



D8.3 Data Management Plan (DMP)

Submission date: M6

Due date: M6

DOCUMENT SUMMARY INFORMATION

Grant Agreement No	101091575	Acronym	tExtended
Full Title	Knowledge Based Framework for Extended Textile Circulation		
Start Date	01/12/2022	Duration	48 months
Deliverable	D8.3: Data Management Plan		
Work Package	WP8 – Management		
Type	DMP	Dissemination Level	Public
Lead Beneficiary	VTT		
Authors	Mikko Mäkelä, VTT		
Co-authors			
Reviewers	Pirjo Heikkilä, VTT		



Funded by the
European Union

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 HADEA. Neither the European Union nor the granting authority can be held responsible for them

DOCUMENT HISTORY

Version	Date	Changes	Contributor(s)
Draft	24/4/2023	Draft version for comments	Mikko Mäkelä, VTT
Review	17/5/2023	DMP to review	Mikko Mäkelä, VTT
Review	23/5/2023	Review comments	Pirjo Heikkilä, VTT
Final	26/5/2023	Finalization	Pirjo Heikkilä, VTT

PROJECT PARTNERS

Partner	Country	Short name
TEKNOLOGIAN TUTKIMUSKESKUS VTT OY	FI	VTT
RISE RESEARCH INSTITUTES OF SWEDEN AB	SE	RISE
CENTRE SCIENTIFIQUE & TECHNIQUE DEL'INDUSTRIE TEXTILE BELGE ASBL	BE	CTB
CENTRO TECNOLOGICO DAS INDUSTRIAS TEXTIL E DO VESTUARIO DE PORTUGAL	PO	CIT
AIMPLAS – ASOCIACION DE INVESTIGACION DE MATERIALES PLASTICOS Y CONEXAS	ES	AIM
CARR COMMUNICATIONS LIMITED	IE	CARR
LOUNAIS-SUOMEN JATEHUOLTO OY	FI	LSJH
GREENWAY LO	SE	TEKO
PURFI MANUFACTURING BELGIUM	BE	PURF
UTEXBEL NV	BE	UTEX
ERT TEXTIL PORTUGAL SA	PO	ERT
RESTER OY	FI	RES
MTEX NEW SOLUTION SA	PO	MTEX
ZALA BRIVIBA BIEDRIBA	LT	ZALA
INESC TEC – INSTITUTO DE ENGENHARIADE SISTEMAS E COMPUTADORES, TECNOLOGIA E CIENCIA	PO	INES
NARODNA RECYKLACNA AGENTURA SLOVENSKO	SK	NARA
TEXTEIS J. F. ALMEIDA S.A.	PO	JFA
SOPREMA	FR	SOP
DE KRINGLOOPWINKEL DELTAGROEP	BE	KLW
Tally Weijl Trading AG	CH	TAW

LIST OF ABBREVIATIONS

Abbreviation	Definition
DMP	Data Management Plan
GDPR	General Data Protection Regulation
GenA	General Assembly

Executive Summary

The overall objective of the tExtended project is to develop a knowledge-based Blueprint, which is a master plan for a sustainable textile ecosystem where utilization of textile flows is optimized, aiming for retention of value of materials in a safe and sustainable way. Within tExtended project we will develop technological and digital solutions enabling efficient textile cycles and show potential to reduce textile waste by 80 % within Real Scale Demonstrator. More about the project in <https://textended.eu/>.

This Data Management Plan (DMP) is a public deliverable of the tExtended project funded by the European Union Horizon Europe programme under Grant Agreement no. 101091575. This document is the first version of the DMP delivered during month 6 of the project and will be updated regularly based on the partners needs and updates. The DMP is accompanied by an online Data Description Template, which will clarify in more detail what kind of data will be generated during the project.

The purpose of the DMP is to establish the data management policy, which lays out the guidelines for generating, processing, and storing data during and after the project and making the data findable, accessible, interoperable, and reusable. Data generated in tExtended will consist of **data from third parties, data from experiments and measurements, and data from human sources** and will include an open access dataset on material sorting and/or quality control measurements.

tExtended data will be made findable, open accessible, interoperable, and its reuse will be promoted based on FAIR principles outlined in this DMP. These principles also apply to other research outputs, such as software, algorithms, workflows, protocols, or models, taking into consideration the potential limitations. Each project member is responsible for allocating the necessary resources, for covering the generated costs, and for taking care of data security, which are important aspects of the data management within the project. Ethical aspects, and especially the privacy of data subjects with data from human sources, will be closely followed to take into consideration the General Data Protection Regulation of the European Union.

Table of Contents

Executive Summary	5
Table of Contents	6
1 Introduction	7
2 Data Management	8
2.1 Data Summary.....	8
2.2 Practicalities.....	9
3 FAIR Data	9
3.1 Making Data Findable, Including Provisions for Metadata.....	9
3.2 Making Data Openly Accessible.....	9
3.3 Making Data Interoperable.....	10
3.4 Increasing Data Reuse.....	10
4 Other Research Outputs	11
5 Allocation of Resources	11
6 Data Security	11
7 Ethical Aspects	12
Appendix 1 Data Description Template	14

1 Introduction

The purpose of this DMP is to establish the data management policy of tExtended. The data management policy lays out the guidelines for generating, processing, and storing data during and after the project and making the data findable, accessible, interoperable, and reusable. The DMP is a living document, which will evolve based on the partners' needs and updates. The DMP is accompanied by an online Data Description Template, which will clarify what kind of data will be generated during the project.

The first version of this DMP was prepared based on a VTT template for a DMP deliverable. Relevant additional project documents are:

- Grant Agreement (no. 101091575)
- Project proposal (Description of Action, Annex 1 of the Grant Agreement)
- Project Quality Management Plan¹
- Ethics Plan²

¹ D8.1 Project Quality Management Plan, lead partner: VTT, type: PQMP, dissemination level: SENSITIVE, submission date: M2.

² D8.2 Ethics Plan, lead partner: VTT, type: ETHICS, dissemination level: PUBLIC, submission date: M3.

2 Data Management

The purpose of the DMP is to establish the data management policy of tExtended to ensure that data management is an integral part of the project. The DMP is a living document and will be updated regularly throughout the project. Data management and the DMP are tasks of the appointed Data Manager (Mikko Mäkelä, VTT) as a part of the management work package of the project³.

2.1 Data Summary

tExtended research data will be generated and collected according to the methods outlined in the project proposal and will cover the main subject areas of the research. Datasets will be processed and analysed using appropriate 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 carefully examined. Quality control measures will be taken to maintain the accuracy of data during the project. Potential reutilization of any opened data will be ensured by careful documentation of datasets and by description and publication of data collection methods, protocols, workflows, and models.

tExtended data will include:

- **Data from third parties** such as material and waste flow data from existing databases from public agencies and previous scientific research
- **Data from experiments and measurements** such as numerical data from chemical laboratory analyses and spectral measurements
- **Data from human sources** such as online surveys and stakeholder interviews

Exact types and formats of these data have not yet been defined by the project members and the estimated data volumes are still unknown. A preliminary plan of collected data and types has been provided in the project proposal⁴. A more detailed description of the data will be collected using a Data Description Template, which has been shared among the project members. The Data Description Template is an online document, which aims to clarify which data will be collected in the project tasks and will be periodically updated by the project members involved in data collection and generation. The document will be accessible during the project lifetime and is managed by VTT. The information collected in the Data Description Template is provided in Appendix 1 and is further summarized in Table 1.

tExtended deliverables include an open dataset.⁵ The dataset will consist of material sorting and/or quality control measurements performed by VTT at the LSJH sorting line and will include the respective ground truth information derived from, for example, current sorting processes. The dataset will be published as open data to support other research initiatives targeting similar material identification and quality analysis topics.

³ WP 8 Management, T8.3 Data management (lead: VTT, contributors: all partners), duration: M1-M48.

⁴ Project proposal section 1.2.6 Data management and management of other research outputs (page 118 of the Grant Agreement)

⁵ D2.3 Open data set of textile material sample measurements, lead partner: VTT, type: DATA, dissemination level: PUBLIC, delivery date: M48.

Table 1 An overview of the information collected in the Data Description Template

Headlines	Metadata descriptors
General overview	Index number, work package, resource type (dataset), title, version, date created, creator, contributors, file location (path), software, origin and method
Content description	Description and relation to the project objectives, 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/license, dissemination, access, restrictions, repository for open data, permanent identifier

2.2 Practicalities

Data management is a fixed agenda point in the General Assembly (GenA) meetings of tExtended. The Data Description Template and any other data management related issues will be reviewed at the meetings on demand.

3 FAIR Data

3.1 Making Data Findable, Including Provisions for Metadata

Discipline compliant metadata elements will be used to describe the data to aid data discovery and potential re-use. The metadata elements and standards are provided in a separate Data Description Template. 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.

3.2 Making Data Openly Accessible

Decisions concerning the sharing of datasets will be taken by the tExtended GenA. The coordinator in collaboration with the 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.

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 person mandated with the task. Permission for data use will be granted providing 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.

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 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, for example, CSC's [IDA](#)⁶, CERN's [Zenodo](#)⁷ or EUDAT's [B2SHARE](#)⁸, will be used to enhance long-term accessibility and re-usability of the data. Metadata of the datasets will be opened under public and open copyright license, CC0.

Justification for possible case-specific embargo for published data will be decided by the project consortium. Embargo will be sought, if necessary, in connection with possible IPR protection or any potential patent, utility mode etc. application based on the 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 for foreseeable future the 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 general data processing conventions and standards common to the research subject. List of variable names and used vocabulary will be provided in a separate list. Examples of vocabulary information to be managed within the project will be, e.g., units of observation, list of variables with the name and label of each variable as well as its values and value labels, information on the classifications used and meanings of abbreviations used.

3.4 Increasing Data Reuse

After project completion 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.

Data quality will be assured by following appropriate quality control and curation methods, for example, 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 changes are made in them.

⁶ <https://www.fairdata.fi/en/ida/> (accessed 24 April 2023).

⁷ <https://zenodo.org/> (accessed 24 April 2023).

⁸ <https://b2share.eudat.eu/> (accessed 24 April 2023).

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 opened alongside the corresponding data taking into consideration the potential limitations outlined in Section 3.2 Making .

5 Allocation of Resources

Making research data quality-controlled, FAIR-compatible, and as open as possible should be considered by the consortium members while allocating resources to the project. Costs related to making data and other research outputs FAIR may include direct and indirect costs. Direct costs will be included in the budget estimates, considering the eligibility rules and the usual accounting practises of each project participant.

The direct costs may include, for example, personnel costs and costs related to storage, archiving, re-use, and security. The costs should be estimated and included in the Grant Agreement budget, and actual costs reported in the financial statements in each periodic report.

Each consortium member is responsible for covering their costs during pre- and post-grant phases with own funding. During the project, consortium partners will be responsible for managing and curating datasets at their possession. At the project ending each consortium member will take appropriate measures to ensure long-term preservation and sharing of opened datasets.

6 Data Security

At the beginning of the research project, the project consortium will decide and agree on the tasks, roles, responsibilities, and rights relating to data collection, dataset management and data use.

During the project, datasets will be available only to those project participants or consortium members, who have been accredited by and their data usage has been approved by the consortium members, which have generated the data. The organisations of the consortium members 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.

Data collected or acquired within the project will be stored in a secure IT environment behind a firewall at the tExtended consortium members' premises or in secure cloud environment provided by the tExtended consortium members' authorized and security cleared IT service providers. Access to it will need registration and authentication. Responsible project participants of the consortium members will check applications for the use of data. Where access is granted to research data, this will be provided through secured telecommunications channels. EU GDPR regulation will be followed in storage and transfer of sensitive or personal data.

Long-term and secure preservation of published research data will be ensured by using only certified and OpenAIRE guidelines⁹ compatible repositories.

⁹ <https://guidelines.openaire.eu/en/latest/> (accessed 24 May 2023).

7 Ethical Aspects

Privacy of data subjects will be secured by following closely the General Data Protection Regulation (GDPR, Regulation (EU) 2016/679 of the European Parliament and of the Council¹⁰). The project consortium has appropriate technical and organizational measures in place to carry out data protection during the project.

Processes that handle personal data have been designed and built with the GDPR principles considered. Specifically, informed consent for data sharing and long-term preservation is always included in questionnaires dealing with any personal data. Processes provide safeguards to 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 identifiable directly or indirectly in the datasets. Any indirect reference to sensitive personal information or, for example, 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 closure 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 European economic area. 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, who is responsible for managing compliance with the GDPR.

Research integrity and ethical principles related to data collection and use are also covered in the Ethics Plan of tExtended¹¹. Rules on the processing of personal data by the project members are also outlined in the Grant Agreement under Article 15 – Data protection:

The beneficiaries must process personal data under the [Grant] Agreement in compliance with the applicable EU, international and national law on data protection (in particular, Regulation 2016/679⁹).

They must ensure that personal data is:

- *processed lawfully, fairly and in transparent manner in relation to the data subjects*
- *collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes*
- *adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed*
- *accurate and, where necessary, kept up to date*

¹⁰ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679> (accessed 24 April 2023).

¹¹ D8.2 Ethics Plan, lead partner: VTT, type: ETHICS, dissemination level: PUBLIC, submission date: M3.

- *kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data is processed and*
- *processed in a manner that ensures appropriate security of the data.*

The beneficiaries may grant their personnel access to personal data only if it is strictly necessary for implementing, managing and monitoring the [Grant] Agreement. The beneficiaries must ensure that the personnel is under a confidentiality obligation.

The beneficiaries must inform the persons whose data are transferred to the granting authority and provide them with the Portal Privacy Statement.

Appendix 1 Data Description Template

The Data Description Template is an online Excel file with input fields for metadata. The exact fields and *example inputs* are given below.

General overview:

1. Index: *0*
2. Work package: *0*
3. Resource type: *Dataset*
4. Title: *Fiber_comp*
5. Version: *1.0*
6. Date of creation: *29.3.2023*
7. Creator: *Mikko Mäkelä*
8. Contributors: *-*
9. File location (file path): *TBD*
10. Software used to create the data: *LUMO*
11. Origin and method: *Generated using a Specim SWIR3 imaging spectrometer at VTT imaging lab*

Content description:

12. Description and relation to the project objectives: *Dataset for T2.2.1: Textile composition study with fiber blends, relates to SO2*
13. Keywords: *hyperspectral imaging, near infrared, fiber blends*
14. Geographical location: *Espoo, Finland*
15. Code list used: *NA*
16. List variables: *pixels (frames x columns), wavelengths*
17. Metadata schema used: *NA*

Technical description:

18. Software used to create the file: *LUMO*
19. File format(s): **.raw, *.hdr*
20. Necessary software: **.raw and *.hdr compatible*
21. File size: *ca. 2 GB*

Sharing and preservation:

22. Use / Users: *Research*
23. Rights / License: *TBD*
24. Dissemination: *Upcoming research article*
25. Access: *Open, permission based*
26. Restrictions: *None (but some part of data might be proprietary)*
27. Repository for open data: *CSC IDA and Zenodo*
28. Permanent identified (e.g. DOI, URN): *TBD*