Document title: Data Management Plan

Version	Status	Date
1.0	Final	2023-11-21

Author
Pär Erik Martinsson

Contact
par-erik.martinsson@ltu.se, 0920-493637

Deliverable D12.1
Data Management Plan (DMP)

Work package leader: Pär Erik Martinsson
par-erik.martinsson@ltu.se

Abstract

This document constitutes deliverable D12.1 of the Arrowhead fPVN project. The Data Management Plan (DMP) outlines the life cycle of data within the AFPVN Project, covering its production, collection, processing, storage, and preservation throughout the project's development and beyond. It details how data will be collected, identifies the primary beneficiaries, and outlines the methods for storage and management within the project. Additionally, it addresses whether the project plans to make the data accessible, findable, and reusable. The document specifies the required resources for ensuring data openness and finalization, considering security and ethical aspects within the project context. This represents version 1 of the Data Management Plan, subject to updates as the project evolves to incorporate changes and improvements Project data management plan and report on established data management facilities..

Grant agreement no.	**101111977**
Project acronym	**Arrowhead fPVN**
Project full title	**Arrowhead flexible Production Value Network**
Dissemination level	PU
Due Date	30-11-2023
Date of Delivery	21-11-2023
Deliverable Number	D12.1
Deliverable Name	Data Management Plan
AL / Task related	LTU
Author/s	Pär Erik Martinsson,
Contributors	Lama Alkhaked
Reviewer	Jerker Delsing
Keywords	Data management plan, Data distribution, data privacy, fair data
Abstract	The Data Management Plan (DMP) outlines the life cycle of data within the AFPVN Project, covering its production, collection, processing, storage, and preservation throughout the project's development and beyond.

Table of contents

1. Abbreviations

Abbreviation	
AFPVN	Arrowhead Flexible Production Value Network
DMP	Data Management Plan
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
IPR	Intellectual Property Rights
DPO	Data Protection Officer
DPIA	Data Protection Impact Assessments

2. Introduction

The Arrowhead flexible Production Value Network (fPVN) project aims to establish autonomous and evolvable interoperability among stakeholders in the flexible production value network. It is built on three key pillars: the microservices paradigm, the utilization of major industrially accepted data models, and automatic translation between these data models. The project is expected to have a significant impact on manufacturing productivity and flexibility. This deliverable focuses on the management of the data in the project. The following data will be discussed and handled carefully through the DPM.

- Administration Data (OwnCloud*)
- Coding Data (GitHub and GitLab)
- Use Case Data

As an effort to enhance the findability, accessibility, interoperability, and reusability (FAIR4) of research data, this deliverable encompasses details related to

- The management of research data during and after the project concludes.
- The nature of data collection, processing, and/or generation.
- The methodologies and standards to be employed.
- Considerations regarding data sharing and open access.
- The approach to data curation and preservation, extending beyond the project's completion.

* OwnCloud is **an open source file sync, share and content collaboration software that lets teams work on data easily from any where**

2.1 **Objective of the deliverable**

The aim of this deliverable is to elaborate on and establish the data management procedures for all the mentioned types of data that will be created, gathered, processed, stored, and preserved throughout the development of the AFPVN project and beyond.

2.2 **Report Structure**

The document follows the established Horizon Europe template for a Data Management Plan (DMP) .

3. **AfPVN Data**

AFPVN endeavors to create a unified data infrastructure and collaborative platform ecosystem that fosters data openness and sharing, enhancing cooperation among different interested industrial partners. The report will focus on the following different types of data:

- Administration Data (OwnCloud)
  - Data Types: Administrative documents, project reports, communication records.
  - Data Storage: Data of this category will be stored on OwnCloud, a secure cloud storage platform.
  - Access Control: Access will be restricted to authorized project personnel.
  - Data Backup: Regular backups of administration data will be maintained on OwnCloud.
- Coding Data (GitHub and GitLab)
  - Data Types: Source code, documentation, scripts.
  - Data Storage: Code repositories will be maintained on both GitHub and GitLab to ensure redundancy and accessibility.
  - Access Control: Repositories will be accessible to project collaborators and contributors, with version control and issue tracking.
  - Data Backup: GitHub and GitLab provide automatic versioning, backup, and redundancy.
- Use Case Data
  - Data Types: Use case data will vary depending on the specific use cases, and the responsible partner for each use case.
  - Data Storage: The partner responsible for a particular use case will decide the most appropriate storage method, like OwnCloud, GitHub, GitLab.
  - Access Control: Access and sharing policies for use case data will be determined on a case-by-case basis by the responsible partner.
  - Data Backup: Data backup and preservation for use case data will be the responsibility of the respective partner.

3.1 Data Security and Privacy

The fPVN project acknowledges its responsibility to comply with relevant data protection regulations, such as the General Data Protection Regulation (GDPR).

All data, especially personal data, will be treated with the utmost care and in accordance with applicable laws.

3.1.1 Access control

Access to project data, irrespective of the storage platform, will be controlled and restricted to authorized personnel only.

Access permissions will be configured to limit access to only those who require it to perform their project duties.

3.1.2 Encryption

All sensitive data, including personal information, will be encrypted during transit and at rest. This ensures that even if unauthorized access occurs, the data remains protected.

3.1.3 Data minimization

Personal data will be anonymized or pseudonymized to minimize the risk of identification. Anonymized data will be used wherever possible to preserve privacy.

3.1.4 Anonymization and Pseudonymization

Collection and storage of personal data will be limited to what is strictly necessary for the project's goals. Data that is not essential will not be collected.

3.1.5 Data transfer

When data is shared among project partners or with external collaborators, secure data transfer methods will be used to maintain data integrity and security.

3.1.6 Data retention and deletion

A data retention policy will be established, specifying the duration for which data will be retained. Once data is no longer required for the project's objectives, it will be securely deleted in compliance with data protection regulations and under the principal investigator(s)' archival department's guidance. If data is deleted, a decision and documentation for this should be preserved, explaining why data was deleted, and what data was deleted.

3.1.7 Incident response plan

The project will develop an incident response plan to address data breaches or security incidents promptly. The plan will outline the steps to take in the event of a breach and the responsible individuals. For any matter regarding the response plan, you can contact the following data protection officer (DPO):

- LTU: Kajsa Borgström, dataskydd@ltu.se

3.1.8 Data protection impact assessment (DPIA)

The project will conduct Data Protection Impact Assessments (DPIAs) for all activities involving the processing of personal data. DPIAs will help identify and mitigate potential risks to data subjects' rights and freedoms.

3.1.9 Consent and data subject rights

- If the project involves the processing of personal data, clear consent mechanisms will be established for data subjects.
- Data subjects will be informed about their rights regarding their personal data, including the right to access, rectify, or erase their data.

3.1.10 Regular data audits

- Regular data audits will be conducted to ensure compliance with data security and privacy measures.
- Audits will help identify and rectify any vulnerabilities or non-compliance issues.

3.1.11 Data security in coding

Code contributors will follow best practices for secure coding to minimize vulnerabilities and data security risks in the software.

3.2 Data Sharing and Collaboration

The fPVN project will utilize various data-sharing platforms, including GitHub, GitLab, OwnCloud, and other relevant platforms as needed. Each platform will serve a specific purpose in the project.

3.2.1 Collaborative workflow

Collaborators within the project will follow a defined collaborative workflow when sharing and contributing data. The workflow will include stages such as data creation, review, approval, and publication.

3.2.2 Version control

For code and documentation, version control will be implemented using Git. Collaborators will use branches and pull requests for code contributions and reviews, ensuring that changes are tracked and reviewed.

3.2.3 Code of conduct

A clear code of conduct for data sharing and collaboration will be established and communicated to all project partners and contributors. The code of conduct will include guidelines for respectful and professional communication, open collaboration, and addressing disputes.

3.2.4 Data licensing

Code repositories will specify open-source licenses, and documentation will include licensing details for any data that is shared. Clear licensing terms will be provided to external contributors.

3.2.5 Data review and approval

Before data is incorporated into the project, a review and approval process will be implemented. This process may involve code reviews for code contributions and content reviews for documentation and use case data.

3.2.6 Data access policies

Access policies will be established for data shared on the project's platforms, specifying who can access the data, under what conditions, and for what purposes. Partners will follow access policies as defined by the responsible partner for use case data.

3.2.7 Data ownership and intellectual property

Ownership and intellectual property rights will be clearly defined in data-sharing agreements. Collaborators and contributors will acknowledge and respect these rights. Ownership may vary for different types of data, with code typically having an open-source approach and use case data following the responsible partner's terms.

3.2.8 Data dissemination

Data, especially code and documentation, will be made publicly available through appropriate platforms. A project website will link to code repositories and documentation for easy access. One of the suggested tools that can be used is Zenodo (https://zenodo.org/communities). Use case data dissemination will follow the decisions of the responsible partner.

3.2.9 Date retraction and removal

Provisions for data retraction and removal will be established. Data that is found to be inaccurate, outdated, or in violation of any terms will be retracted or removed in accordance with data protection regulations and data sharing agreements. If data is deleted, a decision and documentation for this should be preserved, explaining why data was deleted, and what data was deleted.

3.2.10 Date contribution guidelines

- Clear guidelines for external contributors will be provided, outlining how they can participate in the project, and contribute code, documentation, or other relevant data.
- These guidelines will help streamline the contribution process.

3.2.11 Date archive

- The project will maintain an archive of data versions, ensuring that historical data is accessible and retrievable.
- Archived data will be preserved for reference and transparency.
- Data related to research is deposited within the university/principal investigators' archives for at least 10 years for validation purposes.
- If data is deleted, a decision and documentation for this should be preserved, explaining why data was deleted, and what data was deleted.

3.2.12 Date access committee

A Data Access Committee may be established to oversee access requests and ensure compliance with access policies, particularly for sensitive data.

3.2.13 Date sharing agreement

Data-sharing agreements will be formalized with project partners to outline terms, responsibilities, and compliance requirements.

3.2.14 Communication and support

Regular communication and support channels will be established to facilitate collaboration and address data-sharing issues, inquiries, and disputes.

3.3 Data Preservation

The fPVN project is committed to preserving its data for the long term to ensure its availability and usability beyond the project's duration. Data preservation will follow established methodologies and best practices, adapting to the specific needs of different data types.

3.3.1 Data retention policy

The project will establish a data retention policy specifying how long data will be preserved. Data retention periods will be determined based on the type of data and legal requirements, such as those imposed by data protection regulations.

3.3.2 Access to archive data

Archived data, whether code, documentation, administrative data, or use case data, will be accessible to project personnel and, where appropriate, the wider community. Access to archived data will be granted according to the specified data access policies and any legal or ethical restrictions. Research data is considered a public/ official document in Sweden and public documents can be requested by the public. (A secrecy assessment always applies when a document is requested) More information about The Public Access to Information and Secrecy Act here: https://www.government.se/information-material/2009/09/public-access-to-information-and-secrecy-act/

3.3.3 Data format and metadata preservation

- Data will be preserved in open, widely accepted formats to ensure long-term accessibility.
- Metadata and documentation associated with the data will also be preserved to facilitate understanding and use.

3.3.4 Data format and metadata preservation

The project will periodically evaluate data formats to ensure they remain accessible as technology evolves. If necessary, data formats will be upgraded or migrated to maintain usability and compatibility.

3.3.5 Data backup and redundancy

Data backup and redundancy measures will be implemented for all data types, ensuring that multiple copies are securely stored. These measures will safeguard against data loss due to hardware failures, accidents, or other unforeseen events.

3.3.6 Regular edits and verification

The project will conduct regular audits to verify the integrity and accessibility of preserved data. These audits will help identify and address potential issues.

3.3.7 Transition and responsibilities

Clear procedures for the transition of data preservation responsibilities will be established to ensure that data remains preserved and accessible if project roles change or the project concludes.

3.3.8 Data disposal

When data is no longer needed, it will be securely disposed of in compliance with data protection regulations and other legal requirements.

3.4 Metadata and Documentation

The fPVN project will employ recognized and standardized metadata schemas, such as Dublin Core, to describe and provide context for various types of data. These standards will be adapted as needed for specific data types. Metadata schemas will be chosen to support interoperability and discoverability across different data platforms and repositories. Metadata of deposited data must be open under a Creative Common Public Domain Dedication (CC 0) or equivalent (to the extent legitimate interests or constraints are safeguarded), in line with the FAIR principles ((Findable, Accessible, Interoperable and Re-usable data)) and provide descriptive information as explained in the sections for the different types of the provided data.

3.4.1 Metadata elements

Metadata will include essential elements that describe each dataset comprehensively, including but not limited to:

- Title and description: Clear and concise titles and detailed descriptions of datasets to convey their purpose, contents, and relevance.
- Creator and contributor information: Names, affiliations, and contact details of those who created or contributed to the dataset.
- Date of creation and modification: The dates when the dataset was created, updated, or modified.
- Data format and structure: Details about the data format, structure, and file types.
- Data source: Information about the source of data, such as instruments, sensors, or data collection methods.
- Licensing and access information: Information about data access rights and usage restrictions, including licensing terms.
- Keywords and subject categories: Keywords and subject classifications that enable effective searching and categorization.
- Version and history: Documentation of dataset version history, updates, and any related datasets.
- Data quality and provenance: Information on data quality, accuracy, and the origin of the dataset.
- Related publications and references: Links to relevant publications, research papers, or documentation that provide additional context.

3.4.2 Documentation for code and software

The project will establish clear guidelines for creating, updating, and maintaining documentation for data, code, and use cases. Documentation standards will be tailored to the specific needs of each data type, with a focus on completeness, consistency, and clarity.

3.4.3 Documentation and guidelines

Code repositories will include comprehensive README files that provide essential information for collaborators and users. This documentation will encompass:

- Installation and setup instructions.
- Code structure and organization.
- Usage examples and sample code.
- Dependencies and prerequisites.
- Troubleshooting and FAQs.
- Contribution guidelines for external contributors.

3.4.4 Documentation for administrative data

Administrative data will have structured documentation to facilitate efficient data management. This may include:

- Records of communications and meetings.
- Project reports and progress updates.
- Meeting minutes and agendas.

3.4.5 Documentation for use-case data

Partners in charge of use case data will be responsible for creating detailed documentation specific to their data. Documentation may include data collection methods, data preparation procedures, and any unique considerations for the data's use.

3.4.6 Collaboration on documentation

Collaboration tools, such as collaborative document editing platforms, will be used to foster effective collaboration in creating and maintaining documentation.

3.4.7 versioned documentation

Documentation will be versioned and synchronized with code repositories to ensure alignment with code updates.

3.4.8 Documentation review and verification

Documentation, especially for code, will undergo regular review and verification by designated team members to maintain accuracy and relevance.

3.4.9 Accessibility and discoverability

- Metadata and documentation will be structured in a way that promotes discoverability and access to project data.
- Metadata and documentation will be made publicly accessible to facilitate collaboration and sharing.

3.4.10 Data dictionary

A data dictionary or glossary will be created to provide a clear definition of terms and concepts used in data and documentation.

3.4.11 Data catalogue

A data catalog may be established to provide a centralized index of datasets, metadata, and documentation, aiding in data discovery and access.

3.5 Roles and Responsibilities

- Project Coordinator: Oversees the overall DMP implementation.
- Code Maintainers: Responsible for code review, merge, and maintenance.
- Responsible Partners: Ensure proper handling and storage of use case data.
- Data Manager: Ensures the overall compliance with the DMP.

3.6 Review and Update

The fPVN project will establish a schedule for periodic reviews of the DMP to ensure that data management practices align with project goals and evolving requirements. The initial DMP review schedule will be defined, and subsequent reviews will occur annually at regular intervals.

3.6.1 Responsible parties

All project partners are encouraged to provide input and feedback during the review process to ensure that diverse perspectives and insights are considered.

3.6.2 Review criteria

The review will assess the effectiveness and compliance of data management practices. Key criteria for review will include:

- Legal and regulatory compliance: Ensuring that data management practices align with relevant laws and regulations, such as data protection regulations.
- Consistency: Verifying that data management practices are consistent with the DMP's original goals and principles.
- Alignment with evolving project needs: Assessing whether data management practices remain in line with the project's changing objectives and priorities.
- Data quality: Evaluating the quality and accuracy of data and associated metadata.
- Accessibility and discoverability: Confirming that data remains accessible to project personnel and external collaborators.
- Data preservation: Verifying that preservation practices continue to meet the project's long-term data accessibility and integrity goals.
- Documentation completeness: Assessing the comprehensiveness and accuracy of metadata and documentation for data, code, and use cases.

3.6.3 Reporting and feedback mechanism

Review findings will be documented, and a report will be generated. This report will summarize the review's results, highlighting areas of strength and opportunities for improvement. Feedback mechanisms, such as surveys or direct communication channels, will be established to collect input from project partners and stakeholders.

3.6.4 Action plan for updates

Based on the review findings, an action plan will be developed to address identified gaps or areas needing improvement. The action plan will outline specific steps, responsibilities, and timelines for implementing necessary updates.

3.6.5 Documented updates

All updates and changes to the DMP will be documented and tracked to ensure transparency and accountability. Changes to the DMP will be accessible to all project partners.

3.6.6 Communication of updates

Updates to the DMP will be communicated to project partners, collaborators, and stakeholders to keep them informed about changes in data management practices.

3.6.7 Compliance checks

The project will conduct periodic compliance checks to verify that data management practices remain consistent with legal and regulatory requirements.

3.6.8 Training and guidelines

If new practices or updates to existing practices are implemented, the project will provide training and guidance to ensure that all project personnel understand and can adhere to the updated processes.

3.6.9 Continuous Improvement

- The review and update process is part of the project's commitment to continuous improvement in data management practices.
- Feedback received from project partners and stakeholders will be considered in shaping future updates to the DMP.

3.7 Training and Support

The fPVN project will establish comprehensive training sessions to educate project personnel, collaborators, and external contributors about data management practices and compliance with the DMP.

3.7.1 Training objectives

The training programs will aim to achieve several key objectives:

- Ensure all project personnel understand the importance of data management and its role in achieving project goals.
- Provide clear guidance on how to perform various data management tasks, including data submission, version control, metadata creation, and documentation.
- Educate participants on data security and privacy best practices, especially when handling sensitive or personal data.
- Familiarize participants with relevant laws and regulations, such as GDPR, and their implications for data management.
- Train external contributors on code contribution processes and documentation standards.

3.7.2 Targeting audiences

Training programs will be tailored to specific audiences, such as project partners, data managers, code contributors, and use case owners. Separate training modules may be developed to address the unique needs of different roles within the project.

3.7.3 Delivery method

Training will be delivered through a variety of methods to accommodate different learning styles and accessibility needs. These methods may include:

- Workshops and seminars.
- Online courses and webinars.
- On-demand video tutorials.
- Documentation and guides.
- One-on-one or group mentoring.

3.7.4 Continuous training

Training is not a one-time event but an ongoing process. Regular refresher courses and updates will be provided to ensure that project personnel remain up-to-date with the latest data management practices.

3.7.5 Documentation and resources

A central repository of documentation and resources related to data management will be created to serve as reference materials for training and support.

- These resources may include best practice guides, checklists, templates, and FAQ documents.
- Training materials, including slides and course notes, will be made available for review and self-study.

3.7.6 Role-specific guidance

Training and support will be tailored to the specific roles within the project. For example, code contributors will receive guidance on code submission, while data managers will focus on data preservation and metadata creation.

3.7.7 Feedback and evaluation

An evaluation mechanism will be implemented to gather feedback from participants about the effectiveness and relevance of training and support. Feedback will inform updates and improvements to training programs.

3.7.8 Accessibility

Training materials will be designed to be accessible to individuals with disabilities. This includes providing materials in accessible formats and ensuring that virtual training sessions are inclusive.

3.7.9 Language support

Training materials and support mechanisms will be mainly provided in English.

3.7.10 Training records

Training records will be maintained to track who has received training and when. This helps ensure that all relevant personnel have completed necessary training.

.

4. Conclusions

This document outlines the data involved in the project, detailing its utilization, production, and management during and after the project's completion. It encompasses various data types crucial to the project, including pilot data enriched with open data, project deliverables (both confidential and public), publications, software artifacts, research-related artifacts from the project infrastructure, and working documents.

Furthermore, the deliverable provides insights into methodologies ensuring that research data adheres to the principles of being findable, accessible, interoperable, and re-usable (FAIR). Special attention is given to handling research data throughout and after the project, specifying the data collection, processing, and generation methods, the application of methodologies and standards, considerations for data sharing and open access, and the curation and preservation of data post-project completion.

5. Revision history

5.1 Contributing and reviewing partners

Contributions	Reviews	Participants	Representing partner
1	X	Pär Erik Martinsson	LTU
2	X	Lama Alkhaled	LTU
	x	Jerker Delsing	LTU

5.2 Amendments

No.	Date	Version	Subject of Amendments	Author
XX	xxxx-xx-xx			

5.3 Quality assurance

No	Date	Version	Approved by
1	2023-11-21	1.0	Jerker Delsing