

OpARA -
Establishing a multi-institutional
research data repository
for the TU Dresden and the TU BA Freiberg

Stefan Fritzsche, Christian Löschen, Holger Mickler, Ralph Müller-Pfefferkorn

Usage Scenarios

Usage scenarios for the chosen repository platform/product in your organization (who interacts with it, and how)

- **Open Access Repository and Archive**
- Institutional repository – wide variety of disciplines
- Multi-institutional repository – support for multiple organizations
- Map existing organizational structure in repository for automatic access rights management
- Scenarios
 - Interactive work: researcher uploads data and describes it via Web-Interface
 - Batch ingest: researcher creates new item and describes it via Web-Interface, data can be packaged and uploaded by a command-line/background tool
- 2-stage review process for submissions: 1) community-based content review, 2) technical review from repository operator

Selection criteria your institution has employed to choose that research data repository platform/product

- Survey within the two Universities on requirements from user perspective
- Requirements from development and operations perspective
- Developed our own matrix (2 years ago)

Software Selection

- Working repository software – nearly out-of-the-box
 - Open Source product – no licensing costs
 - International deployments
 - Large community
 - Ongoing development and therefore a low risk of extinction
 - Web-based user interface
 - Support user roles and groups, shibboleth-authentication (DFN-AAI)
 - Standardized interfaces (OAI-PMH, REST, RDF, SWORD)
 - Support persistent IDs (DOI)
 - Support for public (Open Access) and non-public data
 - ...
- Decision: DSpace

Experiences with DSpace

Positive and negative attributes of the chosen repository platform/product, based on your experiences

- + Many things work out of the box
- but many need (complex) adjustments!

- + Overlay system for incorporating own/additional code – easier code management
- Debugging with overlay code not always working properly
- Incorporating updated code from upstream more complex

- Complex Web Interface programming (XML-UI) → Hopefully improved by new DSpace7 Web-Interface
- Often poor code design/quality

Experiences with DSpace

Positive and negative attributes of the chosen repository platform/product, based on your experiences

- Metadata Handling

- Only plain metadata possible – no hierarchy or complex metadata

- Limited metadata schema layout

- Both limitations fixed for OpARA by introducing a JSON-based metadata format for complex metadata (large effort)

- Submission forms very limited in form type and content due to metadata limitations

- Additional complex form types introduced for OpARA in conjunction with the JSON-metadata improvement

Experiences with DSpace

Positive and negative attributes of the chosen repository platform/product, based on your experiences

- + Hierarchical layout of repository content
- Support for long term archiving
 - No support for content verification and preservation out-of-the-box
 - Adverse layout of assetstore (files of a single item are spread over different directories in the file system)
- + Configurable submission workflow

Experiences with DSpace

Positive and negative attributes of the chosen repository platform/product, based on your experiences

+ Out-of-the-Box Shibboleth integration

- Adaptions to integrate several identity providers (multi-institutional)

+ Fine-grained access control possible but

– OpARA modification of (Open Access) rights handling necessary to make clear when an item is publicly available (make sure no implicit Open Access is possible)

Experiences with DSpace

Wishes, visions and ideas for how research data repositories should be developed or enhanced in the future

Better and more flexible metadata handling

Better integration of archiving, e.g. archival workflows

More support for validation (e.g. formats)

More focus on (research) data

- Submission of data sets
- Assetstore design for large data sets (many files)

Support for automated external ingest (e.g. from command line)

Integration of external tools, e.g. Metadata Catalogue

Interfaces to directly access data

It worked out –
will start operation in May 2017.

But took longer than expected –
more implementation and adaption work needed!