HORIZON-EIC-2022-PATHFINDERCHALLENGES-01

PEARL DNA

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 5.2
Data Management Plan

WP5 – Coordination and innovation management

Authors	Stephanie Kristin Schröder (LUH)
Lead participant	LUH
Delivery date	27 March 2024
Dissemination level	PU = Public
Type	Data Management Plan

Version 1.0

European Innovation Council

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Innovation Council and SMEs Executive Agency (EISMEA). Neither the European Union nor the granting authority can be held responsible for them. Swiss participants in this project are supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract numbers 23.00332 and 23.00328.

Revision history

Author(s)	Description	Date
Stephanie Kristin Schröder (LUH)	Draft deliverable	19.02.2024
Luka Zupancic (BioSis)	Revision 1	27.02.2024
Stephanie Kristin Schröder (LUH)	Draft 2	12.03.2024
Luka Zupancic (BioSis)	Revision 2	14.03.2024
Stephanie Kristin Schröder (LUH)	Draft 3	20.03.2024
Anna Ziemele, Andreia Cruz (accelCH)	Formatting and formal check	22.03.2024
Stephanie Kristin Schröder (LUH)	Final version	27.03.2024

Contents

Partner short names

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

Abbreviations

Abbreviation	Term
DMP	Data Management Plan
EC	European Commission
EU	European Union
HE	Horizon Europe
IPR	Intellectual Property Rights
R&D	Research and Development
RDM	Research Data Management
WP	Work Package

Executive Summary

Background

Deliverable 5.2 is a crucial component of Work Package 5, which outlines the strategy for managing project data throughout the duration of the project. This deliverable details the procedures, tools, and techniques used to collect, store, organise, and analyse data. It also includes guidelines on data security, privacy, and access control to ensure that sensitive information is protected. Effective project data management is essential for ensuring that project objectives are met, and outcomes are achieved within set timelines and budget constraints. Furthermore, effective data management also forms the basis for high-quality research and publications.

Objectives

Deliverable "D5.2 – Data Management Plan" aims to define, project, and outline the data management strategy concerning the PEARL-DNA project. Throughout the project, the consortium will collect and manage various types of data from different sources in compliance with good data management practices and regulatory requirements. Furthermore, the project aims to share valuable insights with the public, except for sensitive data such as personal information, confidential data, and intellectual property. The consortium seeks to define and implement strategies that ensure the highest level of compliance with the FAIR principles of data management and dissemination. It's important to note that the Data Management Plan (DMP) is a dynamic document that may be revised and updated as the project progresses.

Methodology and implementation

According to the latest Guidelines on FAIR Data Management under Horizon Europe, as released by the EC Directorate-General for Research & Innovation, all participants are mandated to ensure research data management (RDM) in line with the principles of **Findability**, **Accessibility**, **Interoperability**, and **Reusability** (FAIR) [1]. This requirement underscores the importance of effective data management as a foundational element for fostering knowledge discovery, innovation, and the integration and reuse of subsequent data and knowledge.

FAIR data management is integral to the European Commission's strategy for enhancing the openness and reuse of research data produced within Horizon Europe projects. This approach is designed to optimize access to and utilization of research data while also addressing considerations related to the openness and protection of scientific information, commercialisation, Intellectual Property Rights (IPR), privacy, security, and data management and preservation.

With the inception of Horizon Europe, the commitment to FAIR data principles has been broadened compared to its predecessor, Horizon 2020. The initiative now encompasses all thematic areas of Horizon Europe, signifying a universal application of open access to research data, subject to the European Commission's acknowledgment that certain circumstances necessitate restrictions on data sharing.

The emphasis on FAIR data primarily concerns the datasets necessary for validating the findings reported in scientific publications. Nevertheless, Horizon Europe encourages beneficiaries to share additional data voluntarily, as outlined in their Data Management Plans. Moreover, expenses incurred in making research data openly accessible can be accounted for as eligible costs within any Horizon Europe grant, reflecting the program's comprehensive support for open science practices.

By expanding the scope and reinforcing the framework for data management from Horizon 2020 to Horizon Europe, the European Commission aims to advance scientific research and innovation further, ensuring that research data management practices contribute effectively to the broader objectives of transparency, accessibility, and societal benefit.

Outcomes

DMPs are crucial for effective data management. This DMP outlines the data management lifecycle for data that will be collected, processed, and/or produced over the course of the PEARL-DNA project. To ensure that research data is findable, accessible, interoperable, and reusable (FAIR), this document provides details on:

- the handling of research data during & after the end of the project;
- what data will be collected, processed, and/or generated;
- which methodology & standards will be applied;
- whether data will be shared/made open access;
- how data will be curated & preserved (including after the end of the project).
-

Impact

The data management strategies outlined in this plan ensure that the knowledge and technologies developed will be accessible, reusable, and beneficial to a broad range of stakeholders, underscoring the project's commitment to promoting open science and collaborative research.

Next steps

Updating the DMP is crucial throughout the project, especially when significant alterations occur. Such changes may encompass, but are not limited to, the introduction of new data, modifications in consortium policies (like identifying new potential for innovation or opting to apply for a patent), and shifts in the consortium's structure or external circumstances (such as the arrival of new consortium members or the departure of existing ones).

The DMP will undergo regular revisions as part of the project's evaluation routine. Given the absence of other scheduled reviews as per the grant agreement, it is imperative to ensure the DMP is updated by the final evaluation at the latest.

1 Data Summary

1.1 Purpose of Data Collection and Generation

The aim of the PEARL-DNA project is to enhance the development of PEARL-DNA data storage technology through comprehensive data collection and generation activities. The project will gather 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. This external data will support the project's understanding of the need for technological development.

Internally generated data will stem 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, detailed in Table 1 under 'Dataset description', will include DNA data synthesised from both internal research findings and external inputs.

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. The project will start by gathering and generating raw data from diverse sources. This data will be 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 will assess the potential for external accessibility and the feasibility of open access for each dataset. This evaluation will guide the decision-making process regarding the public availability of data in the final stage of data reuse. Publicly available data will be accessible for verification and further exploration by the wider community, facilitating transparency and fostering innovation.

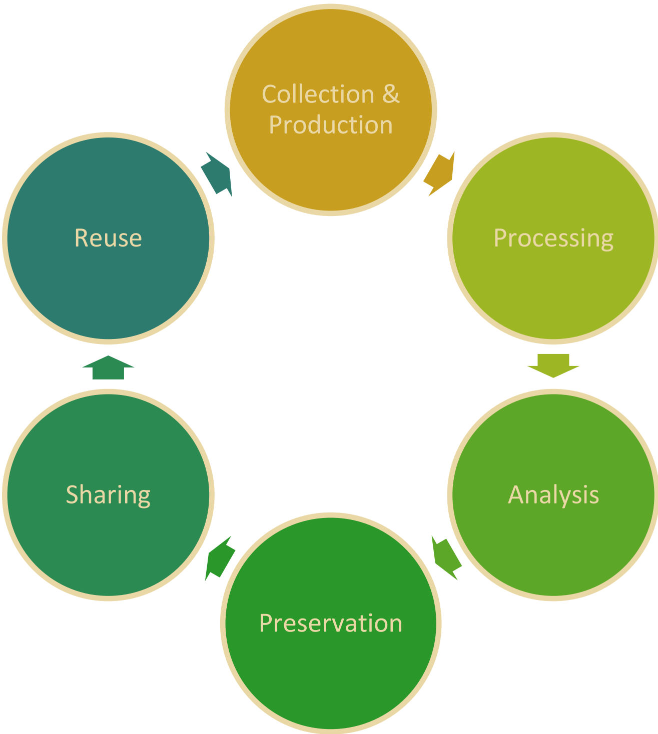

Figure 1: Diagram of the data cycle of the PEARL-DNA project.

1.2 Types and Formats of Data

PEARL-DNA will gather and accumulate a broad spectrum of data from multiple sources (both internal and external) throughout the R&D process. These anticipated data types and formats are outlined in Table 1. The majority of this data will be archived in either physical or electronic formats. To ensure that our data remains accessible, preservable, and usable over the long term, we will follow 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 that we cover all bases and remain compatible with preservation standards. Our adherence to these guidelines demonstrates our commitment to maintaining the integrity of our data, supporting future research, and upholding the principles of open and accessible academic and scientific inquiry.

The DMP is a dynamic document that will undergo revisions throughout the project's lifecycle, especially in terms of the datasets that are amassed and created. Future revisions of this document will reflect any changes, including additions or removals of data.

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

Dataset description	Data Type	Data Format	Associated WPs
DNA oligonucleotide library	Digital data, DNA	FASTA, FASTQ	WP2, WP4
Digital data input for DNA encoding	Digital data	TXT, BIN	WP1, WP2, WP3, WP4
Encoded DNA	DNA	XML, JSON	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

It is currently challenging to make accurate predictions about which data will be reused in the public domain during and after the project. However, we can make some assumptions about which datasets are more likely to be reused based on their sensitivity and value to the general public. These datasets likely include:

- Encoded DNA
- IPR documentation
- Experimental insights (limited)
- Experimental protocols and standard operating procedures (SOP)
- Market research documentation
- Marketing & promotional material

The consortium will assess the public reusability benefits of all data types while considering their sensitivity and the potential risks they pose to partners if fully disclosed. A key focus will be on addressing potential conflicts related to commercial interests, trade secrets, and intellectual property rights (IPR) protection.

It's not mandatory for the consortium to grant open access to all research and development insights gained throughout the project, especially if achieving specific project milestones or deliverables could be compromised by such disclosure. This applies particularly to confidential information such as trade secrets, system design features, product management documents, IPR, etc. In these instances, the consortium will explore three protective measures:

1) retaining the data for internal use only.
2) seeking a patent to leverage the innovation while potentially realising financial benefits commercially.
3) releasing the research data to the public to establish prior art.

1.4 Origin and Volume of Data

Throughout the project's duration, the consortium will produce, collect, and amass data from a variety of internal and external origins. At present, we expect a significant portion of the data obtained from external entities to stem from publicly accessible sources. Forecasting the exact sources and the volume of data that will emerge over the next three years poses a challenge. Therefore, we will update this DMP as and when we come across a new dataset that was not initially anticipated, either through generation or acquisition. Nonetheless, we are able to accurately identify the sources and estimate the volume of the data corresponding to the datasets already outlined (refer to Table 1). Table 2 presents the sources of the datasets we have identified to date, along with their estimated volumes.

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

Dataset description	Data origin	Volumetric scale
DNA oligonucleotide library	Internal & External	B – kB
Digital data input for DNA encoding	Internal & External	B – GB
Encoded DNA	Internal	B – GB
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

1.5 Data Utility

The data produced, gathered, and compiled during the PEARL-DNA project will play a crucial role in enabling the consortium to achieve all established project objectives and deliverables, as well as in formulating a successful approach for further development, integration, prototyping, and eventual commercial launch in the future. Beyond their internal significance, certain datasets could prove beneficial to the broader community, especially for researchers and developers active in various fields, among others:

- DNA synthesis
- PCR/LCR and other forms of DNA assembly
- liquid handling & microfluidics
- DNA sequencing
- biological storage media and data encryption
- inkjet technology
- data storage and data archiving
- data preservation
- data protection

2 FAIR Data

2.1 Data Findability

2.1.1 Metadata Provisions

Metadata plays a crucial role in data management by offering context about the data. Where relevant, metadata will be employed to reveal essential information about the data produced. This encompasses details like the data's origin, the moment of creation, the settings applied during its generation, and the apparatus used for its creation. Access to this information simplifies the process of tracing the data's provenance and history, thus improving transparency and accountability.

2.1.2 Naming Conventions and Unique Identifiers

In data repositories, folders, and files will follow a versioning and structuring system based on a naming convention established by the consortium's internal naming guidelines. Typically, the name of each project document or folder will incorporate the following elements:

- Name
- Project name
- Dataset (where applicable)
- Version
- Date

DOIs will be used as persistent identifiers in open data repositories when applicable.

2.1.3 Search Keywords

As a part of our efforts to improve the accessibility of datasets, we will be adding relevant search keywords to both the data and metadata. These keywords will be chosen based on the content of the datasets and will help quickly discover relevant datasets. By supplementing both data and metadata with these search keywords, we aim to make it easier to find and access the information needed. DNA data storage is a new and rapidly evolving field that lacks standardised terminologies and ontologies. As with other scientific and technical disciplines, terminologies and ontologies evolve as the field matures and the consensus is reached among researchers and practitioners. However, no consensus has yet been reached for DNA data storage. It is possible that efforts to develop standardisation and ontology are underway due to the fast-paced nature of this field. If any ontologies or terminologies are established, we will adhere to them.

2.1.4 Version Numbering

In general, the files stored in depositories will be versioned at two levels:

1. Via the naming convention (e.g., "V1", "V2", ... "VX") and the use of the date as the suffix, indicating the date of last modification.
2. Via the inherent capabilities of our internal data management system (Box) to version all uploaded files.

2.2 Data Accessibility

2.2.1 Repository

All project data will be kept in the PEARL-DNA Box, to which every project team member will have access. Depending on their unique tasks and responsibilities within the project, team members may be granted differing levels of access.

Box is a cloud-based, encrypted data management & storage system that also provides easy sharing of files across platforms and devices. The stored data is encrypted using 256-bit AES encryption.

2.2.2 Public Accessibility

Beneficiaries of Horizon Europe must ensure that research data generated in their projects are made openly accessible as much as possible while also being kept confidential when necessary.

The data will initially be classified as confidential and will not be available to the public to protect sensitive information and intellectual property rights and comply with data protection regulations. This is essential for preserving the opportunity to patent innovations and for ensuring that personal and proprietary information is securely managed. The confidential classification aligns with the Horizon Europe directive of being 'as closed as necessary' to prevent unauthorized access and misuse of critical data.

Once the patenting process is completed or if it is determined that certain datasets no longer require confidentiality protection, then these datasets will be re-evaluated and potentially reclassified for public access. At this stage, they will be transferred to trusted external repositories that facilitate patent documentation and are also committed to open access and adhering to the FAIR principles. This transition signifies a shift towards the 'as open as possible' aspect of the principle, ensuring that valuable research data can be shared with the wider scientific community and the public, fostering innovation and collaboration. Priority will be given to repositories that offer the highest level of accessibility while ensuring the integrity and security of the data.

2.2.3 Open-source Code

Where appropriate, software code designed for effective dissemination in the public domain will be released under an open-source license, which is to be decided on suitable online repositories (such as GitHub). Additionally, we will utilise an archival tool such as Zenodo to archive our releases and obtain a DOI, ensuring our software is easily citable and further enhancing its accessibility and traceability in the academic and research communities.

2.3 Data Interoperability

We anticipate that datasets related to the genetic code (such as oligonucleotide libraries, encoded DNA sequences, etc.) and all research findings poised for potential publication or scientific sharing will be more suited for interoperability compared to other data collected or created during the PEARL-DNA project.

Data designated for public access will be shared in standard or open formats that are compatible with both commercial and open-source software, facilitating extensive data exchange among researchers and institutions.

A standard vocabulary for metadata descriptions will be employed; if this proves unfeasible, a mapping to more widely recognized ontologies will be provided. In such instances, we will seek specific technical input from experts in semantics and logic.

2.4 Data Reusability

2.4.1 Data Licensing

Datasets intended for public access will be licensed under Creative Commons Attribution International Public License (CC BY). For broader reuse, datasets may be licensed under Creative Commons Public Domain Dedication (CC0). Raw research data will be marked with a Creative Commons Public Domain Mark (PDM) or its equivalent unless they qualify for protection under copyright or database rights laws.

Where appropriate, software code designed for effective dissemination in the public domain will be released under an open-source license, which is to be decided on suitable online repositories (such as GitHub).

To address the critical aspects of data quality assurance and documentation necessary for reuse, we propose the inclusion of Readme files alongside the datasets. These files will serve as a repository for comprehensive documentation detailing the dataset's contents, structure, creation process, and any other relevant information to facilitate its use. Moreover, the publication of data papers offers a robust mechanism for emphasizing research data. As these documents undergo peer review, they inherently contribute to the quality assurance process, offering an in-depth look at the datasets, their utility, and methodologies. This approach ensures the datasets' integrity and reliability and enhances their visibility and understanding within the scientific community, promoting a culture of transparency and reproducibility in research.

2.4.2 Reusability by Third Parties

For the purpose of open access, datasets aimed for public availability will be licensed under the latest version of the Creative Commons Attribution International Public License (CC BY), ensuring attribution of authorship or under a license that provides equivalent rights. In cases where broader and unrestricted reuse is advantageous, such as with large datasets or at the discretion of the authors, datasets may be licensed under the Creative Commons Public Domain Dedication (CC0) or a similar license that waives all rights to the data. Raw research data should be marked with a Creative Commons Public Domain Mark (PDM) or its equivalent unless they qualify for protection under copyright or database rights laws.

Furthermore, to support the open dissemination of software code associated with the research, it will be made available under an open-source license via appropriate online repositories, ensuring its accessibility and reuse within the scientific community.

3 Allocation of Resources

3.1.1 Estimation of Costs

The costs for making data FAIR include the costs of the cloud data storage and management provider (Box) as well as the costs related to personnel engaged in data collection and management. These expenses mainly include, but are not limited to:

- Developing the DMP
- Setting up the internal data management system (Box)
- Setting up different layers of access rights
- Onboarding project team members
- Understanding the FAIR principles and applicable regulation
- Defining the consortium's data management policy
- Implementing nomenclature and versioning practices
- Data collection
- Data storage
- Metadata generation
- Transfer to public repositories

The total expense associated with the management of FAIR data, as outlined in the reported cost items, is accounted for and included in the budget for Work Package 5 – Coordination and Innovation Management, as detailed in the Grant Agreement.

4 Data Security

4.1 Data Confidentiality and Integrity

Sensitive data will be treated as highly confidential and for internal use only by default. 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. The consortium is tasked with maintaining these data management policies current, in compliance with all relevant regulations, and enforcing them effectively.

Entry to our internal data management system (Box) is controlled through access limitations, including the use of a username and password. Box also provides optional 2-factor authentication. This can be implemented if required to enhance the security of the cloud storage.

4.2 Data Privacy and Protection

The European Union has recently introduced the General Data Protection Regulation (GDPR). This new legislation imposes additional responsibilities on companies that handle data, with the aim of safeguarding the privacy of individuals whose data is being processed. Under GDPR, data subjects - that is, individuals whose data is being processed - are given greater control over their information. They are granted new rights, such as the right to access their data, the right to request that their data be deleted, and the right to object to their data being processed. The consortium is committed to complying strictly with GDPR. We take the protection of personal data very seriously and will work hard to ensure that the rights of data subjects are respected and upheld at all times.

4.3 Data Availability

Data will be kept and distributed within the project team using the consortium's exclusive, encrypted internal data management system (Box), which limits access to authorised personnel through a username and password. Initially, access to the cloud storage, where datasets and metadata are organized, will be exclusive to consortium members. When it becomes essential to share data with external parties (such as for publication, release to the public, collaboration with potential partners and subcontractors, etc.), reliable external data repositories will be employed for the secure long-term storage and maintenance of the data.

5 Data Storage and Preservation

5.1 Data Storage

During the course of this project, the research data will be deposited in the PEARL-DNA Box to which every project team member will have entry. Depending on their unique tasks and responsibilities within the project, team members may be granted differing levels of access.

Box, the internal data management system, implements a robust infrastructure, including comprehensive backup strategies, to ensure the safety and availability of user data. These strategies typically involve redundancy, data replication, and disaster recovery plans. Here's an overview of how Box might handle backups:

1. **Redundancy:** Box stores multiple copies of data across different servers and data centres. This redundancy ensures that if one copy of the data becomes inaccessible due to hardware failure or other issues, other copies remain available for access and restoration. Box uses data centres with reliable power sources and backup systems to offer 99.9% SLAs and redundancy.
2. **Data replication:** data replication involves creating and maintaining identical copies of data in multiple locations in real time. This ensures that in the event of a significant system failure or a disaster at one location, the data remains safe in another location.
3. **Versioning:** Box offers versioning capabilities for files, allowing users to restore previous versions of a file if necessary. This can be particularly useful for recovering from accidental deletions or changes.
4. **Regular testing:** reliable backup strategies also include regular testing of backup systems and processes to ensure data can be effectively restored when needed.

It is important to note that these backup strategies may vary and evolve over time.

5.2 Data Preservation

After the project is completed, any data that is classified as public access will be available through trusted external repositories. These repositories are also committed to open access and adhere to the FAIR principles. Some of the examples of such repositories include Dryad and Zenodo. After the conclusion of the PEARL-DNA project, the data will be available for a certain timeframe. The exact duration of this timeframe will be decided upon during the course of the project.

To ensure the longevity and accessibility of data, we will utilise a trusted repository and a backup layer with Box, our internal data management system. This streamlined approach is designed to protect information and keep it accessible and usable over time, adhering to best practices in data management. Our internal cloud storage will serve as a secure, efficient backup solution, complementing the external repositories to ensure comprehensive data protection against potential data loss scenarios.

Furthermore, embracing the innovative potential of the PEARL-DNA project, we will investigate the use of DNA as the ultimate medium for critical data backup. DNA data storage stands out for its unparalleled longevity, offering the capability to preserve information for millennia under proper conditions. This pioneering technique not only reflects our commitment to leading-edge data preservation methods but also leverages the intrinsic benefits of DNA technology. Incorporating DNA data storage into our preservation plan positions us at the vanguard of data management innovation, ensuring that our data remains safeguarded and accessible for future generations.

6 Roles and Responsibilities

The latest Guidelines on FAIR Data Management in Horizon Europe, issued by the European Commission's Directorate-General for Research & Innovation, mandate that all recipients ensure their research data is findable, accessible, interoperable, and reusable (FAIR) for proper management. Effective management of research data is not an end in itself but a crucial pathway to unlocking knowledge and innovation, facilitating the integration and reuse of data and knowledge.

6.1 Responsibilities

In the PEARL-DNA project, the Work Package (WP) leads will be responsible for ensuring adherence to these guidelines and all relevant data management laws and regulations within their WP. Each consortium member bears the responsibility to adhere to all relevant data management laws and regulations for the datasets that are created, obtained, or collected during the project's course. Adherence to these guidelines and regulations is critical for the success of the project and the broader scientific community. It ensures that research data is managed in a standardised, systematic, and transparent manner, making it easier for researchers to discover, access, and use data.

6.2 Roles

Several roles are assigned within the PEARL-DNA project to ensure its smooth functioning. These roles are listed comprehensively in Table 3. Moreover, the table also specifies the WP leads who are responsible for ensuring that all regulations and guidelines, including the DMP, are followed and adhered to within their respective WPs.

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	Margaryta Schaltegger (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)

As the project coordinator, Tomaž Karčnik 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. Mirjana Oblak, the project deputy coordinator, supports Tomaž Karčnik in administrative and operational tasks, ensuring smooth project progression and assisting in conflict resolution. As the Data Manager, Jan Voges is tasked with overseeing the data management lifecycle, from collection through to analysis, ensuring adherence to the DMP and FAIR principles. He coordinates with WP leads to facilitate efficient data handling and resolve any data-related issues. Jörn Glökler is responsible for managing the innovation pipeline, protecting intellectual property, and facilitating the transfer of technology and knowledge generated by the project.

7 Ethical Aspects

No ethical concerns regarding data management have been identified since the project will focus on synthetic DNA fragments instead of human/patient genetic data.

8 References

[1] M. D. Wilkinson, et al., "The FAIR Guiding Principles for scientific data management and stewardship, "Sci. Data, https://www.nature.com/articles/sdata201618#citeas, 2016.

[2] ETH Zürich, "File formats for archiving", https://unlimited.ethz.ch/display/DD/File+formats+for+archiving, 2024.