

Communication		April. 2023
<h1>Data Management Plan – Data Retention and Data Storage</h1>		
Thomas Cokelaer, Dimitri Desvillechabrol, Etienne Kornobis, Laure Lemée		



Data Management Plan – Data
Retention and Data Storage
Reviewers: Marc Monot
APRIL, 2023

1 Data Storage and Retention

This excerpt is taken from the internal Data Management Plan of the Biomics Platform. It outlines the duration for which data will be stored on the Biomics server, depending on its origin.

1.1 Different users, different usage

The data is categorized into three types: Biomics, Autonomie, and Externalisation.

data type	description
Biomics	data sequenced in-house by the Biomics Platform itself.
Autonomie	data sequenced by a member of the Institut Pasteur (non-Biomics member) using the shared room facility (salle partagée)
Autonomie - Externalisation	data sequenced by an external company, partner of the Institut Pasteur and the Biomics Platform, for a member outside of the Biomics Platform.

Table 1: Data types

1.2 Data retention

The Biomics Platform stores all sequencing data on Institut Pasteur data storage systems, namely ZEUS or GAIA. The DSI ensures that data duplication takes place. ZEUS is a high-speed data storage system, while GAIA is slower. The duration for which data is retained depends on the type of data:

1. If your project was sequenced by the Biomics Platform, the Fastq files of your sequenced data will be stored on the ZEUS data storage system for a period of 2 years, after which it will be transferred to the GAIA data storage system for another 3 years. After a total of 5 years, the accessibility of your data cannot be guaranteed. Upon the project's closure, you will receive a downloadable link valid for 30 days.
2. In the case of an Autonomie run, your data will be stored for a single year only. Upon availability of your sequencing run on ZEUS (typically within 24 hours), a downloadable link will be sent to you. If you do not receive any communication within 24 hours, please contact biomics-bioinfo@pasteur.fr.
3. If your project was executed through externalization as an independent run, like an autonomous run, the data will be retained for one year.

data origin	ZEUS	GAIA
Biomics	2	3
Autonomie	1	X
Autonomie (externalised)	1	X

Table 2: Data retention and storage place (in years). Policy Jan 2023. Data generated by Biomics are stored 5 years (2 on Zeus, 3 on Gaia)

The Biomics Platform also stores raw data for Illumina, Nanopore, and Pacbio (if available) before demultiplexing. This data is usually large and not sent by default, but can be requested by emailing biomics-bioinfo@pasteur.fr. The raw data is retained for only one year.

If you fail to download the data set within the given time frame, you will be required to fill a request on the PPMS booking system at <https://ppms.eu/pasteur/?Omic> (1h of expertise: 'bioinformatics re-analysis'). This policy may change over time depending on the volume of data generated and the resources available to the Biomics Platform.