

ASSISTED METADATA EDITING

Scientists in the role of a data producer are often frustrated by being asked to provide metadata about their data over and over again. To them, this information seems to be obvious from the context of the research project. Such information can be stored in a database and be used as default entries to seed metadata entry forms.

DATA MANAGEMENT PRÊT À PORTER

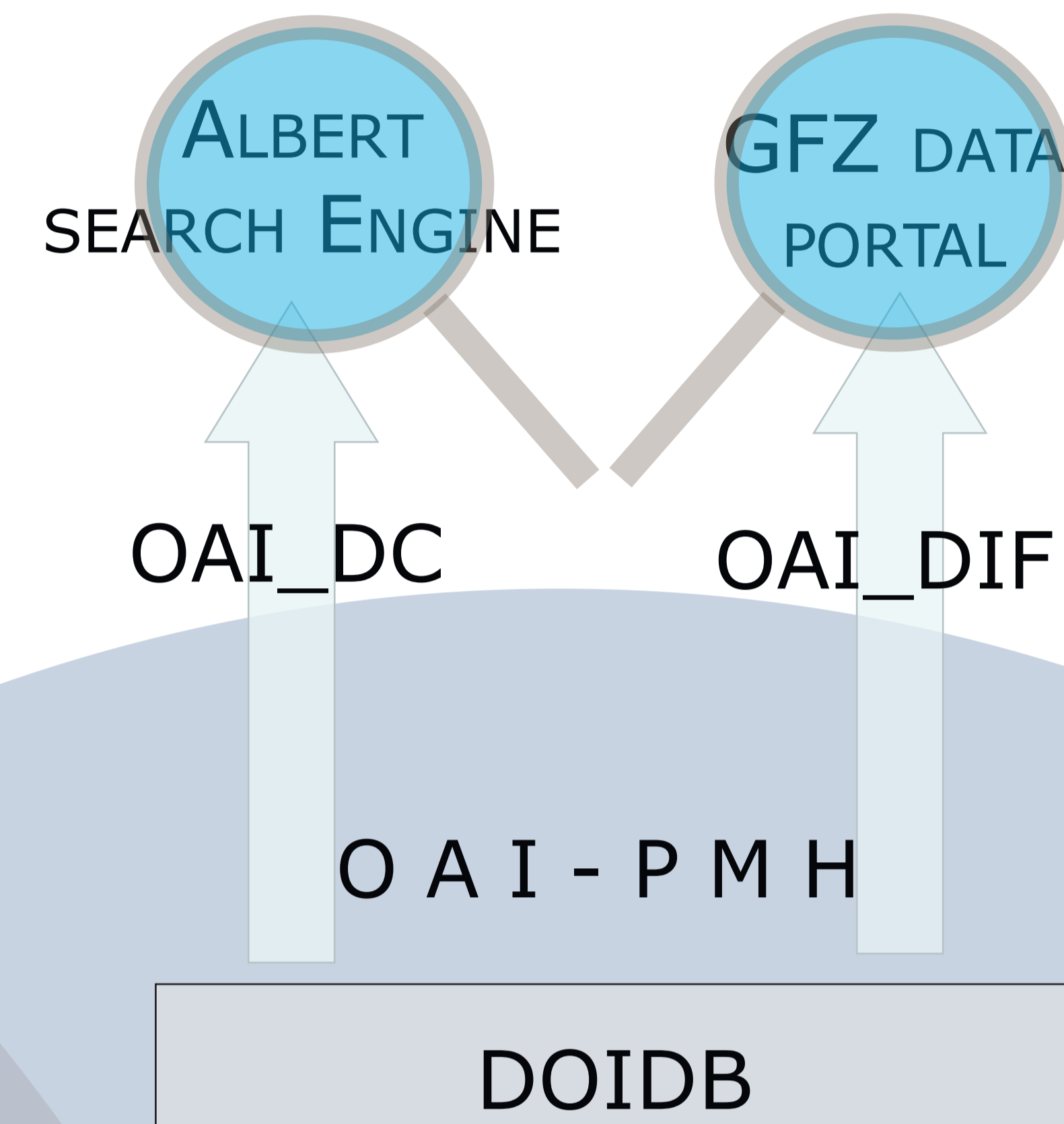
Through the combination of panMetaDocs with eSciDoc research projects can be provided with highly customised software components that support the acquisition and management of data and metadata during the lifetime of the project.

The data acquisition instance is no longer needed once the project has ended. Now the data acquisition instance needs no further maintenance and can be closed. All data are still accessible through the institution data management portal.

DISSEMINATION

DOIDB is an institution-wide registry for online accessible published objects at GFZ. These can either be research papers and posters or scientific datasets, that were registered through DataCite.

Metadata stored in the metadata records of the eSciDoc-items can now be used to produce and syndicate catalogues of published datasets to other data portals. Since standard XML schemas are used, the discipline specific OAI-PMH metadata payload can be easily generated from the existing metadata records by XML stylesheet transformation (XSLT).



DESCRIBING SMALL DATA

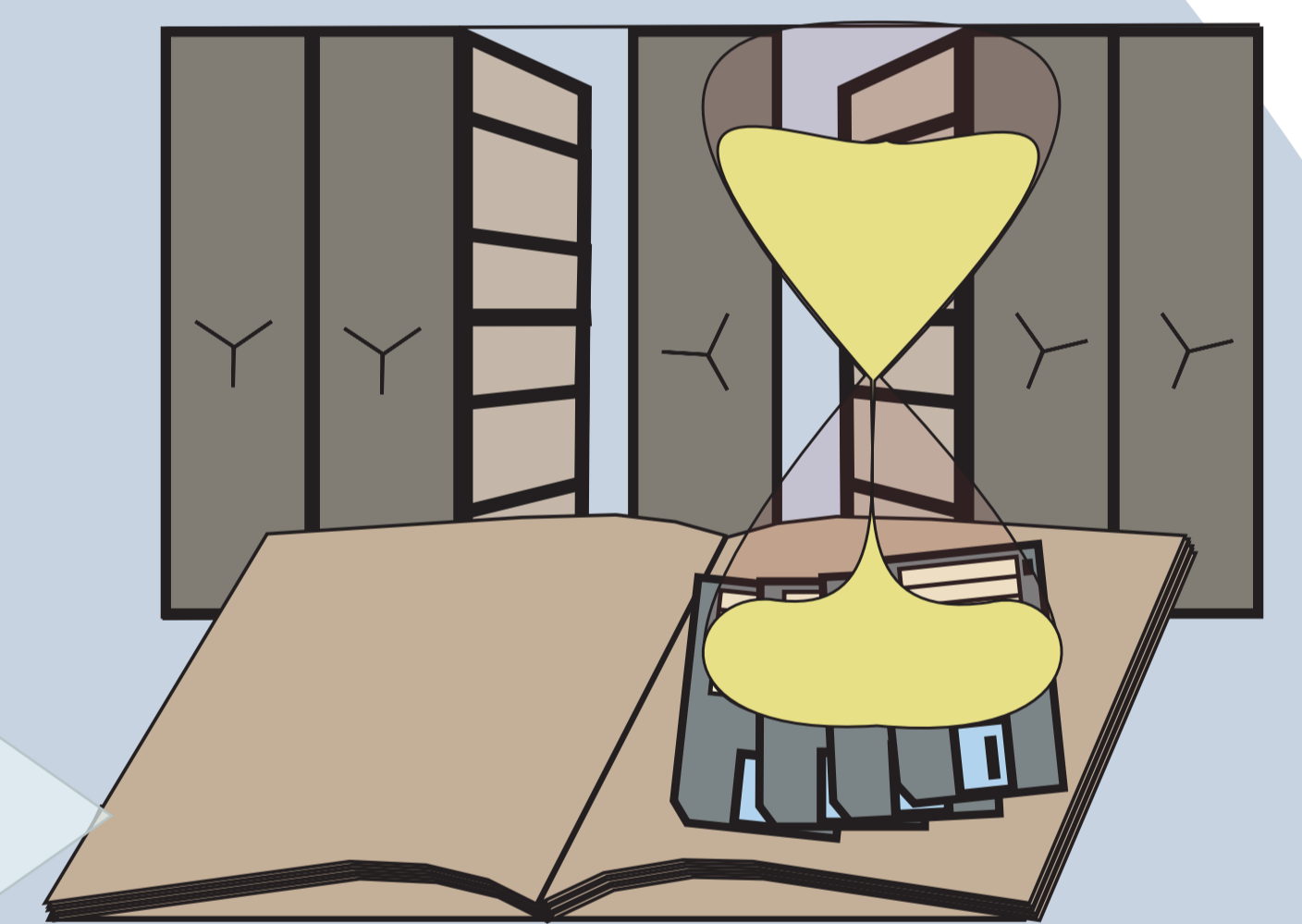
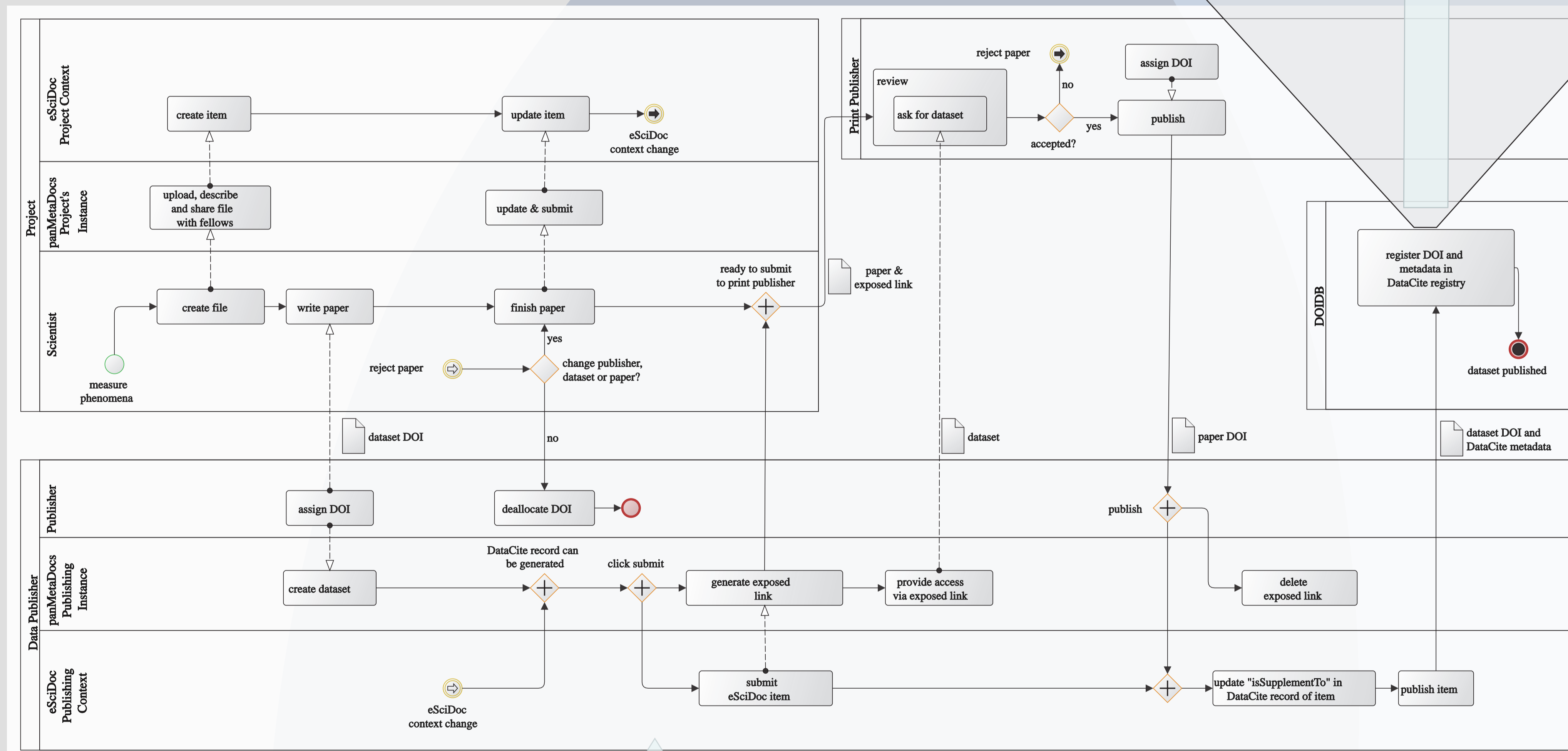
Commonly, the focus is on management of big data, but small data are also very valuable from the perspective of per-item cost. Since small data are generally more diverse than big data, description of small data is more difficult to manage.

Very useful and versatile structures in the eSciDoc content model are its metadata records. These are descriptive XML-subtrees of a repository object. We use eSciDoc items with multiple of these metadata record objects to add information in different metadata schemas that describe our data files in different applications for different purposes.

A DataCite (DOI) metadata record is used for publication of datasets, while georeferenced datasets are put in the right context with a DIF metadata record.

Since different metadata schemas have a common set of elements, like "author", "title" and "description", this does add redundancy to the stored object. On the other hand, this approach allows to store and reuse data without the perils of an overly complex data model that must be maintained and extended with every new arising research application.

PUBLISHING WORKFLOW



ARCHIVING DATA

Once cited in other publications, data and metadata have to be accessible for reference and for future use. To achieve long-term availability, these objects need to be transferred to a long-term archive. OAIS compliant long-term archives are planned for this.

Workflow tools, components and policies required for digital preservation of research data in the geosciences are being developed in our project "Development of workflow components for long-term archiving in geosciences" (DFG Project EWIG). Please also visit our poster IN53B-1632 (Fri PM Session).

eSciDoc Infrastructure