Using a warehouse to disseminate data

The open dissemination of research data guarantees that research is transparent and reproducible. Researchers are now obliged to open some of their data following the axiom “as open as possible, as closed as necessary”.

Storing and depositing data are two distinct operations: while storing pertains to the moment when data is collected or treated, depositing occurs to preserve, share, and disseminate data, and to allow data to be linked to publications. Depositing data thus guarantees that research is transparent and reproducible.

What data to open ?

- The Digital Republic and Valter laws introduced the principle of public data being made open “by default”. But communicating data about defence secrets, professional secrets, the protection of people, public health, and public order is forbidden.
- Disseminating research is subject to conditions for particular cases of personal data
- It is obligatory to open geographical and environmental data (under the INSPIRE directive).

Depositing data in a warehouse

Data deposited in a warehouse exists independently of the scientific article: it needs to be described by the richest metadata possible to make it easy to find, thus encouraging sharing and reuse. A permanent identifier or access number is attributed to each dataset making it visible, accessible, and stable, just like the publication.

How to make your data FAIR :

Findable

make your data findable by :

- Describing it using rich metadata
- Attributing a unique and permanent identifier (a DOI, for instance) to (meta)data.

Tip : most warehouses attribute a permanent archiving identifier to datasets.

Accessible

make your data accessible by ensuring that :

- The warehouse used to share data attributes permanent identifiers enabling data to be recuperated.
- Metadata is, as far as possible, accessible, even if data is not. The access procedure may involve authentication and authorisation steps if necessary.

Interoperable

make data interoperable by using :

- Whenever possible, open-source and widely used formats, software, and languages, enabling exchange between IT systems and increasing the capacity for combining metadata
- Permanent identifiers: DOI, PMID, SWHid, arXiv ID
- Repositories: idRef, ORCID, RNSR
- Controlled vocabularies: DC, RDF, FOAF, SKOS, BILBO, Fabio

Reusable

make data reusable by ensuring that :

- Data is well documented to support correct interpretation.
- That a clear and accessible user licence is attributed so that other researchers may know what types of reuse are authorised.
- Information on provenance is available to clearly indicate how, why, and by whom the data was created and processed.

Choosing a warehouse

A Research Data Warehouse or Repository is a database for gathering and conserving research data, and making it visible and accessible. Its role is to enable data to be collected or deposited, accessed, and shared for reuse.

Each warehouse tends to have a policy for depositing, describing, and disseminating data. One of the criteria for choosing a warehouse may be the possibility of attributing a licence imposing citation of those who created the data when it is used.

There are several types of warehouse :

- disciplinary
- multidisciplinary
- institutional
- publisher-specific
- project-specific

To choose a reliable warehouse, it is recommended that you

- Check if a warehouse is recommended by one of the parties involved in your project (your funder, publisher, or institution)
- Find a warehouse adapted to your needs by using **warehouse directories** and/or looking for **certified warehouses**.

For HASS, a particularly noteworthy warehouse is Nakala, run by Huma-Num, which meets most needs.

Contact
For any question about research data
guichet-ardoise@groupes.renater.fr
Fiona Edmond
Research data officer, University Library
fiona.edmond@univ-rennes2.fr
Morgane Mignon
Digital humanities platform coordinator, MSHB
morgane.mignon@mshb.fr
Tools

- How to choose a data repository
- Nakala Documentation (French)
- Sharing data on scientific publications (French)
- Guide to reviewing a dataset before publication (French)
- Research data and source code that must be made available to the public (French)
- CoopIST guide (French) : Do you have the right or the obligation to distribute your data?