, from the researchers side. The support of EUDAT provides a perception of the results of their deposit as a long-term solution rather than a project-supported one.

The second is the perception and acknowledgement at the university policy level. We expect decision makers in the university to appreciate and recognize the added value of proposals grounded in international platforms.

The third is the possibility of increasing the impact of the results of a forthcoming funded project by partnering with the EUDAT platform and tools. The project already includes effort in data publication, and assumes that several international platforms will be tested and evaluated against the requirements of the panel. The data pilot

will allow more focused work on EUDAT.

Technical Implementation

The pilot coordinates with a funded Portuguese project, TAIL, started May 2016 and running for 3 years. The DataPublication@U.Porto pilot started right after the EUDAT User Forum, 3-4 February 2016, Rome. The first goal of the pilot is the implementation of an interface between Dendro, U.Porto's platform for data organisation and description, and B2Share as a data publication repository. The second goal is to set up a OAI-PMH server on the U.Porto side and expose metadata created in Dendro to B2FIND.

Preliminary Results

The results for Phase 1 are now complete: 1) the interface between Dendro, the platform for data organisation and description in U.Porto, and B2Share as a data publication repository; and 2) the setup of the OAI-PMH server in Dendro to allow the automatic collection of metadata by B2Find. This has also involved the definition of new data and metadata visibility features in Dendro.

Work corresponding to the first prototype is presented in MTSR'2016 on November 25, 2016 ("End-to-end research data management workflows: A case study with Dendro and EUDAT" by Fábio Silva, Ricardo Amorim, João Castro, João Rocha da Silva and Cristina Ribeiro).

Three goals are established for Phase 2. The DataPublication@U.Porto pilot will beta-test the new B2Share API (v2) in the U.Porto RDM workflow. The second goal concerns B2Find and the OAI-PMH server. The service will be tested with the datasets that are deposited in this phase. The third goal concerns work with research groups and communities. In the scope of the TAIL national project, the pilot will support several research groups in the deposit of recently created datasets. These cases will be documented with feedback from researchers with respect to B2Share and B2Find.

In Phase 3, and according to the pilot proposal, the work with researchers started in Phase 2 will continue and the goals of the pilot will be accomplished.

Contacts

- Cristina Ribeiro, University of Porto, mcr(a)fe.up.pt

Further Information

Scientific Challenges behind the Pilot

Read more