



EUDAT services to guarantee long time archiving and visibility to the repository of IST Austria

IST DataRep is the institutional repository for publishing research output of IST Austria affiliates. IST DataRep was implemented to help scientists fulfil the requirements from funding bodies and to meet the growing impact of publishing research data. Therefore, the deposited data collections will be mainly open access.

The Scientific Challenge

The repository is mainly designed for the demands of data publication. This was the main aspect we were focusing on regarding the data life cycle. Each data collection is assigned a DOI to grant it's citeability. But a DOI doesn't only enable citation it also facilitates persistence, which asks for longevity of the data collection. IST Austria has an internal back up strategy running but a truly safe strategy is only guaranteed with offsite data storage. Scientists at IST Austria are encouraged to deposit data at established subject repositories (I.e. Dryad, Gene Bank) but for many long tail of science domains, these repositories are not available.

Who benefits and how?

IST Austria affiliate researchers & scientists dealing with data publication.

Technical Implementation

IST DataRep is the institutional repository for a small scientific operation and even though the content is publicly accessible it needs to be indexed in international platforms/search engines to obtain sufficient visibility. B2SAFE and B2FIND are planned to be additional services to guarantee long time archiving and visibility. Therefore the technical preconditions have to be fulfilled.

On the one hand this is the capability of generating bundles (data collections + metadata) via a REST API and develop a workflow and technical features for the transfer to the EUDAT B2Safe service. On the other hand the metadata has to be collected and indexed by EUDAT. Regarding B2FIND we assume that the implementation of the service won't need any technical development because IST DataRep is an OAI-PMH compliant repository.

Preliminary Results

Implementing B2SAFE: We are collaborating with KIT (EUDAT CDI node) for long term storage of our data collections. We agreed on the export of our data via XML packages which Eprints offers as a feature. After the IRODS server was set up at KIT we tested to transfer/push XML packages from IST DataRep to KIT. The first issue was to rewrite the script for the data export insofar that each data collection is exported in a single XML package (XML metadata + base64-encoded data files) and not all collections packed in one. For new uploaded or updated data collections a script will activate the export. The name of the packages is generated via the Eprints ID so an updated data collection (metadata update) has still the same ID and will therefor simply be replacing the old package at the KIT server.

Implementing B2FIND: This task was quite straightforward. Via the OAI-PMH the metadata is collected and made available in B2FIND. We only had to adjust the protocol insofar that also DOIs are indexed and therefor represented in the metadata record.



Contacts

- Barbara Petritsch, IST Austria, [barbara.petritsch\(at\)ist.ac.at](mailto:barbara.petritsch(at)ist.ac.at)
- Heinrich Widmann, DKRZ, [widmann\(at\)dkrz.de](mailto:widmann(at)dkrz.de)

Further Information

[Scientific Challenges behind the Pilot](#)

Learn more about the IST DataRep/EUDAT collaboration [here](#)

[Read more](#)