



FLEX 4 FACT

Industrial flexibility platform
for sustainable factories

DATA MANANGEMENT PLAN

DATE OF SUBMISSION: 30/11/2022

Deliverable No.: D8.2

Deliverable lead: Paul K. Wan

Authors(s): Ragnhild Eleftheriadis, Paul Kengfai Wan

DELIVERABLE DESCRIPTION

Project Acronym	FLEX4FACT
Call ID	HORIZON-CL4-2021-TWIN-TRANSITION-01-21
Project title	FLEX4FACT - Industrial Cluster FLEXibility platform for sustainable FACTories to reduce CO2 emissions and to enable the Energy Transition
Grant Agreement No	101058657
Start of Project	01.06.2022
Project Duration	42 months
Deliverable No.	D8.2
Work Package	WP 8
Associated Task	Task 8.2
Deliverable Lead	Ragnhild Eleftheriadis
Deliverable Authors	Ragnhild Eleftheriadis, Paul Kengfai Wan
Status	Draft
	Verified
	X Final version
Due date	30/11/2022
Submission date	30/11/2022

DOCUMENT HISTORY

Date	Version	Author	Comment
12/10/2022	V01	Paul Kengfai Wan	Comment
08/11/2022	V01	All partners	Add information, comments and iterations, review

30/11/2022

V1.0

Ragnhild
Eleftheriadis

Corrections and last finish

DISSEMINATION LEVEL

PU – public, fully open	X
SEN – sensitive, limited under the conditions of the Grant Agreement	

TABLE OF CONTENT

LIST OF TABLES	4
LIST OF FIGURES.....	4
SUMMARY	5
DEVIATIONS	5
1 INTRODUCTION AND DATA SUMMARY	6
1.1 INTRODUCTION	6
1.1.1 Intended readership.....	6
1.1.2 Deliverable structure.....	6
1.2 DATA SUMMARY	6
1.2.1 Data related actions in flex4fact	7
1.2.2 Data format and size.....	8
2 FLEX4FACT-FAIR DATA MANAGEMENT	8
2.1 MAKING DATA FINDABLE	8
2.2 MAKING DATA ACCESSIBLE	9
2.2.1 Means of publication.....	10
2.3 MAKING DATA INTEROPERABLE	10
2.4 RE-USABLE DATA	11
3 ALLOCATION OF RESOURCES.....	12
3.1 ALLOCATION OF RESOURCES.....	12
4 DATA SECURITY	12
4.1 DATA SECURITY	12
5 ETHICAL ASPECTS.....	12

5.1 ETHICAL ASPECTS 12

6 OTHER ASPECTS 13

7 CONCLUSIONS 14

8 ANNEX 15

LIST OF TABLES

Table 1 Datatypes and means of publication 10

LIST OF FIGURES

No figures entries found.

ACRONYM	DESCRIPTION
CA	Consortium Agreement
GA	Grant Agreement
DMP	Data Management Plan
HEU	Horisont Europa

SUMMARY

The Data Management Plan (DMP) describes how the Flex4Fact project manages data throughout its life cycle, in order to be compliant to the regulatory framework. This deliverable outlines how the data will be managed and shared with.

This deliverable sets the initial guidelines for how data will be generated in a standardised manner, and how data and associated metadata will be made accessible. DMP is intended to be a living document in which information will be made available on a finer level of granularity through updates, as the project implementation progresses.

DEVIATIONS

This deliverable will affect all the Work packages in FLEX4FACT project.

1 INTRODUCTION AND DATA SUMMARY

1.1 INTRODUCTION

The DMP provides an easy overview of research data the project was expected to generate, the types and formats of this data, and how this data is processed and stored to make them findable, accessible, interoperable, and re-usable, according to the principles of FAIR data management.

The purpose of the DMP is to contribute to good data handling during the project's lifetime, and to describe how such data will be curated and preserved.

1.1.1 INTENDED READERSHIP

Internally in the project:

- All participants in FLEX4FACT are responsible for, or in any way involved with-in collecting data and data handling can use this document for instructions on how to handle, store and process data.
- Participants can use this document to get an overview of data collected in the project and how this is processed, stored and made accessible.
- External audience:
- The Data Management section can be used by all relevant stakeholder who are interested in Quality related to FLEX4FACT project.
- How to reuse of data in own activity will also be provided in the end face of the project.

1.1.2 DELIVERABLE STRUCTURE

This document is structured as follows:

- Section 1: will introduce a chapter describing the main purpose, structure and intended readership of the DMP, as well as relationships to project deliverables.
- Section 2: provides an overview of the research data generated in the Flex4Fact project, with details on types, formats, origin, and metadata provisions.
- Section 3: describes how the project will comply with the principles of data management in HEU 2020 and SINTEF DMP policies
- Section 4: describes the resources allocated to making as much as possible open data.
- Section 5: give a detailed description of data security arrangements
- Section 6: deals with ethical and other aspects related to data management

1.2 DATA SUMMARY

The FLEX4FACT project aims to is to develop an end-to-end ecosystem that increases the flexibility in manufacturing processes and energy flows from North to South in Europe while enabling the increase of (a) awareness and knowledge about consumption patterns, energy behaviours, generation/demand forecasts and (b) local intelligence via properly established digital twin-based

and consumer-centric energy management and decision support mechanisms that locally optimize demand response. Through a structured methodology, the project partners will work towards achieving the project objectives, which will be expressed via the formulated use cases and evaluated mainly on the basis of primary data received from five pilot demonstrations, making use of cutting-edge methods and tools.

The purpose of the current Data Management Plan (DMP) deliverable is to provide updated information concerning the data that will be collected and used by the project partners and, eventually, be openly accessible.

In the present document, the term 'data' refers to five main categories of information that will be used and/or produced during the project. These are:

1. Datasets, as a result of data collection through the operation of the Flex4Fact system modules during pilot testing and/or as analysis results of the procedures that will take place as part of the modules operation.
2. Project deliverables that will be openly accessible to people interested in the project insights (the first DMP is considered as a public deliverable).
3. Dissemination material, such as technical reports, workshops, videos, blogs, presentations, posters, and flyers.
4. Software artefacts, such as open-source code, produced by the various components that make up the Flex4Fact system and accompanied by the appropriate documentation, could be considered sensitive.
5. Scientific publications referring to the procedures and/or the results of the research conducted.

Note that the afore mentioned data may include both personal and non-personal data.

1.2.1 DATA RELATED ACTIONS IN FLEX4FACT

Data management constitutes an integral part of the work conducted throughout the Flex4Fact project lifecycle. The Project Data term refers to all proprietary data generated within the project operations and transactions, e.g., documents, reports and related information, and, more importantly, user data obtained, possessed, and processed in the context of the provided services

There has been a broad discussion among the project consortium members regarding the project data needs. The work performed under WP1 'Definition of industrial use cases and FLEX4FACT Reference System Architecture, which is anticipated to provide outcomes as part of deliverable D1.1 'Industrial Use cases description' (due in M9).

This is considered as a first attempt to understand the status at the start of the project for each industrial setting, including mapping of the processes, known constraints, current IT infrastructure, measurements, and data flows already available, control systems, challenges, and objectives.

1.2.2 DATA FORMAT AND SIZE

A dataset may include different types of formats. Flex4Fact will use widely accepted formats for data collection and generation (with the appropriate extensions) as follows:

- Documents/Reports/Publications: .PDF/A, txt, doc/docx
- Spreadsheets: .xls/.xlsx
- Databases: .cvs
- Audio files: .mp3, .wav, .wma, .ra
- Pictures: .jpg, .png
- Videos: .avi, .flv, .mov, .mp4, .wmv

The exact data size cannot be pre-determined, although this kind of information will be requested for each developed dataset. The expected size of individual datasets is in the range of a few MB and 20 GB, with a typical value of about 300 MB per file. However, raw datasets can include several files and these data will preferentially be processed offline by individual experts, while the aggregated data will be stored in designated project data (cloud) stores.

2 FLEX4FACT-FAIR DATA MANAGEMENT

The Flex4Fact project has followed the FAIR principles for research data, i.e., data should be **F**indable, **A**ccessible, **I**nteroperable and **R**eusable (**FAIR**). To do so, Flex4Fact Consortium members have agreed on establishing a series of criteria to make data findable by other users (e.g. using metadata standards, adding keywords, and DOIs, etc.); to address which data may be accessible via open repository or should remain confidential according to ethical requirements; to foster the interoperability of the data by allowing data exchange and reuse (e.g. using standards, or open source); and, to establish the licences of the data generated to permit the widest reuse possible.

2.1 MAKING DATA FINDABLE

In Flex4Fact project, standardized naming and identification mechanisms are used to allow effective differentiation of the produced datasets. According to extended practices to standardise file names,

FLEX4FACT complies with the following general specifications for file naming:

- File names should be short but descriptive (<40 characters);
- Use alphanumeric and avoid special characters, such as "/\ : * < > [] \$ & ~ ! # ? { } ' ^ %
- Use underscores _ instead of periods or spaces;
- Standardise dates (according to ISO standards: YYYYMMDD);
- Use leading zeros for sequential numbering, e.g., v01

Moreover, Flex4Fact file names contain the following information:

- Project acronym (i.e. Flex4Fact) – compulsory for deliverables and data openly shared.
- Document name (e.g., Data Management Plan) – compulsory for all documents.
- Researcher/Authors initials – compulsory only for internal documents.
- Acronym institution leading the research – recommended for open data.
- Year of study (e.g. 2018) – when necessary.
- File version (e.g. v01) – compulsory for internal documents;
- File type/extension (e.g. odt) - compulsory for all documents.

Examples:

- For EC deliverables to portal: 20221020 FLEX4FACT DataManagementPlan D8.2 Vx.PDF
- Dissemination materials: Flex4Fact _DataManagementPlan.PDF;
- For public research data: Flex4Fact _DataManagementPlan_FCID_v01.PDF;
- For internal/working documents: 20180518_DataManagementPlan_EM_v01.PDF.

All data will be collected in an own Excel file, sorted on WP's and available for whole the consortium.

2.2 MAKING DATA ACCESSIBLE

All the research outcomes (i.e., journal articles, policy briefs, reports, etc.) are published by default in Open Access. Furthermore, public research data generated have been made openly accessible via the online repository ZENODO, as soon as the research has been published. Flex4Fact has been storing research data (e.g., databases, working documents, reports, research articles, etc.) in ZENODO from the beginning of the project and it will continue until its end. ZENODO is a free of charge repository developed by CERN within the EU Open AIRE project to support Open Data management for EC funded multi-disciplinary research.

In general, the Flex4Fact consortium focuses on ensuring open access of research data. However, because of the sensitive nature of some data parts, in certain cases these parts may remain confidential and not be accessible to third parties. The data status in this regard is going to be indicated by each individual partner responsible for collecting and/or producing the data, with the support of the Data Management Committee and the supervision of the Executive Board.

2.2.1 MEANS OF PUBLICATION

Based on the data types available for publication, diverse suitable channels will be utilized for providing access to the Flex4Fact Open Data. The potential means of publication of the various data types are presented in Table 2. Among all publication means, only the project website and its social media accounts (Twitter, LinkedIn, and YouTube) have been utilized up to this point. All other channels will be activated when appropriate material for publication through them is available.

The draft of a proposed catalogue can be found at annex A and how a publications catalogue can be, this will be found in annex B.

Table 1 Datatypes and means of publication

DATA TYPE	MEANS OF PUBLICATION	DATA TYPE	MEANS OF PUBLICATION
Datasets		Project Website, Zenodo, OpenAIRE	
Deliverables		Project Website, Zenodo, OpenAIRE	
Dissemination material		Dissemination material	
Project Website, Social Media, OpenAIRE, Zenodo		Project Website, Social Media, OpenAIRE, Zenodo	
Software		Software	

Datasets with dissemination level tag "confidential" (non-anonymous datasets) will not be shared due to privacy concerns. Instead, they will be archived in a secure repository with controlled access to authorized consortium members only.

2.3 MAKING DATA INTEROPERABLE

Interoperability of the data (i.e. allowing data exchange and re-use of the research data between researchers, institutions, organisations or countries) has been facilitated by following good practices and standards for research data. This aspect of the DMP focuses on the assessment of data interoperability and defines the data and Metadata vocabularies, standards or methodologies that need to be followed to enable interoperability. In addition, it defines whether standard vocabulary will be used for each data type to allow inter-disciplinary interoperability. Within the context of Flex4Fact, the following guidelines are to be followed for the various types of data towards ensuring interoperability:

- If the data are part of a domain with well-known open formats that are in common use, these formats should be selected.
- If the data do not fall in the previous category, an open and easily machine-readable format should be selected.

2.4 RE-USABLE DATA

Research data generated by the project can be made publicly available for reuse according to the type of data and journal embargo policies or in common understanding related to the CA and GA. In most of the cases, data have will be distributed as soon as research has been accepted for publication by the partners involved in FLEX4FACT. The reuse of FLEX4FACT research data by third parties (i.e. other researchers, policy-makers, other societal actors) is expected during and after the project ends. Whenever possible, public research data have been shared under the licence **Creative Commons (CC) Attribution (BY) version 4.0 (CC BY 4.0)**

In order to allow the widest reuse possible. As the process of decision making related to Data Use licences is currently in progress, as part of the work in several WP. This will be listed in the next version of D8.4 Data Management Plan Update since some of these data can be sensitive for the end-user and contain personal data. That version will be marked sensitive.

3 ALLOCATION OF RESOURCES

3.1 ALLOCATION OF RESOURCES

The costs for making Flex4Fact research data FAIR (i.e. findable, openly accessible, interoperable, and reusable) while securing any personal data collected, are detailed in the GA and CA. These costs is covered by the financial budget of the project, and include Open Access (OA) publications if possible, ICT services such as secure servers and Internet domains, and the development of code for a self-administered questionnaire (web-based).

4 DATA SECURITY

4.1 DATA SECURITY

Data security is of major importance in the Flex4Fact project. Although Flex4Fact has not dealt with any sensitive data, personal data of participants have been collected and stored during the first part of the development of the project (e.g., name, profession, email), being destroyed after the completion of the research. The protection of personal data has been ensured through appropriate procedures for data collection, storage, protection, and destruction, according to the current EU data protection regulations for data protection and ethical principles (i.e. GDPR).

This means written material is stored in secured locations (institutional servers and hard drives), and personal data will be anonymized (coded) in order to make it impossible to identify individuals. Only the person in charge of the study had access to the list with names and codes to allow participants to withdraw the data at any point in time during the study. Moreover, personal background-related data have been stored separately from research relevant data.

5 ETHICAL ASPECTS

5.1 ETHICAL ASPECTS

The ethical and data protection frameworks that guide the project's research practices are in line with the EU key ethical principles and research codes of conduct, as well as the current regulations on Data Protection (General Regulation on Data Protection from 25 May 2018, and the EU Directive 2016/679). Thus, ethical principles and research codes of conduct concerning privacy and personal data protection have been addressed.

These interrelated issues were respectively addressed in detail under the Grant Agreement and the Consortium Agreement.

The Flex4Fact project involves human participation, mainly in the activities related to the pilot demonstrations, which are to be conducted in real buildings of different uses. During these activities, it might be necessary to collect basic personal data (e.g., name, background, contact details). The collection and processing of personal data will be performed in accordance with the applicable principles of the European Data Protection Law, including the data minimization principle and will be handled and stored safe.

Raw data coming out of the pilot cases and possibly including personal data are not considered for sharing at this point. If personal data is considered for sharing at a later project stage, this will be performed following the instructions of the pilot case. In any case, all necessary organizational and technical measures will be implemented for digitising any manually collected data and storing both, manually and automatically collected data, in the Flex4Fact central Database. Based on the project Methodology, the operating partner of the Core Data Management Platform is technically responsible for the collection, security, storage, and handling of the related data.

6 OTHER ASPECTS

The overall responsibility for data management lies with the Project Coordinator's DPO and can be reached at privacy@sintef.no.

The coordinator of the FLEX4FACT project, SINTEF, follow ethical guidelines in its work, and all work conducted by SINTEF is subject to the SINTEF Ethics Council and the appointed Ethics Representative. SINTEF will also ensure that all participants in the FLEX4FACT project follows the ethical guidelines of SINTEF as good as possible.

Important aspects with respect to this are:

- The ethical guidelines are based on the vision of using science and technology to create a better society and are reviewed continuously to ensure they stay up to date with developments in society and the challenges of today. They generally fall into these categories: research ethics, business ethics, and ethics in interpersonal relationships.
- SINTEF is a member of the UN Global Compact and Transparency International, and SINTEF's ethics are guided by the principles highlighted by these organisations, as well as based on the regulations of the national ethics committees, the principles promoted by the European Group on Ethics in Science and New Technologies (EGE), and on international conventions such as the Vancouver Convention. When dilemmas of research ethics require an assessment beyond the scope of our guidelines, our Ethics Council and Ethics Representative, we refer to statements from the EGE.
- All SINTEF's employees are expected to act in accordance with the ethical guidelines and principles. As coordinator of the FLEX4FACT project, SINTEF will ensure that any ethical issues, which may arise, will be handled appropriately and in a transparent and fair manner.

7 CONCLUSIONS

As coordinator of the FLEX4FACT project, SINTEF will ensure that any data management issues which may arise during the project will be handled appropriately and in a transparent and fair manner.

The DMP is a living document that will expand as the project evolves and new information on data collection, generation and handling arise.

Day to day data management will happen through the SharePoint Excel tool described in this document, and through continuous collaboration between the partners responsible for data collection and generation in the project. A revised and extended version of this DMP will be prepared in M23 when all data is collected and reflects the current status of data management in the project.

8 ANNEX

Annex A

An Internal Excel File that contains information of all deliverables with specific information on the data collected in each WP.

A full list from each WP will be shared in the Update version of this DMP.

No	Deliverable	Due Date/ Month	Deliverable Type	Dissemination Channels	EC Portal	Dissemination Status
1.	D8.1 Project Handbook	M3	Report and online manual	Share Point	Uploaded	Project only

Annex B

All kind of data collected and shared in the dissemination phase. Pictures, videos, and other personal data that can be sensitive will be listed, most of them in WP7. All partners are informed about this type of activities and can make a reservation of exposition.

Type of Publication	Title	Description	Means of Publication	Release date	Link
Videos	Flex4Fact	A video about the objectives and activities of Flex4Fact	YouTube All informed	Date no decided yet	SharePoint
Press release	Flex4Fact Press Release	General Information from kick off Flex4 Fact	Pictures, all informed on a release	01/06/2022	SharePoint