

Researcher Guidance
for the Selection of
Trustworthy Repositories

TEMPLATE FROM THE SCIENCE EUROPE PRACTICAL GUIDE TO

THE INTERNATIONAL ALIGNMENT OF
RESEARCH DATA MANAGEMENT

The following guidance for the selection of trustworthy repositories is structured according to four main topics:

# Provision of Persistent and Unique Identifiers

A trustworthy repository should:

1a. Allow data discovery and identification

* ensure that PIDs are included in the corresponding metadata.

1b. Enable searching, citing, and retrieval of data

* consistently assign PIDs (for example a DOI,[[1]](#footnote-1) URN,[[2]](#footnote-2) ARK[[3]](#footnote-3)) to the data it holds, allowing the corresponding data and metadata to be found, referred to, and retrieved, even if the location where the data are stored changes.

1c. Provide support for data versioning

* ensure that the version of the data stored in the repository is clearly specified and documented via a permanent audit trail in order for the provenance to be traced.

*Note: Not all repositories use an accepted and universal PID system such as the ones mentioned above. Instead, they use a local identifier or administrative number that the repository itself maintains. This increases the risk that the data cannot be found anymore if they are moved to another location, or if the repository ceases to exist, reorganises, or becomes governed differently.*

# Metadata

The data should be accurately described with rich metadata. The metadata should document how the data were generated, under what license and how they can be re-used, and provide the context for proper interpretation by other researchers.

2a. Enable finding data

* ensure interoperability and re-use of data by others by providing the data and metadata in an accessible language, based on a well-established formalism. Data and metadata should be described using standard vocabularies and formats allowing computer systems to search for them, combine them in an automatic way, and distinguish the metadata from the research data file(s).

2b. Enable referencing to related relevant information

* ensure that in the metadata information it is possible to declare links to other relevant or associated information by providing the PID and a description of the scientific relation. One particular kind of information are details on the associated researcher, for which permanent research IDs exist (such as ORCID,[[4]](#footnote-4) ISNI,[[5]](#footnote-5) or DAI[[6]](#footnote-6))

2c. Provide information that is publicly available and maintained, even for non-published, protected, retracted, or deleted data

* ensure that metadata are archived for the long term and that metadata always remain retrievable, even if the corresponding research data are not or no longer available (for example due to privacy restriction, legal obligations, or other protective measures).
* ensure that retracted data due to poor research practices or misconduct are still findable through the metadata and preserved in order to allow examination of the research record.

2d. Use metadata standards that are broadly accepted (by the scientific community

* ensure that the metadata maintained by the repository are machine-retrievable and use standards that are broadly accepted (by the scientific community).
* ensure that community standards or best practices for data handling are followed if these exist. Note that repositories that are specialised in a particular research field may have community standards regarding the data and metadata that are uploaded.

2e. Provide information that is publicly available and maintained, even for non-published, protected, retracted, or deleted data

* encourage that the information contained in the metadata are structured in a way that allows machines to retrieve it, for example by providing a form with specific fields to be completed.

# Data Access & Usage Licenses

A trustworthy repository should:

3a. Enable access to data under well-specified conditions

* be clear about the terms under which the data can be re-used. Such (license) information is usually included in the metadata.

3b. Ensure data authenticity and integrity

* ensure that the metadata contain detailed information about the provenance of data, including how they were generated, how they were processed, in which context they may be re-used, and how reliable they are.

3c. Enable retrieval of data

* allow retrieval of data or at least metadata using an open and standardised protocol (not a proprietary communication protocol).

3d. Provide information about licensing and permissions (in ideally machine-readable form)

* allow license information to be referred to in a structured way, so that the conditions of use are clear, preferably to humans as well as to machines. Where possible, common or broadly accepted licensing systems should be used (such as Creative Commons) which can be referred to by URL.

3e. Ensure confidentiality and rights of data subjects and creators

* provide a way for authentication and authorisation of human and machine-users, allowing to set user (or group) specific access rights to account for data with confidentiality issues and other restrictions.

# Preservation

A trustworthy repository should:

4a. Ensure persistence of metadata and data

* ensure the preservation and continued availability and access to the data and metadata entrusted to it by its users.

4b. Be transparent about mission, scope, preservation policies, and plans (including governance, financial sustainability, retention period, and continuity plan)

* manage the preservation of data and metadata in a documented way. In particular, it should have a preservation policy that details the mission and scope of the repository, governance aspects, financial sustainability, outsource partners, and retention periods (the timeframe of preservation).
* have a publicly available contingency plan and ensure preservation of data and metadata beyond the lifetime of the repository (for example by allowing easy extraction and transfer of data and metadata to another repository).
1. Digital Object Identifier [↑](#footnote-ref-1)
2. Uniform Resource Name [↑](#footnote-ref-2)
3. Archival Resource Key [↑](#footnote-ref-3)
4. Open Researcher and Contributor ID [↑](#footnote-ref-4)
5. International Standard Name Identifier [↑](#footnote-ref-5)
6. Digital Author Identifier [↑](#footnote-ref-6)