Geophysical Research Abstracts
Vol. 13, EGU2011-10482, 2011
EGU General Assembly 2011

panMetaDocs – A tool for collecting and managing digital objects in a scientific research environment

Damian Ulbricht and Jens Klump
German Research Centre for Geosciences, CeGIT, Potsdam, Germany (jens.klump@gfz-potsdam.de, 49 331 2881703)

Data management in scientific projects is a challenging task. In many cases projects operate with a limited budget for data management that does not allow the development of customized software for data curation. On an institutional scale research data in the earth sciences are described by a number of different metadata schemata. panMetaWorks [1], which is the precursor to panMetaDocs [2], was developed to collect metadata and data in collaborative projects situated at more than one institution. Internet browsers allow easy interaction with panMetaWorks' PHP-based web user interface. Metadata are entered and data objects uploaded through this graphical user interface. A key feature of panMetaWorks is its ability to accommodate any metadata schema. The metadata fields can be filled with static or dynamic default entries to make use of the information implicitly available from the project context. This feature reduces the number of fields that require manual entries to a minimum.

The business logic of panMetaWorks is reused in the development of panMetaDocs, except for authentication and data management functions of panMetaWorks, which are delegated into the repository framework eSciDoc [3]. The eSciDoc repository framework is designed as a Service Oriented Architecture and can be controlled via a REST interface that is accessed by panMetaDocs to create eSciDoc repository items. The framework is designed as an institution-wide data archiving infrastructure and can be accessed by more than one application instance. Once data objects are uploaded to the eSciDoc infrastructure it is even possible to drop the software instance that was used to collect the data while the collected data and metadata reside in the eSciDoc infrastructure and are available for use in other applications. This approach of expendable data curation tools allows for a significant reduction in costs for software maintenance.

panMetaDocs' intention is to allow easy collaboration within a project, collect and curate experimental and measurement data and transfer data objects from a shared into a persistent data curation domain. To accomplish this, only a subset of the lifecycle of eSciDoc items is used. During the workflow starting from the state "pending", through stage "submitted" to the final status "released", objects will be moved from the shared data curation domain to the persistent domain and become available for publication of their data and metadata through data portals. Review and publication of data is, in the case of GFZ Potsdam, a service of its library and therefore the transfer of items to the status "released" is not part of the initial panMetaDocs development. With a RSS interface for recent datasets the reused business logic of panMetaWorks allows project members to be informed about data contributions by other project members. The implementation of the Open Archives Initiative Protocol for Metadata Harvesting interface (OAI-PMH) [4], which is also part of panMetaWorks, can be used to syndicate data catalogues to research data portals. panMetaWorks and panMetaDocs are optimised to serve the panFMP data portal framework [5].

[1] http://metaworks.pangaea.de , Dr. R. Huber, MARUM, Univ. Bremen, Germany
[2] http://panmetadocs.sf.net , only sourcecode, GFZ Potsdam, Germany
[3] http://www.escidoc.org , FIZ Karlsruhe, Germany
[4] http://www.openarchives.org/pmh
[5] http://www.panfmp.org , Uwe Schindler, MARUM, Univ. Bremen, Germany