



## D7.2 Data management plan

*Flying Forward 2020 is funded by the European Union H2020  
Research and Innovation Programme under Grant Agreement No. 101006828*

**Contents**

- CONTENTS..... 2**
- DOCUMENT IDENTIFICATION ..... 3**
- ACRONYMS AND ABBREVIATIONS..... 4**
- 1. EXECUTIVE SUMMARY ..... 5**
- 2. INTRODUCTION ..... 5**
  - 2.1 FF2020 PROJECT OVERVIEW ..... 6
  - 2.2 FF2020 DATA LIFE-CYCLE ..... 6
- 3. FF2020 DATA MANAGEMENT PLAN GENERATION..... 8**
  - 3.1 DATA COLLECTION AND GENERATION ..... 9
  - 3.2 FAIR DATA ..... 9
- 4. ALLOCATION OF RESOURCES..... 11**
- 5. DATA SECURITY ..... 12**
- 6. ETHICAL ASPECTS ..... 13**
- 7. LIST OF FIGURES ..... 15**
- 8. LIST OF TABLES..... 15**

## Document Identification

Authors	Eva Elena (BRAIN); Ruud van Iwaarden (BRAIN)
Internal reviewers	Arthur Dallau (DIG); Guido Maglione (VERSES)
Work package	WP7 “FF2020 awareness-raising activities and business development”
Task(s) involved	T7.2 Dissemination and Communication
Nature	R (document): Report
Dissemination level	CO : Confidential

Version	Date	Contributor	Description
V0.0	02/02/2021	Eva Elena (BRAIN)	Description of table of content and initial inputs for all the sections.
V0.2	12/05/2021	Ruud van Iwaarden (BRAIN)	Input on all sections. Version for review by internal reviewers.
V1.0	30/05/2021	Ruud van Iwaarden (BRAIN)	Changes from comments and input. Final version

## Acronyms and Abbreviations

Acronym	Title
FF2020	Flying Forward 2020
DMP	Data Management Plan
FAIR data	data which meet principles of findability, accessibility, interoperability, and reusability
GDPR	General Data Protection Regulation
LLAB	Living Lab

## 1. Executive Summary

In line with promoting “Open Science” , Horizon 2020 projects are required to develop a Data Management Plan (DMP) to specify how research data will be handled both during and after the project. This report presents the first version of the DMP for the FF2020 project. Since the DMP is intended to be a living document, further updates on Data Management will be provided in the next reporting periods as the implementation of the project progresses and when significant changes occur.

In the end, the main goal of this document will be to provide the DMP of the FF2020 project, specifying all data used in the project, both collected and generated during the research, what data will have an open access, how they will be preserved and how they will be made accessible and reusable. Guidelines provided by the European Commission on FAIR Data Management in Horizon 2020 manual have been followed for the generation of this document (available on the Participant Portal<sup>1</sup>).

All partners involved in the Use Cases and research activities will contribute in the development of this deliverable by providing information about the data collected and/or generated throughout the project.

In this first version, we report an initial analysis on how the amount of data produced in the project is intended to be managed. With this aim, we elaborated a template to identify first the data sets that will be produced by each partner, and then, to define the life cycle of those data sets. Some questions regarding FAIR data, costs, or IPR will be answered in following versions.

As a starting point, we present the following elements of the Data Management Plan:

1. Introduction of the document in the context of the FF2020 project
2. Guidelines for the generation of the DMP. A data summary is presented: Type of data generated, collected and stored, and relevant data features description.
3. Data security, data sharing and specifications on security aspects.
4. Ethical and legal aspects regarding data.

## 2. Introduction

This document corresponds to the deliverable D7.1, belonging to the framework of WP7 (Exploitation, Training, Dissemination & Standardization), and describes the Data Management Plan (DMP) of the FF2020 project. DMP is a short plan that outlines what data will be collected and generated and how to ensure having well managed data and FAIR data: Findable (access, storage, backup, ...), Accessible, Interoperable and Reusable Data.

---

<sup>1</sup> [https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-accessdata-management/data-management\\_en.htm](https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-accessdata-management/data-management_en.htm)

The purpose of this DMP is to set standard procedures and formats to be followed by the partners involved in each of the use cases, while defining platforms and repositories for data to be shared and/or stored. Moreover, the living aspect of the document makes it possible for partners to update any relevant or missing information that was not accessible at the time the document was created, resulting in a more flexible and complete Data Management Plan.

## **2.1 FF2020 project overview**

An overview of the FF2020 project is presented in this section, so that this document can be aligned with the project's objectives and evolve regarding the necessities that may emerge related to data and their management.

The FF2020 project's main objective is to help society to fly forward in a safe, secure and effective way to make life easier, cheaper and providing more opportunities by getting products faster and more efficiently, relying on how new digital technologies can help create a new balance between air and ground mobility, adopting an innovative "whole-of-society" approach, which we believe is the best way to integrate safety, security, sustainability and business-efficient services. In doing so we can create a competitive edge for Europe and develop a fair and inclusive footprint that would last for generations to come whilst engaging private sector, civil society, communities and individuals.

As our physical world is interconnected with our digital world, society has endured growing pains. In FF2020, our vision relies on how new digital technologies can help create a new balance between air and ground mobility, adopting an innovative "whole-of-society" approach, which we believe is the best way to integrate safety, security, sustainability and business-efficient services. In doing so we can create a competitive edge for Europe and develop a fair and inclusive footprint that would last for generations to come whilst engaging private sector, civil society, communities and individuals.

One of the goals of the project is the Digital Toolbox development to tackle these challenges and create a technology agnostic platform in a form of marketplace, whereby the goals of cities to be sustainable and adaptive and tech providers to be agile can meet each other. To meet the need, the platform serves as a procurement tool for cities and as an integration and implementation tool for the tech vendors. Besides the digital toolbox, there are other project parts where research data might play a role. For instance the regulatory framework; governance framework; geospatial framework and interoperability framework. Underlying data and results that are considered open data will be considered according to this data management plan. During the project assets will be added in the form of the excel data management sheets.

Collecting of data management sheets from partner starts from the publication date of this document.

## **2.2 FF2020 Data Life-cycle**

Characteristics and features of all the data generated, collected, processed, analyzed, stored, and/or eliminated during any of the data lifecycle stages must be defined. Herein, a workflow

to be followed to make data FAIR (Findable, Accessible, Interoperable and Reusable) and specifications about data preservation (or elimination) are defined.

The management of research data is a fundamental part of any research process in order to provide a clear overview of the data life cycle and to facilitate accessing and reusing data, not only during the project, but also in the future.

The data lifecycle of a project includes six phases: creation of data, data storage (repositories), use of data, sharing (between users, partners, or external researchers), archive (whenever data are stored for future activities) and destruction of data (if necessary). See Figure 1



*Figure 1. Data lifecycle*

In the case of FF2020, we must address data management during every life-cycle stage: (1) from the collecting data point on the edge (See Use Cases), (2) the generation, modification, or cleaning of data during its processing (FF2020 platform), and (3) the storing relevant data and metadata to promote sharing and reusing data for the aim of collective intelligence.

As far as it concerns use cases, different data types must be dealt with depending on the individual requirements of each use case. This fact makes it necessary for different processing and storing tools to be available and defined, as well as what file formats to be used.

An efficient Data Management must go deeper in this context and also define other data characteristics, i.e. container architecture (if necessary), usage of standard file formats, data loss risks, among others, to ensure that full comprehension and management of data are accomplished during the project.

### 3. FF2020 Data Management Plan Generation

Each of the partners involved in the FF2020 project will collaborate in order to fulfil and update the DMP document along the project. Keeping the document updated is a key aspect in this regard, as some relevant information about data is not available nor well defined at the start, but in further stages of the project.

sharing and collaboration between partners is also of key importance to achieve the objectives of the project. For this reason, research data and metadata obtained as outputs from the use cases must be stored and shared among researchers in order to ensure that a more robust and agile research is performed.

Due to the number of partners participating in the project, some guidelines are defined so the DMP completion procedure is standardized. Large data sets will be gathered and stored according to national and European legislation frameworks and standards. The guiding principle must be the Horizon 2020 effort to create re-usable datasets for sustainable, comparable, and growingly valid and reliable research outcomes.

Given the variety of data types involved in each of the Use Cases, a common table format has been prepared that is filled by all partners, including information about all data collected and generated throughout the project (Context Data, Research Data and Aggregated Data), as well as the FAIR Data issues. These tables will be included in Annex I.

For each partner, a set of questions are presented (See Annex I), related to some general aspects of the data to be used. After completion, and as a guide for partners, the following questions must have been answered:

- What is the purpose of the data collection/generation and its relation to the objectives of the project?
- What types and formats of data will the project generate/collect?
- Will you re-use any existing data and how?
- What is the origin of the data?
- What is the expected size of the data?
- To whom might it be useful ('data utility')?

Additionally, a second table must be filled by each partner. The purpose of this second table is to specify the steps followed to make the data FAIR. More information about FAIR data can be found in section 3.2.

These tables are meant to ensure an efficient management and governance of the data, and set common standards, formats and requirements for all partners to use. The partners are asked to try to use common, open, and widely known data formats, as well as facilitate interoperability between external datasets from other projects. Notice that other data aspects like the ownership of the data generated during the project will be updated once the project advances.

Lastly, the partners may recall that FF2020 operates in the open data pilot<sup>2</sup>. This means that data generated and collected should be shared and openly accessible by researchers, through open access repositories.

In this regard, a common repository for data to be shared between partners during the project will be available, using Teams, SmartSheet and where applicable Github another platform suited for sharing research data. For now we plan on using Github, though this choice might change during the project. (see Section 5 - Allocation of Resources). The access to the Github platform will be granted once it is available. Partners can freely decide what data, code or information to share, including the possibility to keep data for themselves.

### 3.1 Data Collection and Generation

Each of the Use Cases in the FF2020 project has a common pathway for data treatment. On the other parts of the project, the cloud can be accessed and used for storage purposes, but most of the data life-cycle occurs at the edge.

There is still a very wide spectrum of data types involved and the purpose of each data type may vary from one to another. For simplicity, they have been classified in three types: Context Data, Research Data and Aggregated Data.

- **Context Data:** These refer to any data obtained during measurements in the Use Case environment, i.e. environment specifications (limitations, necessities, presence of humans in the scenario, obstacles, etc), and they are used to obtain any kind of information from the environment. Context data can be collected, generated or retrieved from already existing data. The available dataset is specified for each partner in Annex I.
- **Research Data:** Any data that is collected during the course of the study, for research purposes. They also include any derived data that may result from the direct transformation of the original data (by analysis, fusion, filtering...).
- **Aggregated Data:** Data created in order to answer research questions. They are generated from data reduction processes, which mainly summarize the most important aspects in the raw data into relevant parameters that can be used for analysis. These reduction processes can be validation, curation, conversion, etc.

### 3.2 FAIR Data

This document should ensure that the partners follow the “FAIR Guiding Principles for scientific data management and stewardship”<sup>3</sup>. These principles set up the standards for data to be easily findable, accessible, interoperable and reusable not only by humans, but also for

---

<sup>2</sup> [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

<sup>3</sup> Wilkinson, M.D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., ... Bourne, P.E. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018, doi: 10.2038/sdata.2016.18

computational systems. These FAIR Guiding Principles are presented below Table 1, as they were published in the original source:

*Table 1 FAIR principles*

The FAIR Data Guiding Principles
To be Findable
F1. (meta)data are assigned a globally unique and persistent identifier
F2. data are described with rich metadata (defined by R1 below)
F3. metadata clearly and explicitly include the identifier of the data it describes
F4. (meta)data are registered or indexed in a searchable resource
To be Accessible
A1. (meta)data are retrievable by their identifier using a standardized communications protocol
A1.1 the protocol is open, free, and universally implementable
A1.2 the protocol allows for an authentication and authorization procedure, where necessary
A2. metadata are accessible, even when the data are no longer available
To be interoperable
I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation
I2. (meta)data use vocabularies that follow FAIR principles
I3. (meta)data include qualified references to other (meta)data
To be reusable
R1. (meta)data are richly described with a plurality of accurate and relevant attributes
R1.1 (meta)data are released with a clear and accessible data usage license

R1.2 (meta)data are associated with detailed provenance
R1.3 (meta)data meet domain-relevant community standards

These FAIR principles provide a definition of the practices and methods that data resources and infrastructures should follow in order to support discovery and research.

As said before, in this first version of the DMP some information regarding FAIR data will be provided, however, most of the issues related to this question will be accessible in future updates of the document.

In Annex I, information from each partner regarding the FAIR data specifications can be found.

Upon completion of the tables in Annex I, the following questions must have been answered:

**Findability:** ‘Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism?’, and ‘What metadata will be created?’

**Accessibility:** ‘Which data produced and/or used in the project will be made openly available as the default?’, and ‘How will the data be made accessible?’

**Interoperability:** ‘Are the data produced in the project interoperable?’

**Re-usability:** ‘How will the data be licensed to permit the widest re-use possible?’

#### 4. Allocation of resources

In the FF2020 Project we distinguish two types of information, the project management information, e. g. documents, meeting minutes, deliverables, templates, dissemination material, etc.; and the software developed, use case data and test data. Given the characteristics of each type of information, each of them will be allocated in a different repository.

The BRAINPORT’s Microsoft Teams and SmartSheet repository, is the tool that is being used as a private area for preparing and sharing documents and knowledge among the members of the FF2020 Project. BRAIN will give access, as a guest, to the individual members of F2020 partners. WP leaders and use case leaders are free to create specific working spaces to deal with their day-by-day activities. This repository is also used for backups or for archiving of documents.

As mentioned before, it is a BRAIN private repository, so it doesn’t involve any cost to the project. It is managed by BRAIN, and the costs derived from this activity are already considered the efforts of the project. Long term preservation of this information will be discussed later in the project and it will be included in an updated version of the DMP.

For hosting any applicable developed software and hardware (source code), use cases data and test data, GitHub a collaborative development platform based on Git, an open-source distributed version control system, will be used.

GitHub has been selected as the code repository for the project because it is the most widely used repository, offers secure cloud storage, and is free for teams, offering unlimited public and private repositories for unlimited collaborators.

Regarding hardware and software integration two possible scenarios are considered:

- Integration from partners that use their own infrastructure. In this case, the basic idea is that at the end of the development iteration, they upload their changes to the centralised GitHub, for further integration with other partners development.
- Integration from partners that do not use their own infrastructure. They can use the centralized GitHub to host their day-to-day Git repositories (and commit to the centralised Git repository).

As Git is free and GitHub is a free online service (the version that will be used in the project), it won't involve any cost to the project. The costs related with the management of the repository are included in the efforts of the project. This repository will be managed by LKS.

The use of free platforms such as GitHub reduces the cost of long-term preservation to zero. Moreover, since Git is a distributed version control system each user that clone for their own use the data repository has a complete copy of the data set (history, changes, meta-data, etc.). However, long term preservation of this information will be discussed later in the project and it will be included in an updated version of the DMP.

## 5. Data Security

As said in previous sections, the two main tools for sharing data are:

- Microsoft Teams and SmartSheet repository for documentation related with FF2020.
- GitHub for the code required for some of the work packages.
- The projects or living labs cloud infrastructure (during project)

One of the most important concepts in security is that there is no completely secure system. It is simply not possible to prevent any current and future attack. If a new way to attack a system is found in the future, a system designed in the past can do nothing against it. Nevertheless, what a system can do is to have a state-of-the-art security when made public and allow to be updated when new vulnerabilities are found before it affects its users. And this is what our solutions do.

According to Microsoft: "Microsoft Teams is designed and developed in compliance with the Microsoft Trustworthy Computing Security Development Lifecycle (SDL), which is described at Microsoft Security Development Lifecycle (SDL). The first step in creating a more secure unified communications system was to design threat models and test each feature as it was designed. Multiple security-related improvements were built into the coding process and practices. Build-time tools detect buffer overruns and other potential security threats before the code is checked in to the final product."

For enterprise customers, Microsoft stores the data from their 365 services, like Teams or Sharepoint, in datacenters nearest to the business location provided when the company create its tenant.

In case of tenants created with a billing address in any European Union country, the datacenters are located in the European Union under European laws.

GitHub also has extremely high security standards. Their solution is GDPR compliant, they encrypt all the data in transit and store the hash of all passwords using bcrypt, one of the most used hashing solution in the industry. It also prevents bruteforce attacks by limiting the number of attempts before locking.

One of the advantages of using these solutions is that both are prepared to handle multiple teams geographically separated and concurrent working. For this reason, all the information is secured in a third party storage, without requiring trust among partners. As said before, all the information is end-to-end encrypted, assuring privacy.

All the data in GitHub has redundancy and history, meaning that losing all the data at once is really unlikely and changes can be tracked among partners.

Also, Microsoft Teams corporate not only integrates Office365 in their solution, but also has redundancy in the backend, making the data accessible even if a given server fails.

Obviously, as the data is stored at a trusted third party, it has the required trusted certificates. Both solutions promise long term preservation and as the biggest players in each of their segments of the markets, they are trusted by millions.

## 6. Ethical Aspects

Data sharing and the Open Access implementation to the scientific research process is clearly enhancing scientific progress, while also benefiting transparency and reproducibility. However, this also questions the implications and impact that making specific kinds of data public and accessible may have. Several discussions have taken place on this respect, however a general consensus in the scientific community has not yet been reached.

This chapter describes what impact the data in the FF2020 project will have on ethical aspects, specifically when these data are related to personal information. The partners involved in the project may be aware that neither during or after the project, any of the data shared/stored will have an impact on ethical, gender, or other personal circumstances.

To this regard, the partners must follow a set of practices while involved in data treatment:

- Ensuring that the citizens' rights on their own data are protected.
- Confidentiality and privacy must be managed so that shared datasets are anonymized and people cannot be identified from shared data.
- Consent from people whom data are being obtained from must be validated, keeping clear the objective and purpose of the data to be obtained from them.
- Analyze the potential impact and implications on further research, in a way that further results obtained from shared data can be validated and trustworthy.

As a rule of thumb, partners and participants shall follow the statement: **“All participants in the FF2020 project will conform to GDPR and the current legislation and regulations in the countries where the research will be carried out”**.

Partners must ensure that the ethical and societal aspects are incorporated into the design process from the beginning. They must also allow security and validation of operational efficiency and transparency, possible interpretative distortions, liability for possible damage, patentability, anonymisation of user data and privacy in compliance with the guidelines developed by the European Commission Expert Panel<sup>4</sup> that has developed and continues to develop guidelines for the design of reliable artificial intelligence systems, respecting the centrality of the human being.

Living lab demonstrators might collect and use personal information. LLab partners will ensure compliance of the UE 679/2016 of the European Parliament and of the Council which entered into force from May 2018 on the General Data Protection Regulation (GDPR) of individuals with regard to the processing of personal data and on the free movement of such data. Collected data at all stages will be treated with respect to personal anonymity. Only when necessary, participants’ personal data will be legally obtained after an informed consent. Data will be securely accessed and privacy protection measures will be undertaken proportionally to the risks involved and the sensitivity of the data, such as password protection, encryption in all transmissions, etc. Identification data will be encrypted and strictly separated from sensitive data such as health data. Users’ personal data will be collected for processing, and undergo such processing, only if they are adequate, relevant and not excessive in relation to the scope and the specified, explicit and legitimate purposes for which they will be obtained.

Also, personal data subjected to processing will not be used for purposes incompatible with those for which they will be collected. Further processing of the data for historical, statistical or scientific purposes shall not be considered incompatible. In addition, personal data will be erased when they will have ceased to be necessary or relevant for the purpose for which they will be obtained or recorded. They will not be kept in a form which permits identification of the data subject for longer than necessary for the purposes for which they will be obtained or recorded. On a regular basis, the procedure will be determined by which, exceptionally, it will be decided to keep the entire set of particular data, in accordance with the specific legislation, because of their historical, statistical or scientific value. The researchers will ensure that outcomes will be reported and will not contravene the right to privacy and data protection. They also will carefully evaluate and report the personal privacy implications of the intended use or potential use of the research outcomes.

Additional information on each of the individual Use Cases can be found below. This information accounts for any sensitive data that may be involved in any of the Llabs, either during research, development, or storage of data for further purposes, and is only available for those Llabs that deal with data that may have any ethical impact.

---

<sup>4</sup> <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52011DC0882&from=EN>

## **7. List of figures**

Figure 1. Data lifecycle.....	7
-------------------------------	---

## **8. List of tables**

Table 1 FAIR principles.....	10
------------------------------	----

# FF2020



This document (including any enclosures and attachments) has been prepared for the exclusive use and benefit of the addressee(s) and solely for the purpose for which it is provided. Unless we provide express prior written consent, no part of this document should be reproduced, distributed or communicated to any third party. We do not accept any liability if this document is used for an alternative purpose from which it is intended, nor to any third party in respect of this report.

© 2020 FLYING FORWARD 2020. ALL RIGHTS RESERVED.



*Flying Forward 2020 is funded by the European Union H2020  
Research and Innovation Programme under Grant Agreement No. 101006828*