HORIZON-EIC-2022-PATHFINDERCHALLENGES-01



Project No. 101115115

Interoperable end-to-end **P**latform of scalabl**E** and sustain**A**ble high-th**R**oughput techno**L**ogies for **DNA**-based digital data storage

Deliverable 5.8
RP2 Update: Data Management Plan

WP5 – Coordination and Innovation Management

Authors	Stephanie Kristin Schröder (LUH), Jan Voges (LUH)
Lead participant	LUH
Delivery date	08 September 2026
Dissemination level	PU = Public
Type	Data Management Plan

Version 1.0

Revision History

Author(s)	Description	Date
Jan Voges (LUH)	Draft deliverable	01.07.2026
Jan Voges (LUH)	Draft 2	11.08.2026
Luka Zupancic (BioSis)	Revision 1	26.08.2026
Jan Voges (LUH)	Final version	26.08.2026
Andreia Cruz (accelCH)	Formal check	08.09.2026

Contents

Revision History 2
Contents 3
Partner Short Names 4
Abbreviations 4
Executive Summary 5
1 Data Summary 6
1.1 Purpose of Data Collection and Generation 6
1.2 Types and Formats of Project Data 7
1.3 Reuse of Data 7
1.4 Data Origin and Scale 8
2 FAIR Data Management Implementation 9
2.1 Data Findability 9
2.1.1 Metadata Provisions 9
2.1.2 Naming Conventions and Unique Identifiers 9
2.1.3 Version Numbering 9
2.2 Data Accessibility 9
2.2.1 Internal Repository and Working Storage 9
2.2.2 Public Accessibility and Access Restrictions 10
2.2.3 Open-source Code 10
2.3 Data Interoperability 10
2.4 Data Licensing and Reusability 10
2.5 Reproducibility and Provenance 11
3 Data Release, IP and Open Science Decision Process 11
4 Allocation of Resources 11
5 Data Security 12
5.1 Data Confidentiality and Integrity 12
5.2 Data Privacy and Protection 12
5.3 Data Availability, Backup, and Recovery 12
6 Data Storage and Preservation 12
6.1 Data Storage 12
6.2 Data Preservation 12
7 Roles and Responsibilities 13
7.1 Responsibilities 13
7.2 Roles 13
8 Ethical and Legal Aspects 14

9 References..... 14

Partner Short Names

No.	Organization	Short name
1	**BioSistemika**	BioSis
2	**Gottfried Wilhelm Leibniz Universität Hannover**	LUH
3	**Imagene SA**	IMG
4	**Technische Hochschule Wildau**	THWi
5	**Haute École Specialisée de Suisse Occidentale**	HES-SO
6	**accelopment Schweiz AG**	accelCH

Abbreviations

Abbreviation	Term
DMP	Data Management Plan
DOI	Digital Object Identifier
EC	European Commission
EU	European Union
FAIR	Findability, Accessibility, Interoperability, and Reusability
GDPR	General Data Protection Regulation
HE	Horizon Europe
IP	Intellectual Property
IPR	Intellectual Property Rights
R&D	Research and Development
RDM	Research Data Management
WP	Work Package

Executive Summary

Background & Objectives

Deliverable "5.8 – Data Management Plan" (DMP), as an update of Deliverable 5.2, outlines the updated strategy for data collected, stored, organised, analysed, and shared during the PEARL-DNA project. The consortium manages diverse data streams in compliance with robust research data management (RDM) standards and regulatory requirements.

Methodology & Implementation

In alignment with the Horizon Europe (HE) open-science requirements, the consortium manages research data in accordance with the principles of Findability, Accessibility, Interoperability, and Reusability (FAIR) [1]. Research data is managed according to the principle of "as open as possible, as closed as necessary". Openness is assessed for each dataset rather than applying a blanket confidentiality classification. Restrictions may be applied where justified by intellectual property rights (IPR), confidentiality, personal-data protection, security, third-party rights, or other legitimate constraints. Where access is restricted, the reason, scope and expected duration of the restriction is documented.

Outcomes

This document defines the implementation frameworks for FAIR data management throughout the data lifecycle. It introduces a dataset-level management approach, a pre-publication IP and data-release review, defined metadata and versioning practices, a repository strategy, measures for reproducibility and provenance, and clearer responsibilities for data release,

Impact

The data management strategies outlined in this plan aim to maximise the scientific and societal value of PEARL-DNA research outputs while protecting legitimate commercial, contractual, intellectual-property and privacy interests.

1 Data Summary

1.1 Purpose of Data Collection and Generation

The aim of this document is to enhance the development of PEARL-DNA data storage technology through comprehensive data collection and generation activities. The project gathers data from external sources, such as academic literature and market studies, to gain insights into the current technological landscape and the potential for technological advancements. Such external data is reused only where the applicable licence, terms of use, copyright/database rights and contractual conditions permit such use and, where relevant, redistribution.

Internally generated data stems primarily from the project's research and development (R&D) efforts, aiming to improve the technology and leverage the project team's expertise. This data is detailed in Table 1.

The DMP encompasses the entire lifecycle of research data within the PEARL-DNA project (Figure 1), from initial data collection and production to processing, analysis, preservation, and, ultimately, making data available for further use. Raw data generated from diverse sources are processed and transformed into usable formats, including documents, datasheets, code, and reports, adhering to standardised naming and metadata conventions agreed upon by the consortium.

In terms of data access, the project assesses the potential for external accessibility and the feasibility of open access for each dataset. This assessment is repeated when the dataset is prepared for publication or deposition and when relevant circumstances change, for example following an IP assessment. Publicly available data is made accessible for verification and further exploration by the wider community, facilitating transparency and fostering innovation.

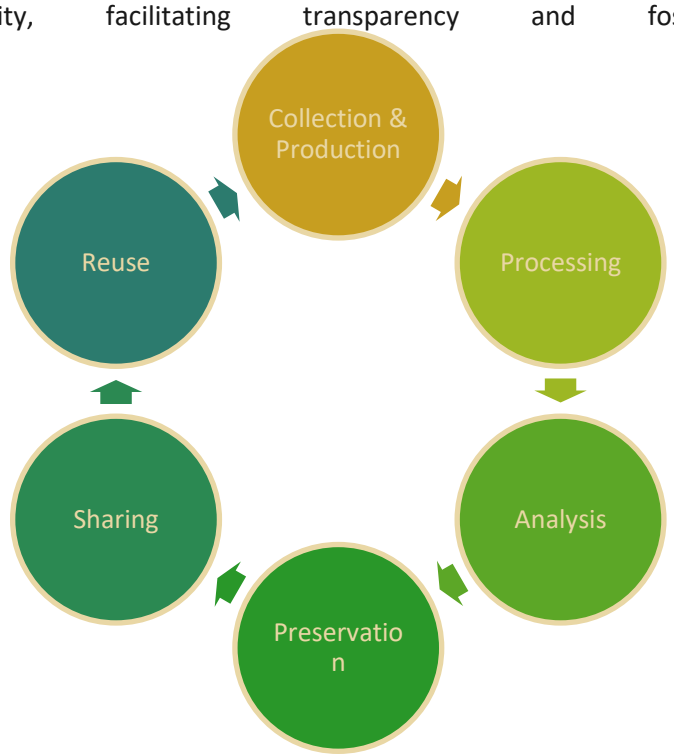


Figure 1: Diagram of the data cycle of the PEARL-DNA project.

1.2 Types and Formats of Project Data

PEARL-DNA gathers and accumulates a broad spectrum of data from multiple sources (both internal and external) throughout the R&D process. The data types and formats are outlined in Table 1. To ensure long-term usability, the consortium follows the file format recommendations provided by ETH Zurich's guidelines [2]. These formats include a broad range of data types, such as text documents, spreadsheets, presentations, images, audio, and video files, ensuring compatibility with preservation standards.

Table 1: Types and formats of data produced and collected during the PEARL-DNA project.

Dataset description	Data Type	Data Format	Associated WPs
Digital data input for DNA encoding	Digital data	TXT, BIN	WP1, WP2, WP3, WP4
DNA oligonucleotide library	Digital data, DNA	FASTA, FASTQ	WP2, WP4
DNA sequencing data	Digital data	FASTQ	WP1, WP2, WP3, WP4
Employee data	Documentation	CSV, XLSX, DOCX, PDF	WP5
IPR documentation	Documentation	DOCX, CSV, XLSX, PDF	WP5
R&D insights	Documentation	DOCX, CSV, XLSX, PDF	WP1, WP2, WP3, WP4
Experimental insights	Documentation	XLSX, CSV, Electronic lab notebooks	WP1, WP2, WP3, WP4
Technical specifications & drawings	Documentation	PDF, CAD Formats	WP1, WP2, WP3, WP4
Project management documentation	Documentation	XLSX, DOCX, PDF	WP5
Internal processes documentation	Documentation	DOCX, PDF	WP5
Marketing & promotional material	Digital data	PPT, DOCX, PDF	WP6
Market research documentation	Documentation	DOCX, CSV, XLSX, PDF	WP6
Legal documentation	Documentation	DOCX, PDF	WP5

1.3 Reuse of Data

The datasets prioritized for public release following scientific, quality, IP, confidentiality, and legal review include:

- DNA sequencing data
- Experimental protocols
- Marketing and promotional material

The consortium assesses the public reusability benefits of all data types while considering their sensitivity and the potential risks they pose to partners if fully disclosed. Attention is given to commercial interests, trade secrets, system design features, personal data, third-party rights and IPR protection.

Where a dataset cannot be fully opened, the consortium considers proportionate alternatives, including delayed release, restricted access, release of a redacted/aggregated/derived dataset, or release of metadata sufficient to make the existence and nature of the research output findable. The reason for restricting access and, where feasible, the expected duration of the restriction is documented.

For potentially exploitable results, publication or data release is coordinated with the project's IP management process. Where patent protection is considered appropriate, the relevant IP assessment and filing decision takes place before public disclosure that could affect novelty. After the relevant protection decision, the dataset is reassessed for public release.

Public disclosure may also be deliberately selected as part of the project's IP strategy where establishing prior art is considered appropriate. Such a decision is made deliberately and in coordination with the Innovation Manager and relevant consortium partners; public disclosure is not treated as a generic substitute for IP protection.

The DNA oligonucleotide library constitutes a core element of the PEARL-DNA technology and is therefore not expected to be made publicly available in the near term due to its commercial and IP sensitivity. However, selected information or the library itself may need to be disclosed at a later stage where required to demonstrate compliance with applicable data security, biosafety and biosecurity standards and good practices, including frameworks such as ISO/IEC 27001 and IBBIS guidance.

DNA sequencing data generated for technology development and validation may be made publicly available where appropriate to provide evidence of platform performance, reproducibility and data integrity, thereby supporting scientific validation and future end-user trust. Sequencing data originating from customers, end-users or other confidential third parties will not be publicly disclosed and will be handled in accordance with applicable confidentiality, data protection and information security requirements, including GDPR and ISO/IEC 27001.

1.4 Data Origin and Scale

Table 2 presents the sources of the data, along with indicative volumes, corresponding to the datasets outlined in Table 1.

Table 2: Origin and volume of data produced and collected during the PEARL-DNA project.

Dataset description	Data origin	Volumetric scale
Digital data input for DNA encoding	Internal & External	B – GB
DNA oligonucleotide library	Internal & External	B – kB
DNA sequencing data	Internal	GB – TB
Employee data	Internal	MB
IPR documentation	Internal	MB – GB
R&D insights	Internal	GB – TB
Experimental insights	Internal	GB – TB
Technical specifications & drawings	Internal	MB – GB

Project management documentation	Internal	MB – GB
Internal processes documentation	Internal	MB – GB
Marketing & promotional material	Internal & External	MB – GB
Market research documentation	Internal & External	MB – GB
Legal documentation	Internal & External	MB

2 FAIR Data Management Implementation

2.1 Data Findability

2.1.1 Metadata Provisions

Metadata structures are generated systematically alongside research assets to maintain structural provenance. At minimum, dataset metadata includes, where applicable: title, description, creator(s), organisation, project acronym and Grant Agreement number, WP/task, creation date, version, keywords, methodology or protocol reference, data format, processing status, access conditions, licence, related publications, related software, persistent identifiers and provenance information. Automated logs may additionally store data origin, timestamps and processing events.

For public datasets, repository metadata is made openly available to the extent permitted by legitimate interests and constraints.

2.1.2 Naming Conventions and Unique Identifiers

File hierarchies follow a unified naming convention established by the consortium's internal naming guidelines. Where applicable, names incorporate project, WP/dataset identifier, descriptive name, version and date. Dates use ISO format (YYYY-MM-DD) and version numbers use a consistent format such as v01, v02, etc. Spaces and ambiguous special characters are avoided.

A typical research-data filename may follow the pattern: PEARLDNA_WP02_<dataset>_v01_2026-08-11.<ext>. Filenames are operational aids and do not replace machine-readable metadata.

DOIs or other persistent identifiers are used for publicly deposited research outputs. Related datasets, publications, software and protocols are linked where possible.

2.1.3 Version Numbering

Research datasets are versioned at the dataset level. The dataset version recorded in the dataset metadata and repository record is authoritative. File-system or Google Drive revision histories may be used for operational recovery and collaboration but do not replace the dataset-level version.

2.2 Data Accessibility

2.2.1 Internal Repository and Working Storage

During the project, Google Drive/Google Workspace serves as the consortium's collaborative working-storage environment. It is not considered the long-term public research-data repository.

Access follows the principle of least privilege. Team members receive access according to their tasks and responsibilities. Access rights are reviewed periodically and removed when a person no longer requires access.

2.2.2 Public Accessibility and Access Restrictions

Beneficiaries of HE must manage research data in accordance with FAIR principles and the principle of "as open as possible, as closed as necessary". PEARL-DNA therefore does not apply a blanket confidential classification to all research data. Each dataset is assessed individually.

The assessment considers, at minimum: IP and patentability; confidentiality and trade secrets; personal-data protection; security; third-party rights; contractual obligations; and scientific or quality considerations. Where no legitimate restriction applies, data is made openly accessible as soon as possible and within the applicable project timelines.

Where access is restricted, the repository record remains findable and states the applicable access condition and reason for restriction. Restrictions may include delayed/embargoed access, controlled access, or permanent non-disclosure where justified.

Public datasets are deposited in a trusted repository as soon as possible and within the timeline established by the project and Grant Agreement requirements. The preferred repository is a suitable institutional or disciplinary repository where available; Zenodo may be used where no more appropriate trusted repository is identified. Repository selection considers domain suitability, persistent identifiers, metadata quality, access controls, preservation arrangements and interoperability.

2.2.3 Open-source Code

Where appropriate, software code developed for dissemination in the public domain is released under an open-source licence selected according to the project's IP position, third-party dependencies and intended reuse. Source code is maintained in a version-controlled repository where appropriate. Releases are version-tagged and, where suitable, archived in a trusted repository such as Zenodo to obtain a DOI.

The code release should include sufficient documentation to support reuse, including installation or execution instructions, relevant dependencies and, where necessary, configuration or example files. Software is not released where doing so would conflict with legitimate IP, confidentiality, security or third-party licensing constraints.

2.3 Data Interoperability

Data designated for public access are shared in standard, open or widely supported formats that are compatible with both commercial and open-source software, where such formats are suitable for the scientific purpose.

Sequence data uses established sequence formats where appropriate. Structured tables use machine-readable tabular formats where feasible. Metadata and documentation use formats suitable for long-term access and reuse.

2.4 Data Licensing and Reusability

Datasets intended for public access are assigned an appropriate open licence on a dataset-by-dataset basis. CC BY 4.0 may be used where attribution is appropriate and legally possible. CC0 may be considered where rights can appropriately be waived or dedicated to the public domain. The choice considers copyright, database rights, partner ownership, contractual restrictions and third-party rights.

The consortium does not apply a PEARL-DNA open licence to third-party material unless the applicable rights permit such licensing. Where data cannot be openly licensed, the repository record describes the applicable access or reuse conditions.

Publicly released software is licensed separately under an appropriate open-source license.

2.5 Reproducibility and Provenance

Where research data underpin publications or major project conclusions, the consortium retains or provides access to the data, protocols, software/code and other research outputs needed to understand and, where reasonably possible, reproduce or validate the reported results, subject to legitimate restrictions.

3 Data Release, IP and Open Science Decision Process

To balance open science and protection of legitimate interests, the following decision process is applied before public release of research datasets or associated software where material IP or confidentiality considerations may arise:

1) The data producer and relevant WP lead identify the dataset, its purpose, status, quality and intended reuse.
2) The Data Manager verifies that the dataset has adequate metadata, versioning, provenance and documentation.
3) The WP lead and Innovation Manager assess whether the dataset contains potentially exploitable results, confidential information, trade secrets or other IP that requires protection.
4) Where patent or other IP protection may be appropriate, the relevant protection decision is made before public disclosure that could affect the ability to obtain protection.
5) The consortium assesses constraints regarding personal data, third party rights, contractual matters, and security.
6) The dataset is classified as open, open after embargo, controlled/restricted, or not for public release, with the reason documented.
7) An appropriate repository, metadata record and license/access condition are selected.
8) The dataset is deposited and linked to related publications, software and protocols where applicable.
9) The decision is revisited when circumstances change, for example after an IP filing, publication or expiry of a confidentiality period.

Where public disclosure is intentionally selected to establish prior art, this is a deliberate IP-management decision rather than an automatic consequence of the DMP.

4 Allocation of Resources

The costs for FAIR data management include the costs of the cloud data storage and management provider, trusted repository deposition where applicable, and personnel engaged in data collection and management. These expenses mainly include, but are not limited to:

- developing and updating the DMP;
- setting up and maintaining the internal data-management system, including access rights;
- onboarding project team members;
- understanding FAIR data-management principles and applicable regulations;
- defining the consortium's data-management policy;
- implementing nomenclature, metadata and versioning practices;
- data collection and quality control;
- data storage and backup;
- metadata generation;
- transfer to trusted public repositories;
- software/code release and archival where applicable; and
- long-term preservation and maintenance activities within the project's scope.

The total expense associated with FAIR data management, as outlined in the reported cost items, is accounted for and included in the budget for WP5, as detailed in the Grant Agreement.

5 Data Security

5.1 Data Confidentiality and Integrity

Sensitive data is treated according to its assessed confidentiality and access requirements. Each consortium member is responsible for ensuring the confidentiality and security of the sensitive data they handle by adhering to the consortium's data management guidelines and applicable institutional requirements. The consortium is tasked with keeping these data-management policies current, in compliance with relevant regulations, and enforcing them effectively.

5.2 Data Privacy and Protection

The project does not involve human or patient genetic data as part of its research activities. Nevertheless, personal data may be processed for project administration, contact management and employment-related purposes. Such processing is limited to what is necessary for the relevant purpose and is handled in accordance with the GDPR, applicable national legislation and the responsible beneficiary's institutional policies.

5.3 Data Availability, Backup, and Recovery

During the project, data are stored and shared within the consortium's authorised Google Drive/Google Workspace environment, which serves as the collaborative working-storage environment. Access is restricted to authorised personnel according to their tasks and responsibilities, following the principle of least privilege. Public datasets are deposited in trusted external repositories in accordance with Section 2.2.2 and are therefore not dependent on the internal working environment for long-term public accessibility.

Google Workspace provides the project's active data-storage and collaboration environment, including the service's available redundancy, version history and recovery capabilities. Access to project data is protected through the authentication and access-control mechanisms configured for the consortium's Google Workspace environment. External sharing is restricted to authorised cases.

6 Data Storage and Preservation

6.1 Data Storage

During the project, research data are stored in the PEARL-DNA internal data-management environment, with Google Drive/Google Workspace serving as the consortium's collaborative working-storage environment where appropriate. Data is organised using the agreed folder structure, naming conventions, metadata and access controls.

The internal system is intended for active project work and collaboration. It is not the sole preservation mechanism for research data that are intended for long-term public access.

6.2 Data Preservation

After the project, data intended for public access will be available through trusted external repositories selected according to the criteria in Section 2.2.2. Examples include Zenodo and, where scientifically appropriate, disciplinary or institutional repositories such as Dryad.

Public datasets will be deposited as soon as possible and within the applicable project and Grant Agreement timelines. Repository records will contain sufficient metadata, persistent identifiers and licensing/access information to support long-term findability and reuse.

In compliance with applicable Horizon Europe requirements, the consortium will retain publicly deposited research data for at least the applicable minimum period following completion of the project and in accordance with the Grant Agreement and repository policy.

Furthermore, embracing the innovative potential of the PEARL-DNA project, the consortium investigates the use of DNA as a demonstration medium for critical-data backup. This experimental activity complements conventional repository-based preservation.

7 Roles and Responsibilities

The PEARL-DNA project maintains clear responsibilities for data creation, quality, management, release and preservation. WP leads remain responsible for data-management compliance within their WPs, while the Data Manager coordinates the cross-project implementation of this DMP.

7.1 Responsibilities

In the PEARL-DNA project, the WP leads are responsible for ensuring adherence to these guidelines and all relevant data-management laws and regulations within their WP. Each consortium member bears responsibility for the datasets that it creates, obtains or collects during the project, including appropriate quality, documentation, security and rights management.

7.2 Roles

Several roles are assigned within the PEARL-DNA project to ensure its smooth functioning. These roles are listed in Table 3.

Table 3: Roles within the PEARL-DNA project.

Role	Name
WP1 Lead	Jan Voges (LUH)
WP2 Lead	Tjaša Stare (BioSis)
WP3 Lead	Marthe Colotte (IMG)
WP4 Lead	Tomaž Karčnik (BioSis)
WP5 Lead	Tomaž Karčnik (BioSis)
WP6 Lead	Miriam Frances, Anastasiia Aksonova (accelCH)
WP7 Lead	Tomaž Karčnik (BioSis)
Data Manager	Jan Voges (LUH)
Project Coordinator	Tomaž Karčnik (BioSis)
Project Deputy Coordinator	Mirjana Oblak (BioSis)
Innovation Manager	Jörn Glökler (THWi)

The Project Coordinator oversees the entire PEARL-DNA project, ensuring it remains on track with its objectives, budget, and timeline. He acts as the main point of communication between the project team, funding bodies, and external stakeholders. The Project Deputy Coordinator supports the Project

Coordinator in administrative and operational tasks, ensuring smooth project progression and assisting in conflict resolution.

The Data Manager is tasked with overseeing the data management lifecycle, from collection through analysis, preservation, and release, ensuring adherence to the DMP and FAIR data management principles. The Data Manager coordinates with WP leads to facilitate efficient data handling and resolve any data-related issues.

The Innovation Manager is responsible for managing the innovation pipeline, protecting IP, and facilitating the transfer of technology and knowledge generated by the project.

8 Ethical and Legal Aspects

The PEARL-DNA research activities focus on synthetic DNA fragments rather than human or patient genetic data. Therefore, no ethical concerns related to the processing of human genetic research data have been identified. Nevertheless, the project may process personal data for administration and employment-related purposes. Such processing is subject to the applicable data-protection and legal requirements described in Section 5.2.

9 References

[1] M. D. Wilkinson et al., “The FAIR Guiding Principles for scientific data management and stewardship,” Scientific Data, vol. 3, no. 1, p. 160018, Mar. 2016, doi: 10.1038/sdata.2016.18.

[2] ETH Zürich, “File formats for archiving,” ETH Library Research Data Management, 2024. [Online]. Available: https://unlimited.ethz.ch/display/DD/File+formats+for+archiving. [Accessed: 11-Aug-2026].