

## Long-term archiving/storage of research data

At the end of any research project, the question arises of how best to archive the data or ensure their long-term storage in complete security. The goal of archiving is to ensure that your data are available and readable to other researchers or authorities who wish to reuse them or inspect them in the future. You will need to ask yourself several questions to ensure the correct archiving of your data: What data do you want to keep? Where? And under what conditions? With this practical guide, the CeRIS can help you answer these questions.

### Which data should be stored in the long term?

When sorting through your data to determine which should be stored in the long term, you should take three criteria into account:

#### Usefulness

You should conserve data that:

- Might be reused in future research
- Are unique, non-reproducible, or difficult to reproduce
- Make possible a global understanding of the research project
- Prove that the research has been well conducted
- Can be used to validate the results presented in scientific publications

#### Legal obligation

As an example: for biomedical studies on blood products, research documents and data must be stored for 40 years after the end of the biomedical study or its early cessation (Ministerial order of November 8, 2006).

As an example: personal data must be deleted once the needs for which they were collected are satisfied (right to oblivion enshrined in the French data protection act). The duration for which non-anonymized data can be stored is determined by the CNIL. Contact [rgpd@pasteur.fr](mailto:rgpd@pasteur.fr) for more information.

#### Historic/heritage interest

As an example: lessons can still be drawn from epidemiological data on the 1918 Spanish flu pandemic that are still being analyzed 100 years later.



This document is made available under the terms of the [Creative Commons CC-BY-NC-SA License](https://creativecommons.org/licenses/by-nc-sa/4.0/)

## Where should research data be archived?

To ensure their long-term preservation, the ideal is to store the data in a database that is:

- **Certified** for long-term storage and management: to ensure the long-term readability, integrity, and intelligibility of the data
- **Searchable**: to facilitate the discovery and reuse of the data

### How to find a data repository certified for the long-term storage of data

#### How to find a repository

You can search for a data repository corresponding to your research domain by consulting:

- [FAIRsharing](#): a directory of data repositories for the life sciences
- [Re3data](#): a directory of multidisciplinary data repositories
- [The list of repositories in the biomedical domain proposed by the CeRIS.](#)

#### How to know whether it is certified for the long-term storage of data

You should check the archiving policy on the website of the data repository.

Examples:

- [The preservation policy of Harvard Dataverse](#)
- [The archiving policy of Dryad](#)

The Institut Pasteur does not have such a certified and searchable internal database. For all data that cannot be stored in a repository (confidential or sensitive data, data subject to regulation, etc.), it is recommended that the following good practices be followed.

### Archive the data at the Institut Pasteur

#### Create a preservation space

In your storage entity or project space (Gaïa or Synplicity) :

- Create a new folder at the root called "**Archiving\_unit**" or "**Archiving\_project**"
- Regularly transfer electronic data into this space following good practices
- At the end of the project or when a collaborator leaves, check that all data has been transferred.

#### Good practices to follow

- Create a **file classification plan** and define **naming rules** in this space
- **Associate documentation or [metadata](#)** with the data to make it understandable
- Convert the data to a **long-term file format** to keep the data readable and intelligible for as long as possible (see below)

## How long should data be stored?

There is no archiving table for research data like that available [for documents](#). However, the same logic can be used to determine how long to store data.

**Example:** at the Institut Pasteur, laboratory notebooks are kept for at least 25 years, as they validate research results. Following the same logic, research data associated with publications validating published results should be retained for at least 25 years.

In addition, when determining how long to store data, the following factors must be considered:

- **Legal obligations** and regulations
- **Standard requirements**, if relevant in your context
- **Risks associated with the data being unavailable.** For example, for unique data impossible or difficult to reproduce, the risk is a waste of time and non-negligible costs to retrieve/regenerate the information.

If in doubt, you should always store the data for as long as possible.

## Which archiving format should you choose?

You should choose a **long-term file format**, to ensure that your data remain readable and intelligible in the future. This format should be:

- **Open** (see the definition below)
- **Widely used**
- **Standardized** (if possible).

### What are open formats and closed formats?

Open format	Closed format
The functional and technical specifications are <b>public</b> and available <b>free of charge</b> (or at low cost)	The functional and technical specifications are <b>not public</b>
There are no legal restrictions to its <b>free use</b>	<b>Legal restrictions</b> preclude its use (patent, copyright)
A document saved in an open format is <b>independent of the software</b> used to create, edit, read, and print it. Open formats therefore allow data <b>interoperability</b> .	The format can be read only with a <b>specific program</b> or software suite
It is maintained by a <b>nonprofit organization</b>	Closed formats are often <b>proprietary</b>

**Examples of open and closed formats**

	<b>Closed format</b>	<b>Open format</b>
Text	DOC, PAGES	ODT, TXT
Spreadsheet	XLS, NUMBERS	ODS, CSV
Video	AVI, WMV	MP3, MP4, MKV
Image	PSD, JPEG 2000	PNG, BMP, JPG, GIF
Archiving – compression	RAR	ZIP, TAR
Presentation	PPT	ODP