

Data Handling Processes in Edmond

Dr. Yves Vincent Grossmann
grossmann@mpdl.mpg.de

Max Planck Digital Library



MAX PLANCK
digital library

Steps:

1. Ingest
2. Archival Storage
3. Data Management
4. Access
5. Administration

Ingest

- In Edmond, the first supplied version of a dataset (including a metadata record and one or more data files) is a draft version. It is assigned a draft DOI and is not publicly available yet. This corresponds to the OAIS Submission Information Package (SIP).
- In general, Edmond does not create separate Archival Information Packages (AIPs), which sets it apart from OAIS. Instead, the dataset is prepared as Dissemination Information Package (DIP) at this stage.
- The system does not keep track of changes on the dataset until it is published. In this phase, the depositor is allowed to give other users access to the dataset, e.g. for reviewing purposes.

Archival Storage

As described above, DIPs instead of AIPs are prepared during ingest.

Data Management

- Once published, any change on a dataset (metadata and/or data files) results in a new unpublished version. Thus, any alterations are accurately documented and ensure the authenticity of a dataset.
- Deaccessioning (public access removal) is still possible for important reasons.

Access

- Datasets can only be published when they provide certain required metadata for citation and findability. Every published dataset is registered with a DOI and metadata at DataCite.
- Additionally, different other metadata formats are provided, which can be used by third party catalogues and search engines and result in better dissemination (e.g. JSON-LD Schema.org for Google Dataset Search).

Administration

- In Dataverse, the role of the Administration function can technically be distributed across different internal sections.
- However, in Edmond, the daily operations of the repository will be managed by a dedicated service manager with superuser access rights.

Thank you for your attention

edmond@mpdl.mpg.de