

Archive III

19. desember 2013

UNINETT



Accessing Data

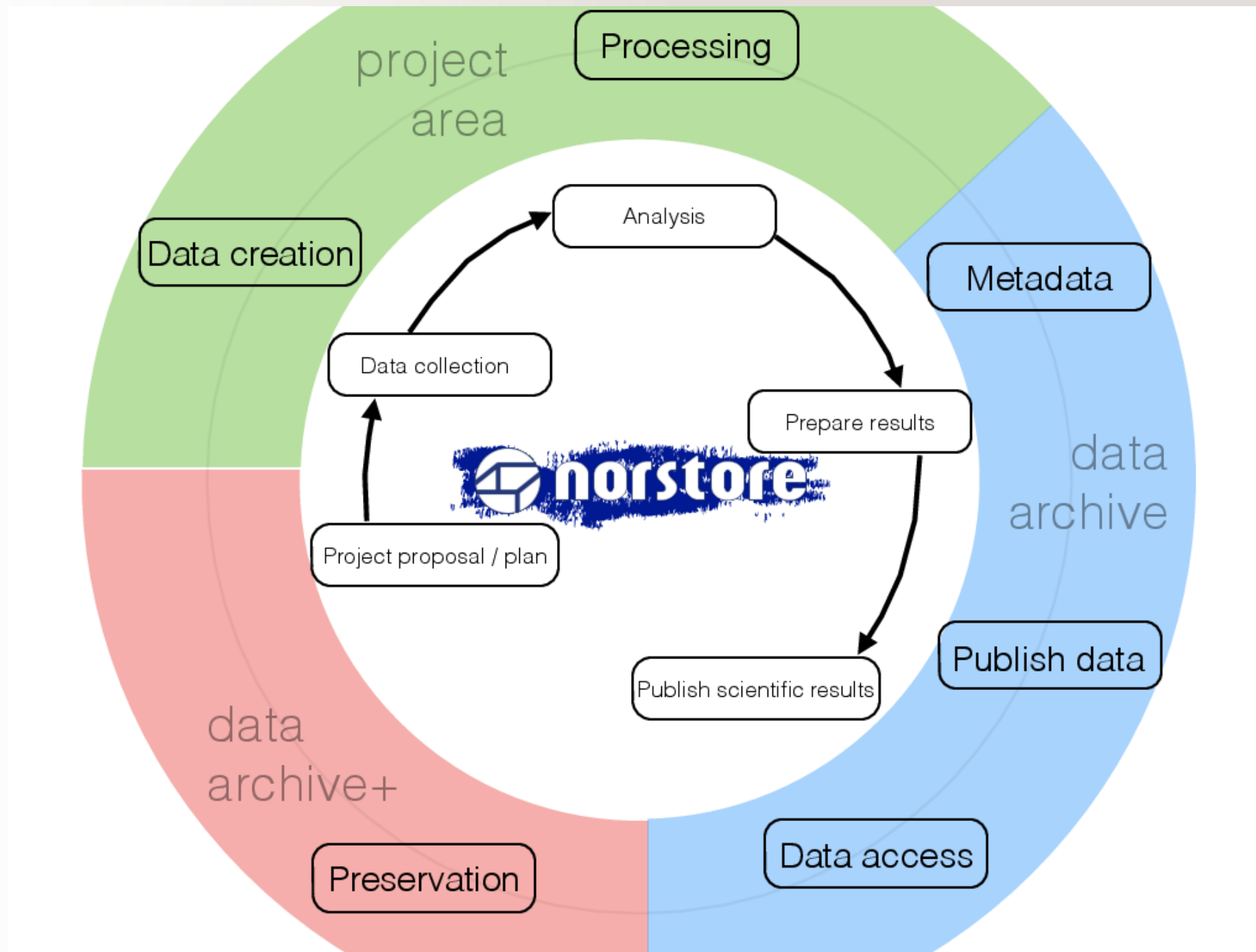
Data Reuse

- Datasets < 5GB can be downloaded via the web interface.
 - Use needs to abide by the access licence.
 - Only an email address needs to be supplied for un-registered users (in order to send the link to download).
- Datasets larger than 5GB can currently be downloaded to the Norstore project area.
 - Users need to be registered Norstore users.
 - Dataset is downloaded to ARCHIVE_STAGEDIR/<username>/<UUID>
 - For other users we are currently working on mechanisms for download.

Data Reuse

- It's good practice to cite the dataset that you used in your publications.
 - Use the DOI in the citation - this will make it possible for other researchers to access the datasets you used.
- If you wish to archive your research findings you should also record the additional datasets using the optional metadata “source” term.
 - If a DOI exists for the dataset use that. If not, try to describe the dataset in documentation included in the dataset.

Preservation



Preserving Datasets

- Digital preservation attempts to ensure digital remains accessible and usable by future users.
- This is addressed by:
 - Ensuring bit-level integrity through data replication.
 - Ensuring data is understandable (may require adding or updating metadata on how to use and interpret the data).
 - *Ensuring data is discoverable (equipped with the right and relevant metadata and description).*
 - Ensuring data in usable format (may require migration from obsolete formats to new formats, or virtual environments).

Migration & Virtualisation

➤ Things to be aware of for Migration:

- What's the best format (most durable, popular, open)?
- What features in the data need to be maintained and how can we check they are?

➤ Migration pros/cons:

- Easy to use new tools with old data, easier to integrate data into new/current workflows
- One-way street. May lose some features/functionality in the migration that may only be relevant later.
- Requires experts to be able to assess what features need to be kept and whether they are indeed kept.

➤ Things to be aware of for Virtualisation:

- What type of virtual machine to use (licensing, rendering, performance)?
- Are all the resources required by the application contained within the VM?

Migration & Virtualisation

➤ Virtualisation pros/cons:

- Preserves original features/functionality (little risk of missing something).
- Can be difficult to integrate with newer tools.
- If large volume of data may not be scalable option.

➤ Choice depends on your circumstances and needs.

Versioning Datasets

- There may be cases where a new data is added to an existing dataset making a new 'edition' or 'version' of the dataset desirable.
- Norstore metadata supports 'versioning' via a few metadata terms:
 - Replaces - a link to the dataset this dataset replaces.
 - IsReplacedBy - a link to the dataset that replaces this dataset.
- Metadata also supports collections of datasets:
 - isPartOf - a link to the dataset that this dataset is a part of.
 - hasPart - a link to the dataset that is part of this dataset.

Command-line queries

- Working on REST-based scripts to query the metadata catalogue.
- Allow users to access the catalog from applications
- Provide OAI-PMH interface for harvesting metadata

Auditing

- Aim to pass Data Seal of Approval (<http://www.datasealofapproval.org/en/>)
- Ensures the archive conforms to best practice
- Allows users to assess how reliable the archive is.