Project Data Management Plan PDMP

Deliverable D1.3

Version N°1

Authors: Hille Rautkoski, Miika Nikinmaa, Camilla Halinoja, VTTNina Jeliazkova, IDEA, Marta Casadesus, AEI

Co-funded by the European Union

Disclaimer

The content of this report reflects only the author's view. The European Commission is not responsible for any use that may be made of the information it contains.

Co-funded by
the European Union

Document information

Grant Agreement	n°101092164
Project Title	Development of verified safe and sustainable PFAS-free coatings for food packaging and upholstery textile applications
Project Acronym	ZeroF
Project Coordinator	Miika Nikinmaa, VTT
Project Duration	1 January 2023 – 31 December 2025 (36 months)
Related Work Package	WP1
Related Task(s)	1.4
Lead Organisation	VTT
Contributing Partner(s)	IDEA, AEI, ALL
Due Date	30.6.2023
Submission Date	26.6.2023
Dissemination level	Public

History

Date	Version	Submitted by	Reviewed by	Comments

Table of contents

Summary

This deliverable is the initial Project Data Management Plan for the ZeroF project. This deliverable is a living document which will be updated throughout the project. An updated version will be submitted in M36 (D1.4 Updated version of Data Management Plan). This report is intended primarily for the ZeroF consortium, and external stakeholders interested in evaluating the data management procedures of the project. Monitoring the Project Data Management Plan implementation and coordinating the data management will be the tasks of the appointed Data Manager (Nina Jeliazkova, IDEA).

ZeroF generates data through administrative and research related tasks, which relate to the main subject areas of the research. Different methods will be used to collect the data, e.g., relevant infrastructure software will be used where needed and typical methods for stakeholder engagement like interviews, questionnaires, and workshops. Datasets will be processed and analysed using relevant software. The operations are GDPR compliant regarding personal data and fulfil the FAIR principles. Personal Data Management is detailed in the ethics deliverable D1.2 which was submitted to the Commission in March 2023.

The main objective of ZeroF is to develop and validate two novel Safe and Sustainable by Design (SSbD) PFAS-free hybrid coating formulations. The developed coating materials will be tested in packaging and textile applications to demonstrate coating performance while building an exploitation pathway through stakeholder engagement, consumer, and regulatory acceptance activities, including certification road-mapping. ZeroF will also develop and demonstrate guidelines and computational tools for SSbD of novel materials in practice.

Abbreviations and acronyms

Acronym	Description
CC-BY	Creative common licence
CC-BY-SA	Creative common license
CERN	European Organization for Nuclear Research
CSC	Standardized access protocols
DOI	Digital Object Identifier
EEA	European Economic Area
EUDAT	Collaborative Data Infrastructure

GDPR	General Data Protection Regulation
FAIR	Findable, Accessible, Inter-operable, Re-use
IDA	Fair data storage
IPR	Intellectual Property Rights
LCA	Life Cycle Assessment
PDMP	Project Data Management Plan
PFAS	Per- and Polyfluoroalkyl Substances
SSbD	Safe and Sustainable by Design
VTT	Technical Research Centre of Finland
WP	Work Package

Co-funded by
the European Union

1 Introduction

This deliverable outlines the initial Project Data Management Plan (PDMP) of the ZeroF project. An updated version of the Data Management Plan will be submitted in the end of the project, M36 as a D1.4 "Updated version of Data Management Plan". This report is intended primarily for the use of the ZeroF consortium, and for external stakeholders interested to evaluate the data management procedures of the project. Monitoring the PDMP implementation and coordinating the data management will be the tasks of the appointed Data Manager (Nina Jeliazkova, IDEA).

ZeroF generates data through administrative and research related tasks, which relate to the main subject areas of the research. Different methods will be used to collect the data, e.g., relevant infrastructure software will be used where needed and typical methods for stakeholder engagement like interviews, questionnaires, and workshops. Datasets will be processed and analysed using relevant software. The operations are GDPR compliant (regarding personal data) and fulfil the FAIR principles. Personal Data Management is detailed in the ethics deliverable D1.2 which was submitted to the Commission in March 2023.

Quality control measures will be taken to maintain the accuracy of data during the project. Potential re-utilization of any open data will be ensured by careful documentation of datasets as well as description and publication of data collection methods, protocols, workflows, and models. The project consortium has appropriate technical and organizational measures in place to carry out data management and protection during the project. Project documents will be stored on Microsoft Teams-based ZeroF workspace to which all partners have access.

The consortium is strongly committed to promoting open science. All the scientific articles, conference papers, public deliverables and project reports produced are published according to the open access principles: publicly available and stored repository. All publishing partners have budgeted funding to cover open access costs, all generated public research outputs will be published openly without delay, and they will be stored in a public repository. Public portals like European Open Science Cloud and OpenAIRE and their public repository Zenodo will be used.

Findability will be supported by using persistent identifiers (DOI) and used in linking to datasets. Creative Commons license CC-BY-SA or CC-BY will be used for any open datasets unless there are compelling reasons to select a more restricted type of CC-license. The data is presented and stored in a way that reusing and interoperability are ensured.

2 ZeroF Data management

2.1 Overview of data

ZeroF generates data through research related and administrative tasks, which relate to the main subject areas of the research. The generated data can be divided into the following two categories:

1) Research related activities and data, such as desktop reviews, stakeholder engagement activities, Life Cycle Assessments (LCA) and via the research infrastructure and equipment used for experimental and development work.
2) Administrative tasks and data (such as minutes, agendas)

Experimental work at laboratory and pilot test units will generate experimental and numerical research data, images, and text. The data will include physical property data, process parameters and qualitative and descriptive data for the substances and processes. Data from material research include e.g., melting point, flash point, degradation temperature and flame retardancy, solubility, rheometric data, spectroscopic data, surface energies and mechanical property data such as strength properties, Martens hardness and abrasion resistance, and barrier property data. Environmental and toxicological data that can be collected include cytotoxicity, biodegradability, mutagenicity, aquatic toxicity using Daphnia magna, and in vitro methods for inhalation and particle toxicity, irritation and respiratory sensitisation. Researchers will examine the possibilities of re-using open research data or other pre-existing data, software or algorithms or standards, as available.

Different methods will be used to collect the data. For example, relevant infrastructure software will be used where needed and typical methods for stakeholder engagement like interviews, surveys, questionnaires, and workshops are utilised.

The data will cover the main subject areas of the research. The main types used for the datasets will be raw laboratory data, qualitative and quantitative data, personal data, digital image and video data, and different reports. The main formats of data will be handwritten and digital (lab) notebooks, logs, Excel, csv, tab, documents, and reports in the format of Word, ppt, pdf, different image, and video formats like png, jpeg, mp4. Personal data will be collected on a minimum level possible, including details like name, organisation, title, and contact details of consortium members. The estimated volume of data will be moderate considering the main outputs are different report documents, laboratory, and experimental data as well as different campaign materials. The videos produced in the project will most likely be the biggest datasets.

The datasets will be processed and analysed using relevant software. The data analysed in connection with the project work will be produced by the project, but the possibility of reuse of any existing relevant open data will also be examined carefully.

Quality control measures will be taken to maintain the accuracy of data during the project. Potential re-utilization of any opened data will be ensured by careful documentation of datasets as well as description and publication of data collection methods, protocols, workflows, and models.

An overview of the data generated and collected as well as how it is handled in ZeroF is depicted below in Table 1 per type of data and linked to Work Package (WP) level activities.

Table 1 ZeroF project type of generated data, and project activities and used software foreseen in the beginning of the project

Type of data	Nature of data	Format of data	Open access	ZeroF activities and used software foreseen in the beginning of the project
Raw laboratory data	Laboratory experiments, results	Notebooks, logs, csv files	NO, metadata possibly	WP2, WP3, WP4: The processing and the analysis equipment will store the raw laboratory data on their system hard drives. There is also paper processing document where notes are handwritten during the processing trials and stored in the appropriate processing trial folder. Laboratory notebooks are kept during measurements
Quantitative data	Lab experimental, material composition data and analytical performance data	Excels, csv, tab (numerical data, units)	Will be decided later by the consortium	WP2, WP3, WP4: Analyzing equipment data, WP5: computational data, quantitative data from lab experiments, chemical and material databases, AMBIT software (ambit.sourceforge.net), REST API, eNanoMapper WP6: LCA data, databases WP7: Quantitative data of the dissemination and communication activities stored in an Excel file. The data includes description of the D&C activity, target audience, communication channel, outcome and link. ZeroF webpage uses Matomo Analytics for analytics, a web analytics service provided by Matomo. Matomo Analytics uses "cookies", which are text files placed on the computer to help the website analyse how visitors use the site. As the plans for experimental data become more concrete, the consortium will decide on the openness, licensing and embargo period of the experimental data. The decision will be made based on

				the FAIR principles and, where relevant, the principle 'as open as possible, as closed as necessary'.
Qualitative data	Interview, survey, questionnaire, workshop results	Notebooks, excel, word	YES (only metadata where required)	WP6: Consumer survey, WP7: Data obtained through interviews, surveys, questionnaires and workshop for dissemination and communication activities in word, Notes files and recordings. The data will be used for press releases, blog notes, social media posts and articles.
Personal data (data minimization)	From participants of workshops, surveys, questionnaires, interviews	Name, title, organization, contact details	NO	In the case personal information is required, this will be treated as confidential and will be handled in accordance with data protection and regulation (GDPR). WP1: Data related to the experts in the Advisory board, project mailing lists and meeting participation lists WP6: Consumer survey, Final workshop and awareness campaign, Questionnaire, ISO experts, WP7: Personal data for organizing two events with stakeholders (mailings), final exploitation workshop and for impact study will be handled. Data relevant for the dissemination, and communication activities, such as mailing and participant lists, with the participants' consent. Awareness campaign
Digital image and video data	Images, videos, workshops, news and media items, communication materials	Image and video formats (jpeg, mp4, etc.)	YES (no confidential content)	WP7: Awareness campaign, videos
Documentation	Project reports, deliverables	Word, pdf, ppt	YES, when public	For documentation, standard Microsoft Office programmes like Word, Power Point and Excel will be used with ZeroF visual identity. WP1: Public deliverables D1.2, D1.3, and D1.4.

				WP6: Public deliverables D6.1, D6.2, D6.6 and D6.7.
				WP7: Public deliverables D7.1, D7.2, D7.4, D7.5 and D7.6.

2.2 Data processing practicalities

These are the main principles of ZeroF data processing

- Project data including research and administrative information will be stored in project's Microsoft Teams channel, whenever possible
- All relevant partners commit to fill in Research Metadata template in Teams or harmonized data entry templates in eNanoMapper database. Selection of the tool will be decided in the next GenA meeting.
- Joint sample labelling will be created and used/utilized
- In case of any question ZeroF Data Manager Nina Jeliazkova, who is in charge of monitoring and advising in data related topics of the project, provides answers
- Research Metadata template or harmonized data entry templates in eNanoMapper database will be checked in mid-term (before M18) and in the end of the project (M36).

3 FAIR Data

3.1 Making data findable, including provisions for metadata

Discipline compliant metadata elements will be used describing the data to aid data discovery and potential re-use. Metadata including descriptions and keywords of opened data will be made available via FAIR compliant repository for searching and discovery after project closure. Persistent identifiers provided by the repository will be used in identifying and linking to datasets.

The WP5 database provides free text and faceted search, which is capable of aggregating search results from local and remote databases (based on the widely used Apache Solr search engine). If the metadata is open, it can also be configured to be indexed by the Google Dataset search.

Each data entry in database is assigned with a release version tag.

User friendly data entry templates with ontology annotation support and lab notebooks workflow can make the metadata creation faster and more reliable.

Annex 1 presents the planned content of research metadata that will be used in the project. Information will be collected frequently. The collected content will be reported on the D1.4 Updated Project Data Management Plan (M36).

3.2 Making data openly accessible

Decisions concerning the sharing of datasets will be taken by the General Assembly. The Coordinator, in collaboration with project participants, will take all the appropriate measures to make relevant data openly available and usable for third parties for study, teaching and research purposes.

During the project, data will be used mainly by the consortium members.

If, after project closure, permission to re-use the data is required, all requests for further use of data will be considered carefully and whenever possible approved by the data owner or by the General Assembly if needed. Permission for data use will be granted providing that there are no IPR or confidentiality issues involved or any direct overlap of research questions with the primary research. Permission will be provided by request using the appropriate procedure described in connection with other metadata. The ZeroF consortium will decide upon the data licenses and management by the end of the project.

The primary focus in data sharing will be on the data underlying prospective scientific publications ensuring the validation of results presented in publications. In addition to the summary data, also operational or raw data will be opened when benefits and possibilities for successful raw data re-use are recognized and there are no confidentiality or commercialization issues involved or identified.

Published and FAIR-compatible data will be archived in a public and trusted repository. Unless no discipline-specific archive platform is available, generic and certified repository services using standardized access protocols, e.g., CSC's IDA, CERN's Zenodo or EUDAT's B2SHARE, will be used to enhance long-term accessibility and re-usability of the data. The metadata of the datasets will be opened under the public and open copyright license, CC0.

Justification for possible case-specific embargo for published data will be decided by the project consortium. The embargo will be sought, if necessary, in connection with possible IPR protection or any potential patent, utility mode etc. application based on project results.

No definite period or time limit is planned for access to data. However, the opened data will be deposited in a repository, which guarantees a foreseeable future for data integrity on bit-level. No perpetual data curation policy to guarantee full long-term digital preservation of datasets is planned at this point.

3.3 Making data interoperable

Variables and value names will be constructed following the general data processing conventions and standards common to the research subject. A list of value names and used vocabulary will be provided separately. WP5 uses open-source FAIR database, which includes mapping the vocabulary to ontology terms and provides programmatic access via REST API for data analysis.

3.4 Increase data re-use

After the project completion, the ownership of datasets will belong to the grant beneficiaries that generated them. Creative Commons license CC-BY-SA or CC-BY or similar public copyright license will be used for any opened datasets, unless there are compelling reasons to select more restricted type of public license. Creative commons licenses will, by default, also include a disclaimer of liability for the re-use of opened data.

The data quality will be assured by following appropriate quality control and curation methods, e.g., rigorous control of any incoming data by well-managed data profiling (formats, value distributions and data consistency and completeness will be assessed for any incoming data); logically defined data pipeline with centralized data management preventing duplicate data entering the system; capturing and documenting data conditions and scenarios with their dependencies and conditions; maintaining data integrity with checksums and triggers, if necessary; enhancing data and metadata lineage traceability for

the pipeline, thus enabling more effective data governance. Research teams will regularly check the quality of not just the data, but also related software, algorithms, and workflows when and if changes are made to them.

4 Other research outputs

Any other project outputs, which will be needed for verifying or analysing the data - software, algorithms, workflows, protocols or models - will be assigned open license alongside the corresponding data.

5 Allocation of resources

Making research data quality-controlled, FAIR-compatible, and as open as possible has been considered by the consortium members while allocating resources to the project. During the project duration, costs related to research data management and opening are eligible as part of the project funding.

Each consortium member is responsible for covering their costs during pre- and post-grant phases with their own funding. During the project, consortium partners will be responsible for managing and curating datasets in their possession while the central data in the Microsoft Teams platform is administered by VTT as the coordinator of the project. Long-term preservation and sharing of datasets are the responsibility of the project coordinator.

At the end of the project, each consortium member will take appropriate measures to ensure the long-term preservation and sharing of opened datasets.

6 Data Security

Each project partner is responsible for their research related data collection and dataset management as well as use. Moreover, each partner must ensure that their datasets are correctly recorded. In addition, a FAIR database e.g., eNanoMapper will be considered as an option.

During the project, datasets will be available on a need-to-know basis only to those project participants or consortium members, who have been accredited by and whose data usage has been approved by the Principal Investigator or other authorized project consortium member. The project participants will be responsible for curating, preserving, disseminating, and deleting in appropriate manner the datasets in their possession. Retention time for curated datasets will be the same as for other project results.

The data collected or acquired within the project will be stored in secure systems, such as Microsoft Teams for business IT environment behind a firewall at VTT's premises or in a secure cloud environment provided by VTT's authorized and security-cleared IT service providers. Access to these platforms need registration and multifactor authentication from their users. A responsible project participant at VTT will check the requests to access to project's Teams channel. When access is granted to research data, it will be provided

through secured telecommunications channels. The General Data Protection Regulation (GDPR, Regulation (EU) 2016/679 of the European Parliament and of the Council) will be followed when storing and transferring sensitive or personal data.

As the project progresses and we start to gather research data, a table will be added here to specify the places in which data is stored. FAIR database eNanoMapper or similar system will be utilized to store research metadata. The personal data collected from the interviews will be made accessible to consortium members strictly on a need basis, and a separate restricted folder will be created in Teams to ensure this.

Data security is ensured according to national legislation, and the members of the consortium operating the data are responsible for fulfilling these requirements. Long-term and secure preservation of published research data will be ensured by using only certified and OpenAIRE guidelines compatible repositories.

Personal data received from the users will be treated confidentially, complying properly at all times the provisions contained in the RGPD and any other current norm that may be promulgated on the subject in the future.

Experimental data from partners to support computational modelling will be collected. This is done through harmonized data entry templates (shared on project Teams server), which are subsequently submitted to a FAIRification process (Kochey, 2020), and imported into a FAIR database (Jeliazkova, 2021). The database has role-based security and implements standard authentication protocols as OAuth2. The database access (web based) is granted to partners requesting access. API access plans, using Oauth2 or API keys, are provided to software developers.

7 Ethics

The privacy of data subjects will be secured by following closely the GDPR. The project consortium has appropriate technical and organizational measures in place to carry out data protection during the project. VTT has an appointed Data Protection Officer who can advise on data protection questions when necessary. Processes that handle personal data have been designed and built with the GDPR principles at the core. Specifically, informed consent for data sharing and long-term preservation is always included in questionnaires dealing with any personal data.

Dedicated processes are implemented to safeguard and protect research data (e.g., using pseudonymization or full anonymization where appropriate), and use the highest possible privacy settings by default. No person or organization involved will unintentionally be identified directly or indirectly in the datasets. Any indirect reference to sensitive personal information or e.g., lines of businesses, branches or industries will be removed and destroyed after the anonymized dataset has been checked and validated. After curation, no person-related data is available publicly without explicit, informed consent, of the data subject and – if no full anonymization is required – publicly available data cannot in any

circumstances be used to identify a subject without additional information stored securely in a separate place.

Project members will always retain unambiguous and individualized affirmations of consent from the data subjects and the subjects will always have the right to revoke their consent at any time. During and after the end of the project, the project members will clearly disclose any datasets, which have been collected during the project and declare the lawful basis and purpose for their processing. In addition, project members will state how long the data in their possession will be retained and unambiguously declare, if it is being shared with any third parties or outside of the EEA. Data subjects of the project will have the right to request a portable copy of the data collected in a common format, and the right to have their data erased under specified circumstances. VTT employs a data protection (privacy) officer (DPO), who is responsible for managing compliance with the GDPR. Research integrity and ethical principles related to data collection and use are covered in detail in the ethics section of the Grant Agreement.

A separate public Deliverable D1.2 Ethics Plan has been submitted on M3 (March 2023). That deliverable covers ethics aspects of the ZeroF project in more detail.

8 References

EC, 2021: Horizon Europe Data Management Plan Template Version 1.0. 05 May 2021

EC, 2016: General Data Protection Regulation (Regulation (EU) 2016/679 of the European Parliament and of the Council)

Kochev, N. *et al.* Your Spreadsheets Can Be FAIR: A Tool and FAIRification Workflow for the eNanoMapper Database. *Nanomaterials* **10**, 1908 (2020).

Jeliazkova, N. et al. Towards FAIR nanosafety data. Nat. Nanotechnol. 16, 644–654 (2021)

Annex I Research Metadata Content

Project's metadata information will contain following headlines and it will be updated frequently. The collected content will be reported on the D1.4 Updated Project Data Management Plan (M36).

General overview

- ID
- Work package
- Resource type
- Title
- Version
- Date of creation
- Creator
- Contributors
- File location
- Software used to create data
- Origin and method

Content description

- Description and relation to the project objectives
- Subjects (keywords)
- Geographical location
- Code list used
- List of variables
- Metadata schema used

Technical description

- Software used to create the file
- File format
- Necessary software
- File size

Sharing and preservation

- Use/Users
- Rights/Licence
- Dissemination
- Access
- Restrictions
- Repository for open data
- Permanent identifier (e.g., DOI, URN)