GFZ



Workflows for ingest of research data into digital archives - tests with Archivematica

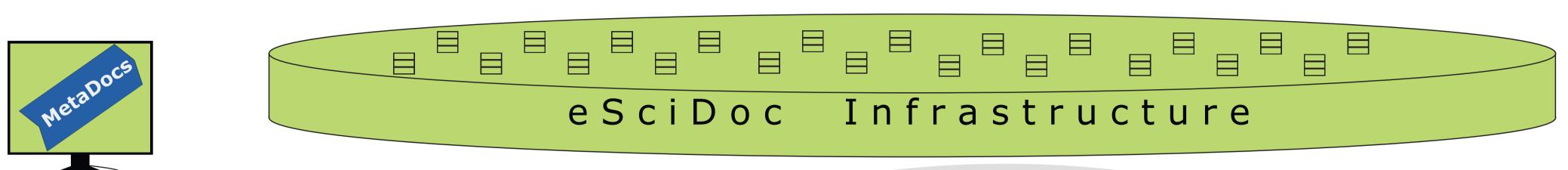


Ingo Kirchner (1), Roland Bertelmann (2), Petra Gebauer (1), Tim Hasler (3), Nora Mettig (1), Jens Klump (2), Wolfgang Peters-Kottig (3), Beate Rusch (3), Damian Ulbricht (2) E-Mail: ewig-projekt@zib.de

- (1) Institute for Meteorology, Free University Berlin, Berlin, Germany
- (2) GFZ German Research Centre for Geoscience, Potsdam, Germany
- (3) Zuse Institute, Berlin, Germany

metadata.

MEVIS



Quality assurance

Both institutions implement quality assurance after the data is collected. In case of GFZ Potsdam the data originate from research projects and are supplementary publication to an article in a peer reviewed journal. So data and metadata have already passed review. To preserve quality checked data Insitute for Meteorology established an exemplary workflow. New tools are developed to support quality checking, to transfer data from MEVIS into the institutional repository, and allow institute's staff to enrich data with

Dataset files & continuous data

GFZ Potsdam and the Institute for Meteorology at FU Berlin hold two different types of data. On the one hand there exist file based data and metadata in GFZ's eSciDoc-repository. On the other hand the Institute for Meteorology operates amongst others an urban network of monitoring stations, that deliver continuing time series of data.

Get data into the archive

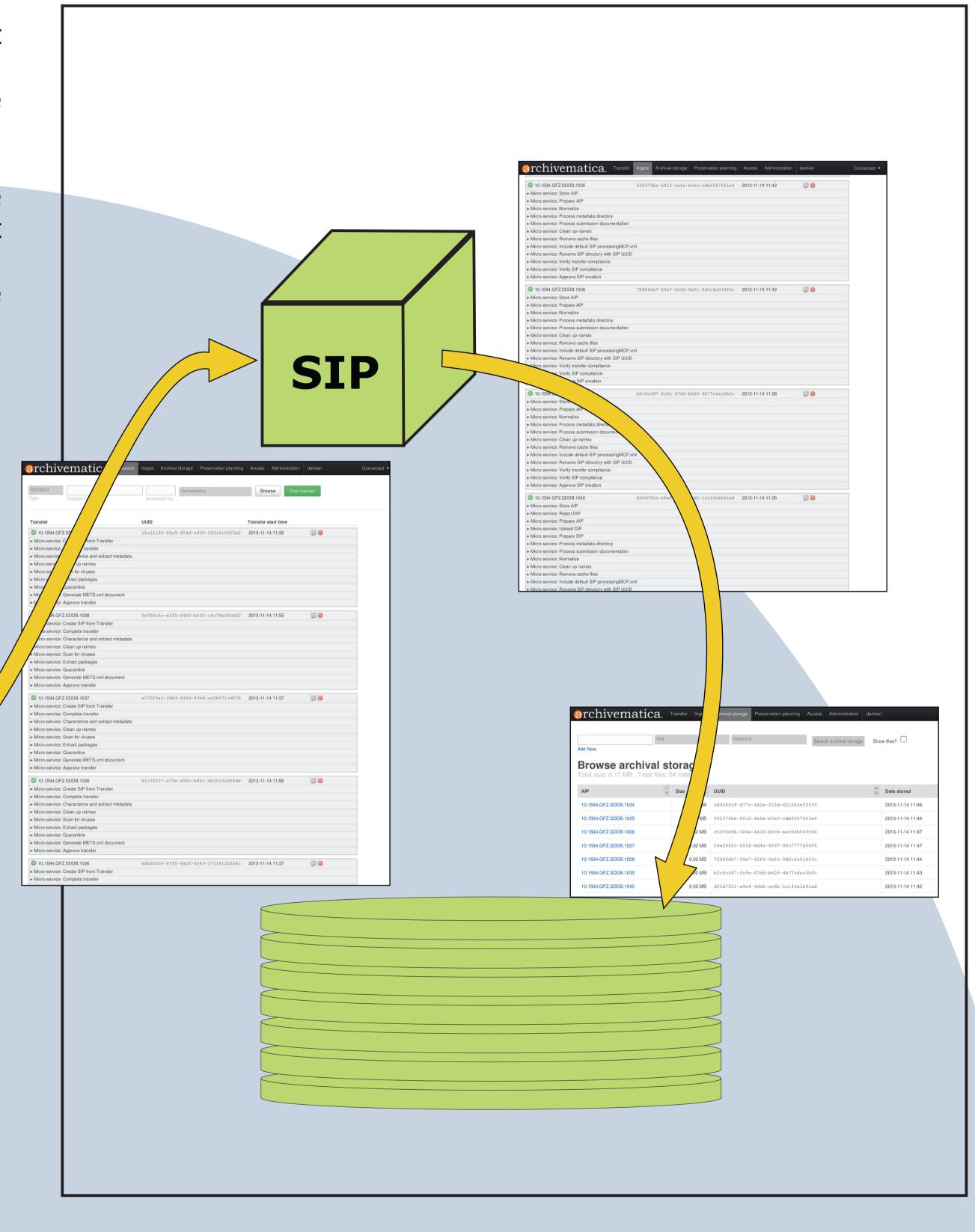
After all preparations are done the data will be packed into transfer-packages and will be transferred to the preservation system at Zuse Institute.

SIP-Packaging and Archivematica

Archivematica is a free and open source preservation system that currently can be installed in version 1.0 on top of Ubuntu Linux. The software allows to generate SIP-Packages with additional preservation metadata through a web-based dashboard. While bagit provides a good way to ingest data into archivematica we are still evaluating software and do software development to convert our metadata into the METS XML-format.

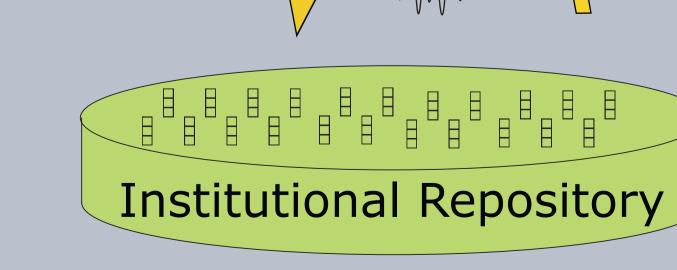
Transfer Package

archivematica。



Training and Policies

Quality of metadata and quality of data are key issues for the re-use of archived data. Policies are needed to establish a common practice in data curation and there is a need to teach data curation basics to young scientists. Developing policies and ideas for teaching material are our task in project EWIG. Please visit our weblog at http://ewig.gfz-potsdam.de.



Resulting AIP-Packages

The AIP-Packages generated with Archivematica are fully qualified for the long term preservation of datasets as the software generates an inventory of stored file types. However, Archivematica is still under heavy development and the system lacks at the moment possibilities to properly transfer metadata of the earth sciences into the METS metadata structures. Import of FGDC metadata and ISO metadata is one of the next goals in the development of the software. It is part of the scheduled "dataset preservation" feature sponsored by the University of Alberta.

