



ICARUS:

“Aviation-driven Data Value Chain for Diversified Global and Local Operations”

D6.3 – Project Data Management Plan

Workpackage:	WP6 – Dissemination, Communication and Stakeholders’ Engagement		
Authors:	SingularLogic (SILO)		
Status:	Final	Classification:	Public
Date:	29/06/2018	Version:	1.00

Disclaimer:

The ICARUS project is co-funded by the Horizon 2020 Programme of the European Union. The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the European Communities. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

© Copyright in this document remains vested with the ICARUS Partners.

ICARUS Project Profile

Grant Agreement No.:	780792
Acronym:	ICARUS
Title:	Aviation-driven Data Value Chain for Diversified Global and Local Operations
URL:	http://www.icarus2020.aero
Start Date:	01/01/2018
Duration:	36 months

Partners

	UBITECH (UBITECH)	Greece
	ENGINEERING - INGEGNERIA INFORMATICA SPA (ENG)	Italy
	PACE Aerospace Engineering and Information Technology GmbH (PACE)	Germany
	SUITE5 DATA INTELLIGENCE SOLUTIONS LIMITED (SUITE5)	Ireland
	UNIVERSITY OF CYPRUS (UCY)	Cyprus
	CINECA CONSORZIO INTERUNIVERSITARIO (CINECA)	Italy
	OAG Aviation Worldwide LTD (OAG)	United Kingdom
	SingularLOGIC S.A. (SILO)	Greece
	ISTITUTO PER L'INTERSCAMBIO ISI SCIENTIFICO (ISI)	Italy
	CELLOCK LTD (CELLOCK)	Cyprus
	ATHENS INTERNATIONAL AIRPORT S.A (AIA)	Greece
	TXT e-solutions SpA (TXT) – 3 rd party of PACE	Italy

Document History

Version	Date	Author (Partner)	Remarks
0.10	04/05/2018	Andreas Kalamaras (SILO)	Initial Draft Version (ToC)
0.20	18/05/2018	Andreas Kalamaras, Varitimos Slavidis (SILO)	Contributions to section 2
0.30	25/08/2018	Andreas Kalamaras, Varitimos Slavidis (SILO)	Contributions to section 3
0.40	01/06/2018	Andreas Kalamaras, Varitimos Slavidis, Stelios Pantelopoulos (SILO)	Contributions to section 4
0.50	08/06/2018	Andreas Kalamaras, Stelios Pantelopoulos (SILO)	Contributions to section 5
0.60	15/06/2018	Andreas Kalamaras (SILO)	Executive summary, Conclusions, Captions & References
0.70	22/06/2018	Andreas Kalamaras (SILO)	Minor corrections
0.75	25/06/2018	Andreas Kalamaras (SILO)	Review ready version
0.75_CINECA	27/06/2018	Claudio Arlandini (CINECA)	v.0.70 of the document reviewed by CINECA
0.75_UCY	27/06/2018	George Pallis (UCY)	v.0.70 of the document reviewed by UCY
1.0	29/06/2018	Andreas Kalamaras (SILO)	Final document version

Executive Summary

Each project in the EC's Horizon 2020 program has to define what kind of results are generated or collected during the project's runtime and when and how they are published openly. Giving particular attention to research data management, the ICARUS consortium follows a series of dedicated activities toward disseminating the project outcomes in a wider audience. This document describes initially the methodological framework taking into account the guidelines from EC. In particular, the Data Management Handling Plan is formulated in accordance with the H2020 guidelines regarding Open Research Data. In alignment with the EC Guidelines for Open Access, the aim of the current document is to clearly define how all public data assets issued by the project consortium will be further disseminated to a wider audience and to outline the methodology which can make the research data generated in the context of the ICARUS project, findable, accessible, interoperable and reusable (FAIR principles). It aims at providing not only the source of the data, the standards and metadata used for data handling, etc. for all the results generated or collected during the ICARUS project, but also at defining the process of how this data will be further exploited and/or shared/made accessible for verification and re-use, along with the data preservation and maintenance processes.

The current version constitutes the initial version of the Data Management Plan. The ICARUS DMP is intended to be a *living document* in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur.

Table of Contents

1	Introduction	8
1.1	Purpose of the Document.....	8
1.2	Structure of the Document.....	9
2	ICARUS Data Management Process.....	10
2.1	Data Management Process.....	14
2.2	Responsibilities and Decision Making.....	15
3	Data Archiving and Publication Infrastructure	17
3.1	ICARUS Nextcloud Repository	17
3.2	ICARUS Website	18
3.3	ResearchGate	19
3.4	Zenodo Project Repository.....	19
3.5	ICARUS platform	19
3.6	GitHub	20
4	ICARUS Datasets and Publications.....	22
4.1	ICARUS project datasets.....	22
4.1.1	Population Data	23
4.1.2	GLEAM Simulation output	24
4.1.3	Virus & infections data	24
4.1.4	Number of Passengers	24
4.1.5	Eurostat Air Transport Statistics	25
4.1.6	Eurocontrol Network Operations Portal (NOP).....	25
4.1.7	ICAO Engine Emissions Databank.....	26
4.1.8	European Aviation Environmental Report	27
4.2	ICARUS project deliverables	27
4.2.1	D1.1 Domain Landscape Review and Data Value Chain Definition	29
4.2.2	D1.2 The ICARUS Methodology and MVP.....	29
4.2.3	D6.1 Plan for Dissemination, Communication and Stakeholder Engagement	30
4.2.4	D6.2 Project Website and Web 2.0 Channels Setup	31
4.3	ICARUS scientific publications	31
5	Compliance with GDPR	32
5.1	Data Subjects	32
5.2	Personally Identifiable Information	34
5.3	Roles	34
5.4	Processing Activities	35

6	Conclusions	37
	References and Applicable Documents	38
	Annex I: FAIR template	39
	Annex II: Adapted FAIR template.....	41

List of Figures

Figure 2-1. Research Data Management according to the FAIR principles	11
Figure 2-2. Granting Open Access Diagram	13
Figure 3-1. ICARUS Next Cloud Repository	17
Figure 3-2. ICARUS Project Web Site.....	18

List of Tables

Table 4-1. Data Asset Attributes Analysis Template	22
Table 4-2. ICARUS Internal Publicly Available Data Assets	23
Table 4-3. ICARUS External Publicly Available Data Assets.....	23
Table 4-4. ICARUS Project Deliverables	28
Table 4-5. Public Deliverables Attributes Analysis Template	29
Table 4-6. Scientific Publications Attributes Analysis Template	31
Table 5-1. ICARUS Data Assets Attributes.....	33
Table 5-2. ICARUS Data Assets with Anonymisation Requirements.....	34
Table 5-3. ICARUS Data Subjects.....	34
Table 5-4. Data Controllers, Data Processors and Data Recipients	35
Table I-1. H2020 FAIR DMP Template	40
Table I-2. H2020 FAIR DMP Template	42

1 Introduction

Each project in the EC's Horizon 2020 program has to define what kind of results are generated or collected during the project's runtime and when and how they are published openly. Giving particular attention to research data management, the consortium follows a series of dedicated activities toward disseminating the project outcomes in a wider audience. These activities should clearly address issues such as the overall methodology for handling the scientific outcomes of the project, the specification of data types that the project generates and/or collects, the standards that will be used, the process of how this data will be exploited and/or shared/made accessible for verification and re-use, the data preservation and maintenance processes etc.

1.1 Purpose of the Document

Deliverable D6.3 is the direct outcome of task T6.4 led by SILO. This document describes initially the methodological framework taking into account the guidelines from EC. In particular, the Data Management Handling Plan is formulated in accordance with the H2020 guidelines regarding Open Research Data, as ICARUS aims to actively take part in the Pilot of Open Research Data. In alignment with the EC Guidelines for Open Access [1], the aim of the current document is to clearly define how the scientific publications issued by the project consortium, white papers published, open source code generated, and mock-up datasets generated will be further disseminated to a wider audience. It aims at outlining the methodology which can make the research data generated in the context of the ICARUS project, findable, accessible, interoperable and reusable (FAIR principles).

The current document is related to the detailed plan regarding which data will be collected and generated during the project, and how it will be shared and opened. This information is essential to decide the best sustainability model for project results and disseminate according to this plan the open data provided by the project. D6.3 aims at providing not only the source of the data, the standards and metadata used for data handling, etc. for all the results generated or collected during the ICARUS project, but also at defining the process of how this data will be further exploited and/or shared/made accessible for verification and re-use, along with the data preservation and maintenance processes.

The current version constitutes the initial version of the Data Management Plan. This initial version aims at defining and documenting the management strategy and methodology, and at answering the aforementioned questions for the data assets that have been preliminary identified and made available, and not at providing detailed answers to all the questions for all data assets that will be made available, produced and published within the lifecycle of the project. Rather, the DMP is intended to be a *living document* in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. This is the 1st version of the document which is going to be updated on a regular basis.

The overall analysis towards the definition of the ICARUS Data Management Plan is provided in the following sections.

1.2 Structure of the Document

The current document comprises of six main sections:

1. Section 1 introduces the document and describes its purpose and structure.
2. Section 2 describes the Data Management Process of ICARUS towards handling the results collected or generated during the project. The overall analysis is inline and in accordance with the EU guidelines, applying the adapted H2020 FAIR DMP Template [2]. It also defines the responsibilities' assignment and the decision-making concerning data that is created and uploaded to the ICARUS platform.
3. Section 3 provides an overview of the platforms exploited within the context of ICARUS to openly publish the project results. It presents the platforms selected to present the datasets/publications during the project and describes their concepts for publishing, storage, and backup.
4. Section 4 documents a list of all existing or foreseeable results for dissemination, separated into public deliverables, publications and open research data, and provides the description, the standards used, the associated metadata, and the open access platform chosen for storage and curation, in accordance to the FAIR data management guideline.
5. Section 5 describes how the ICARUS Data Management Handling Plan and overall procedures and processes comply with GDPR, identifying the data subjects involved and the Personally Identifiable Information exposed, associating the ICARUS roles with the user roles defined by GDPR, and documenting the corresponding processing activities.
6. Section 6 concludes the deliverable.

2 ICARUS Data Management Process

Data Management Plans (DMPs) are a key element of good data management. A DMP describes the data management life cycle for the data to be collected, processed and/or generated by a Horizon 2020 project. As part of making research data findable, accessible, interoperable and re-usable, a DMP should include information about the handling of research data during and after the end of the project [1]:

- what data will be collected, processed and/or generated; what kind of data will the project collect or generate, and to whom might they be useful later on;
- which methodology and standards will be applied; what metadata required to enable data to be found and understood, ideally according to the particular standards of scientific discipline;
- whether data will be shared/made open access;
- how data will be preserved (including after the end of the project); how to archive and preserve the open datasets of the project;

More specifically, for Horizon 2020 projects a FAIR DMP template has been designed to be applicable to any project that produces, collects or processes research data [2]. The respective activities defined as part of the methodology, adopted also in ICARUS project are [3]:

- Data Summary
- FAIR Data Principles
 - Making data findable, including provisions for metadata
 - Making data openly accessible
 - Making data interoperable
 - Increase data re-use (through clarifying licenses as defined during project period)
- Allocation of resources
 - Explain the allocation of resources
- Data Security
 - Address data recovery as well as secure storage and transfer of sensitive data
- Ethical Aspects
 - In the context of the ethics management plan of the project
- Other Issues
 - Refer to other national/funder/sectorial/departmental procedures for data management if any

Figure 2-1 presents the FAIR data principles towards promptly disseminating the data outcomes of a research project.

A DMP is required for all projects participating in the extended Open Research Data pilot. ICARUS wants to be part of the pilot and thus a preliminary version of the Data Management Plan is provided early in the project. The DMP needs to be updated over the course of the project whenever significant changes arise, such as (but not limited to):

- new data are generated
- changes in consortium policies (e.g. new innovation potential, decision to file for a patent)

- changes in consortium composition and external factors (e.g. new consortium members joining or old members leaving).

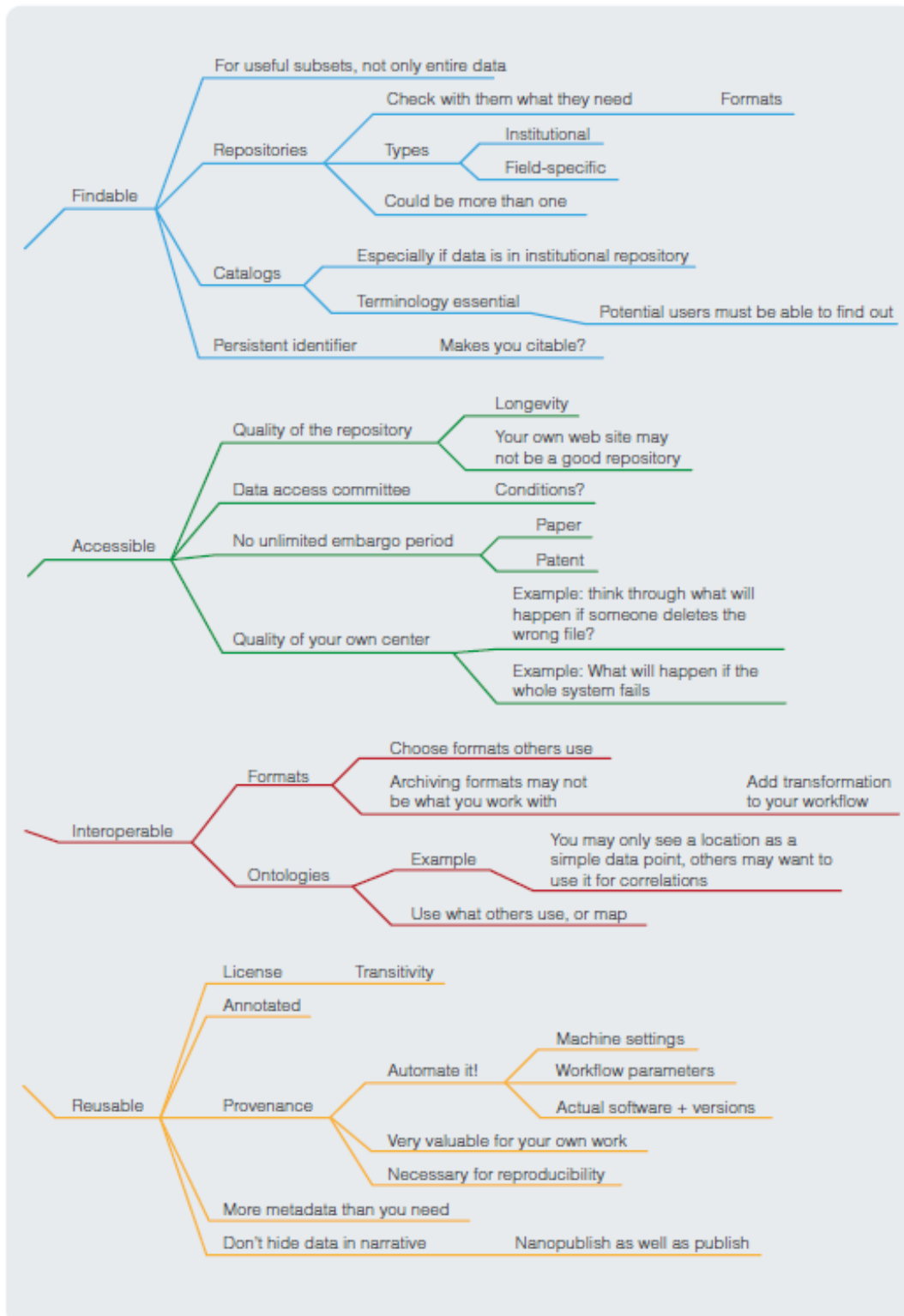


Figure 2-1. Research Data Management according to the FAIR principles

Along with the definition of the datasets, special focus is delivered at the selection of the platform to archive and preserve the datasets. When choosing a repository, it is important to consider factors such as whether the repository [4]:

- Gives the submitted dataset a persistent and unique identifier. This is essential for sustainable citations – both for data and publications – and to make sure that research outputs in disparate repositories can be linked back to particular researchers and grants.
- Provides a landing page for each dataset, with metadata that helps others find it, tell what it is, relate it to publications, and cite it. This makes your research more visible and stimulates reuse of the data.
- Helps to track how the data has been used by providing access and download statistics.
- Responds to community needs and is preferably certified as a ‘trustworthy data repository’, with an explicit ambition to keep the data available in the long term.
- Matches particular data needs (e.g. formats accepted; access, back-up and recovery, and sustainability of the service). Most of this information should be contained within the data repository’s policy pages.
- Provides guidance on how to cite the data that has been deposited.

In addition, a main point of the DMP is the definition of the open access type over the data. Open Access (OA) refers to the practice of providing online access to scientific information that is free of charge to the end-user and reusable. 'Scientific' refers to all academic disciplines. In the context of research and innovation, 'scientific information' can mean:

- peer-reviewed scientific research articles (published in scholarly journals) or
- research data (data underlying publications, curated data and/or raw data).

Open Access to scientific publications means free online access for any user. The two main routes to Open Access are:

1. **Self-archiving / 'green' Open Access** – the author, or a representative, archives (deposits) the published article or the final peer-reviewed manuscript in an online repository before, at the same time as, or after publication. Some publishers request that open access is granted only after an embargo period has elapsed.
2. **Open Access publishing / 'gold' open access** - an article is immediately published in open access mode. In this model, the payment of publication costs is shifted away from subscribing readers. The most common business model is based on one-off payments by authors.

Within the context of the project, the ICARUS consortium opts for the “Green” Open Access model as it is the one commonly until now adopted by most (mainly academic) partners. Nevertheless, this does not exclude the adoption of the “Gold” Open Access model for specific publications in case the authors wish to adopt it and cover the publication costs.

Research data refers to information, in particular facts or numbers, collected to be examined and considered as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images. The focus is on research data that is available in digital form. Users can normally access, mine, exploit, reproduce and disseminate openly accessible research data free of charge. The next figure presents the process flow towards defining the open access type in scientific publications and research data.

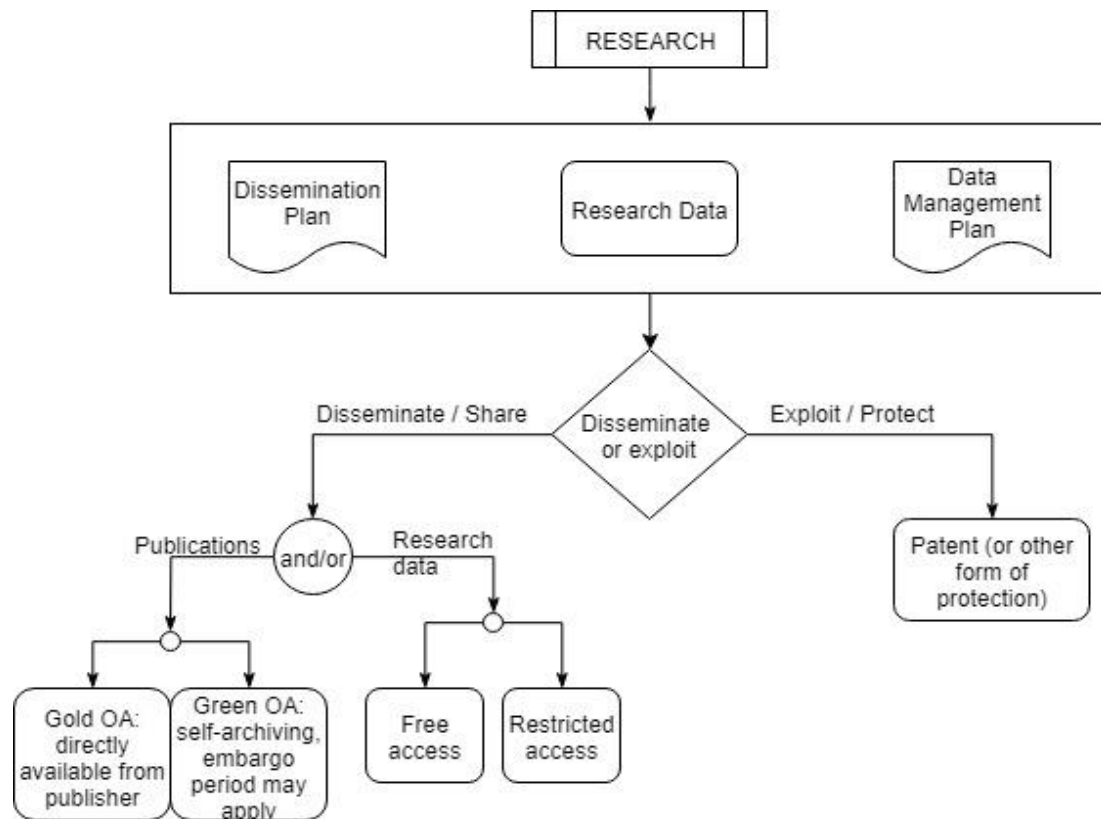


Figure 2-2. Granting Open Access Diagram

The open access mandate comprises two steps:

1. depositing publications in repositories
2. providing open access to them

These steps are explained in the following sections along with the definition of the ways of archiving and preserving the open datasets of the ICARUS project. A methodological process to handle the Data Management Process is provided as part of the framework towards the definition of the Data Management Plan.

In addition to adopting the FAIR principles and to defining the open access type over the data, a thorough DMP needs to also provide information about conformance to the General Data Protection Regulation¹ (GDPR), the newly enacted EU regulation about data. GDPR is a unified regulatory framework that took effect as of the 25th May 2018 and should be implemented across the European Union. Due to its regulatory nature, it is directly binding and applicable for all EU members and does not require an individual enabling state legislation. Intended to replace the 1995 EU Data Protection Directive² in the light of the growing need for clarifying and safeguarding the “digital rights” of all natural persons in EU, GDPR harmonises the relevant national EU States legislations and opens up the scope to cover even companies without a physical presence in the EU.

¹ General Data Protection Regulation, available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN>

² Data Protection Directive, OJ 1995 L 281, available at: <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:31995L0046>

Organisations acquiring and/or processing data of natural persons are required to adopt more robust data management and security systems. At the same time, GDPR empowers citizens, by enhancing monitoring and control over their own data. As stated³ :

- This Regulation lays down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data.
- This Regulation protects fundamental rights and freedoms of natural persons and in particular their right to the protection of personal data.
- The free movement of personal data within the Union shall be neither restricted nor prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data.

For further details on the GDPR and the changes it introduces in data handling and manipulation at research and industrial level, mainly at project level in the context of ICARUS, please consult ICARUS D8.2 – Project Ethics Safeguarding Framework.

As previously noted, significant changes on data, which may arise in the course of the project and the development of the platform, are to be reported in the form of new versions of the present deliverable. Finally, all generated and collected project data will be described in detail at the end of the project within the context of deliverable D6.5 - Final Report on Dissemination, Communication and Stakeholder Engagement to be delivered on M36.

2.1 Data Management Process

Taking into account this preliminary analysis, the ICARUS Data Management Process is defined as a step-wise approach for each result generated or collected during the project runtime. The following questions must be answered to classify the different datasets:

1. Does a result provide significant value to others or is it necessary to understand a scientific conclusion?

If this question is answered with yes, then the result is classified as public (granted for open access). If this question is answered with no, the result is classified as non-public. For example, code that is very specific to the ICARUS platform (e.g. a database initialization) is usually of no scientific interest to anyone, nor does it add any significant contribution.

2. Does a result include personal information that is not the author's name?

If this question is answered with yes, the result is classified as non-public. Personal information beyond the name must be removed if it should be published according to the ethical principles of the project.

3. Does a result allow the identification of individuals even without the name?

This is also a step managed by the ethics management plan of the project as we have committed in ICARUS project to establish anonymization techniques to conceal a single user's identity, e.g.

³ Article 1, GDPR

abstraction, dummy users, or non-intersecting features. If this question is answered with yes, the result is classified as non-public.

4. Can a result be abused for a purpose that is undesired by society in general or contradicts with societal norms and the project's ethics?

If this question is answered with yes, the result is classified as non-public.

5. Does a result include business or trade secrets of one or more partners of the project?

If this question is answered with yes, the result is classified as non-public. Business or trade secrets need to be removed in accordance to all partners' requirements before it can be published.

6. Does a result name technologies that are part of an ongoing, project-related patent application?

If this question is answered with yes, then the result is classified as non-public. Of course, results can be published after the patent has been filed.

7. Does a result break security interests for any project partner?

If this question is answered with yes, the result is classified as non-public.

This is a simple structural approach to determine the different data types defined as part of the DMP. The responsibilities of the ICARUS consortium partners towards disseminating the project outcomes are defined in the following section.

2.2 Responsibilities and Decision Making

The Data Management Plan presented in this deliverable aims to identify the project outputs to be massively disseminated as well as to decide on ways and means of their Open Access (if applicable). To ensure it, a dedicated time slot will be reserved at each of the project plenary meetings and, if needed, at selected consortium audio conferences. EC and project reviewers will be informed about related work done and publications provided in the project management reports.

Individual responsible persons for dealing with, handling and resolving data management related issues within the ICARUS project consortium include:

- Data Management Plan Leader (SILO) – Responsibilities include preparing and leading related discussions at the relevant project meetings and maintaining the channels for dissemination of project outcomes.
- Technical Coordinator (SUITE5) – Responsibilities include defining the specifications and design principles of the ICARUS platform and safeguarding those technical choices taken for all data handling and manipulation activities are according to and do not violate the GDPR regulation and principles.
- Scientific coordinator (UCY) – Responsibilities include defining all theoretical and scientific aspects and mentoring of technical partners on solutions proposed and adopted, as well as identifying data collected by the project and technical project outcomes eventually suitable for publication.
- Dissemination leader (CELLOCK) – Responsibilities include identifying publications suitable for publication in the considered repositories and maintaining the ICARUS inputs for the Open Access.

- Each individual partner – Responsibilities include identifying own project results suitable for publication.

Moreover, each ICARUS partner has to respect the policies set out in this DMP. Datasets have to be created, managed and stored appropriately and in line with applicable legislation. Validation and registration of datasets and metadata is the responsibility of the partner that generates the data in the Work Package (WP). Metadata constitutes an underlying definition or description of the datasets, and facilitate finding and working with particular instances of data. Additional responsibilities undertaken by the ICARUS project partners include:

- Backing up data assets for sharing through Open Access repositories. It is the responsibility of the partner possessing these data assets.
- Quality control of the data assets. It is the responsibility of the relevant WP leader, supported by the Data Management Plan Leader.
- In the case the data assets are updated, the partner that possesses the data has the responsibility to manage the different versions and to make sure that the latest version is available in the case of publicly available data.
- Consulting the concerned partner(s) before publishing data in the open domain that can be associated with an exploitable result. It is the responsibility of all project partners involved in this activity.

3 Data Archiving and Publication Infrastructure

Before providing the detailed analysis of the datasets/publications to be handled within the context of ICARUS DMP, we provide an overview of the platforms to publish our results openly. The following list presents the platforms selected to present the datasets/publications during the project and describes their concepts for publishing, storage, and backup.

3.1 ICARUS Nextcloud Repository

ICARUS Document sharing and collaborative work will be performed via the support of the Nextcloud open source software⁴. Nextcloud is a suite of client-server software for creating and using file hosting services. It is functionally similar to Dropbox, although Nextcloud is free and open-source, allowing anyone to install and operate it on a private server. In contrast to proprietary services like Dropbox, the open architecture allows adding functionality to the server in form of applications and enables users to have full control of their data. Nextcloud files are stored in conventional directory structures. User files are encrypted during transit and can be encrypted at rest (requires encryption to be turned on). Nextcloud can synchronise with local clients running Windows, OS X (10.6 or later), or various Linux distributions. Nextcloud users can manage calendars (CalDAV), contacts (CardDAV), scheduled tasks and streaming media (Ampache) from within the platform. From the administration perspective, Nextcloud permits user and group administration (via OpenID or LDAP). Content can be shared by defining granular read/write permissions between users and/or groups. Alternatively, Nextcloud users can create public URLs when sharing files. Logging of file-related actions, as well as disallowing access based on file access rules is also available. Furthermore, users can interact with the browser-based text editor, bookmarking service, URL shortening suite, gallery, RSS feed reader and document viewer tools from within Nextcloud. For additional extensibility, Nextcloud can be augmented with "one-click" applications and connection to Dropbox, Google Drive and Amazon S3. The ICARUS Nextcloud Repository is mainly intended for internal use, however a set of (properly anonymised) test data will also be available in the repository from the project demonstrator partners including also OAG.

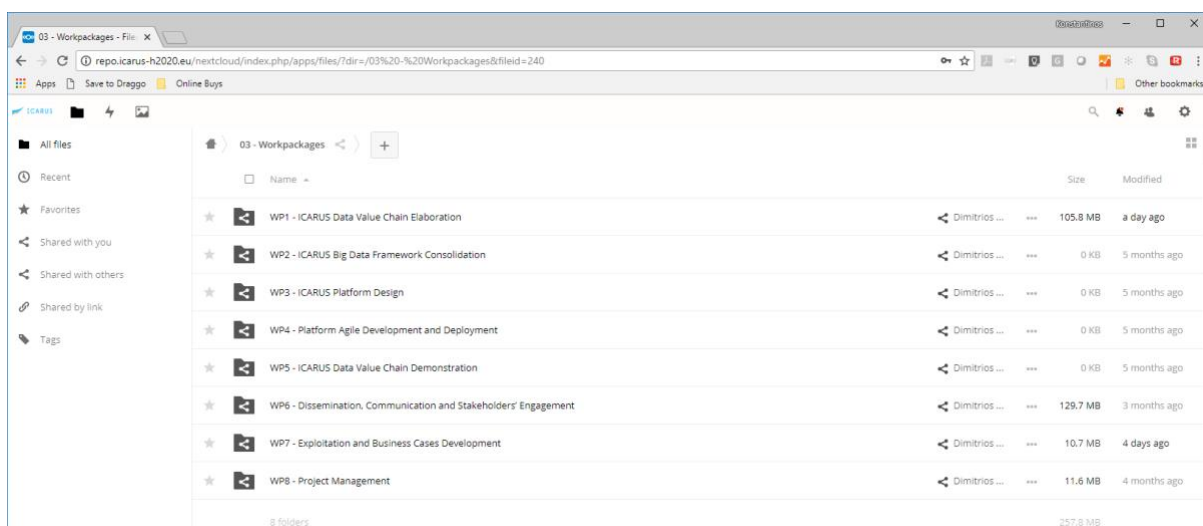


Figure 3-1. ICARUS Next Cloud Repository

⁴ <https://nextcloud.com/>

The ICARUS Nextcloud repository is hosted by the Project Coordinator (UBITECH). ICARUS partners use Nextcloud as the main document repository of all the files exchanged within the consortium, including intermediate versions of the deliverables, meetings’ material (agenda, notes, presentations, demos, minutes, etc.) and any other documents used for gathering inputs from the project’s partners. Credentials are needed to access any of the material, as the platform usage is restricted to the ICARUS consortium and to EC.

The ICARUS Nextcloud repository will be available during the project runtime, and will still be available for at least one year after the official project end.

Web link: <http://repo.icarus-h2020.eu/nextcloud/index.php/login>

3.2 ICARUS Website

The partners in the ICARUS consortium decided early to setup its own project-related webpage, which has been set up since the 1st month of the project. Its purpose is to describe the mission and the general approach of the project and its development status, as well as provide a short description of the project’s objective and its methodology, post news, events and updates that are relevant to the project’s activities. A dedicated page for project’s public documents is available where all the deliverables of the project are published in portable document format (PDF). The webpage was designed by the technical coordinator (Suite5) and was reviewed by the whole ICARUS project consortium. All webpage-related data is backed on a regular basis. All information on the ICARUS website can be accessed without creating an account. The webpage is backed manually every two weeks. The ICARUS Project Portal will be collecting useful (anonymous) data regarding its visitors including unique visitors, countries of origin, time spent on portal etc., and will be available during the project runtime, and for at least two years after the official project end.

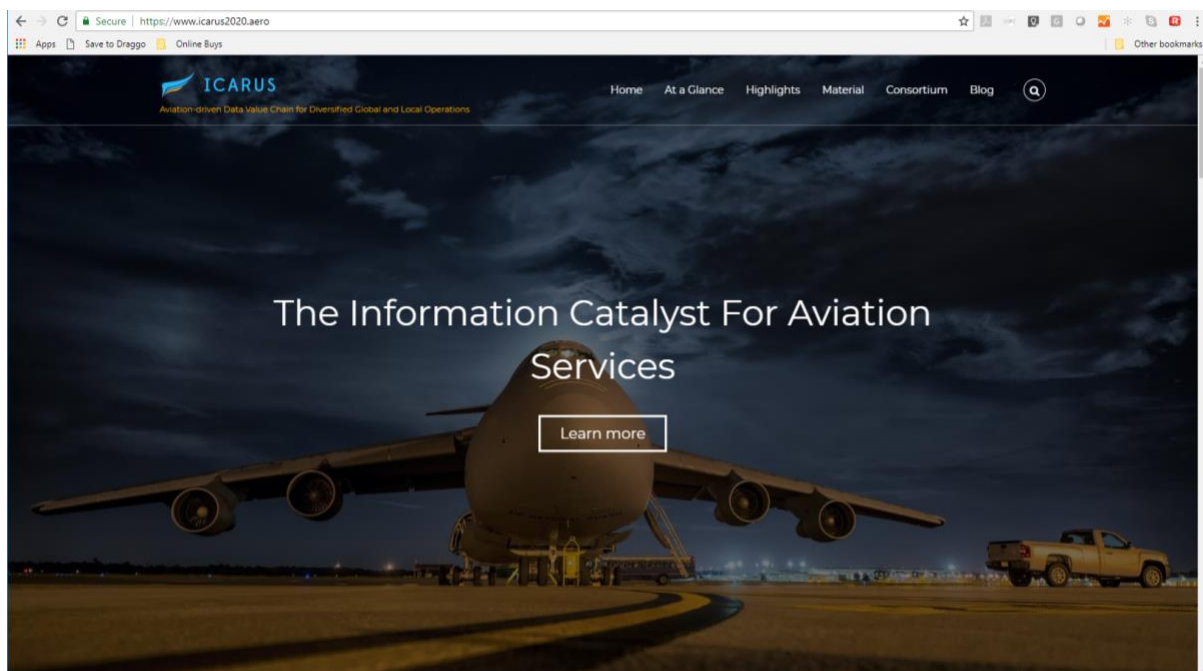


Figure 3-2. ICARUS Project Web Site

Web link: <https://www.icarus2020.aero>

3.3 ResearchGate

Along with the establishment of the project portal, the ICARUS ResearchGate channel has been established to promote the dissemination of scientific publications of the project. Open Access documents are published using the portable document format (PDF). All downloads are enriched by using simple metadata information like the title, a short description and the type of the document. The ICARUS ResearchGate channel is managed by Suite5, while all partners periodically update the material. The link for accessing the ICARUS ResearchGate channel is:

<https://www.researchgate.net/project/ICARUS-Aviation-driven-Data-Value-Chain-for-Diversified-Global-and-Local-Operations>

The aforementioned tools are defined as the platforms for accessing the project scientific results. Towards the wider dissemination of project outcomes and following the recommendation from EC, we intend to use Zenodo service for datasets dissemination, as described in the following section.

3.4 Zenodo Project Repository

Zenodo is an open data repository service maintained by CERN, Geneva and was launched in 2013. Zenodo archives, and makes available, research outputs in all scientific disciplines. Datasets can be located via the Zenodo ElasticSearch engine. Zenodo is compliant with the open data requirements of Horizon 2020, the EU Research and Innovation funding programme and OpenAIRE, the EC-funded initiative in support of the Open Access policies of the European Union. Data are backed nightly and replicated to different locations.

Zenodo not only supports the publication of scientific papers or white papers, but also the publication of any structured research data (e.g. using XML). Zenodo provides a connector to GitHub that supports open collaboration for source code and versioning for all kinds of data. All uploaded results are structured by using metadata, like for example the contributors' names, keywords, date, location, kind of document, license, and others. Considering the language of textual metadata items, English is preferred. All metadata is licensed under CC license (Creative Commons 'No Rights Reserved')⁵. The property rights or ownership of a result does not change by uploading it to Zenodo.

All public results generated or collected during the ICARUS project will be uploaded to Zenodo for long-term storage and open access. At the end of the project, we will consider the option of using an institutional research data repository to further disseminate the project outcomes. The ICARUS project profile has been established in the platform.

Web link: <https://zenodo.org/communities/icarus2020/>

3.5 ICARUS platform

The ICARUS partners will setup a multi-segment platform that will combine data of different velocity, variety and volume under an inter-linked, trusted, multilingual engine to produce a big-data repository of value and veracity. The ICARUS platform will address critical barriers for the adoption of Big Data in

⁵ <https://creativecommons.org/share-your-work/public-domain/cc0/>

the aviation industry (e.g. data fragmentation, data provenance, data licensing and ownership, data veracity), and will enable aviation-related big data scenarios for EU-based companies, organizations and scientists, that will allow exploration, curation, integration and deep analysis of original, synthesized and derivative data characterized by different velocity, variety and volume in a trusted and fair manner. The ICARUS platform will facilitate the execution of four strong pilots that will bring into ICARUS data in the order of TeraBytes in order to deliver a resource of collaborative, data-driven intelligence. ICARUS will give participants the capability to upload both private and public resources of data and interrelate them, facilitating the extraction of insights through powerful analytics and intuitive visualisations. Delivering this platform is the main objective of the ICARUS project and will be realised in the context of WP4 of the project through multiple incremental development releases.

Link: Not available yet. However, SILO will be responsible for setting up and maintaining the ICARUS platform in the infrastructure provided by partner CINECA, which will be available to the consortium in the forthcoming months.

3.6 GitHub

As part of the exploitation process, the members of the consortium will disseminate some of the project outcomes as Open Source products. While Zenodo is the repository for archiving and preserving publications and datasets, it provides also a connector to GitHub that supports open collaboration for source code. Therefore, the aim of the consortium is to preserve a GitHub channel for disseminating the open source outcomes of the ICARUS project.

GitHub is a well-established online repository which supports distributed source code development, management, and revision control. It is primarily used for source code data. It enables worldwide collaboration between developers and provides also some facilities to work on documentation and to track issues.

GitHub provides paid and free service plans. Free service plans can have any number of public, Open Access repositories with unlimited collaborators. Private, non-public repositories require a paid service plan. Many open-source projects use GitHub to share their results for free. The platform uses metadata like contributors' nicknames, keywords, time, and data file types to structure the projects and their results. The terms of service state that no intellectual property rights are claimed by the GitHub Inc. over the provided material. For textual metadata items, English is preferred. The service is hosted by GitHub Inc. in the United States. GitHub using a rented Rackspace⁶ hardware infrastructure where data is backed continuously to different locations.

All source-code components that are implemented during the project and decided to be public will be uploaded to an open access GitHub repository, linked also with the Zenodo account of the project.

Web link: <https://github.com/>

The dedicated link to the ICARUS open source code will be available in due time, as the project evolves and technical tasks initialise.

⁶ <https://www.rackspace.com/>

We have presented above the list of platforms to be considered for disseminating the project outcomes. We proceed in the next section with the early definition of the datasets/ documents that will be disseminated as part of the Data Management Handling Plan.

4 ICARUS Datasets and Publications

In this section, a list of all existing or foreseeable results for dissemination is presented, separated into public deliverables, publications and open research data. For each result and in accordance to the FAIR data management guideline [2] we provide a description, name the standards used for storage and metadata (to make data findable & interoperable), and define which open access platform is chosen. Data Security aspects are also defined in this document, while the detailed ethics management policy of ICARUS project is defined in WP8 (D8.2 – Project Ethics Safeguarding Framework in specific) which provides a detailed analysis on the way to handle the datasets generated in the project. In summary, the ICARUS partners will comply with the ethical principles as set out in Article 34 of the Grant Agreement, which asserts that all project activities must be carried out in compliance with:

- (a) ethical principles (including the highest standards of research integrity - as set out, for instance, in the European Code of Conduct for Research Integrity - and including, in particular, avoiding fabrication, falsification, plagiarism or other research misconduct)
- (b) Applicable international, EU and national law.

4.1 ICARUS project datasets

This section undertakes the provision of the overview of the public datasets that will be made available by the consortium partners for analysis. The template for the management of ICARUS public datasets is provided below:

Dataset reference and name: The name and the ID of the dataset.

Dataset description: A short description of the content of the dataset.

Standards and metadata: The type of the dataset format and any type of metadata associated with the content of the dataset.

Dataset sharing: Where the final version of the dataset is stored.

Archiving and preservation (including storage and backup): How the dataset will be preserved, how it will be archived and preserved.

Table 4-1. Data Asset Attributes Analysis Template

Within the context of ICARUS, datasets can be divided into three main categories, based upon their privacy attribute. These three categories are Public, Proprietary and Confidential. Confidential data are not to be shared, while proprietary data will be made available outside the platform using specific licenses to be defined during the course of the project. Towards this end, in the current document we will only be dealing with the publicly available data, the list of which (as per the holistic list of datasets included in D1.1) is provided below:

ID	Data Asset Title	Data Asset Owned	Privacy	Need for anonymisation
ISI_01	Population data	N	Public	N
ISI_02	GLEAM Simulation output	Y	Public	N

ISI_03	Virus & infections data	N	Public	N
CELLOCK_02	Number of Passengers	Y	Public	N

Table 4-2. ICARUS Internal Publicly Available Data Assets

In addition to these datasets, the consortium has also identified external data sources that are related to aviation. These data sources are mostly collected and released by EUROCONTROL⁷ and range from the European AIS Database and the Network Operations Portal to broader aviation statistics published by Eurostat. The following table includes these datasets with their corresponding attributes.

ID	Data Asset Title	Data Asset Owned	Privacy	Need for anonymisation
ADS_01	Eurostat Air Transport Statistics	N	Public	N
ADS_03	Eurocontrol Network Operations Portal (NOP)	N	Public	N
ADS_04	ICAO Engine Emissions Databank	N	Public	N
ADS_05	European Aviation Environmental Report	N	Public	N

Table 4-3. ICARUS External Publicly Available Data Assets

For each of these datasets, we populate the template provided before.

4.1.1 Population Data

Dataset reference and name: ISI_01, Population data

Dataset description: The Gridded Population of the World (GPW) collection, now in its fourth version (GPWv4), models the distribution of human population (counts and densities) on a continuous global raster surface

Standards and metadata: The data sets are available in ASCII (text) format as well as in GeoTiff format. Five of the eight raster data sets are also available in netCDF format. In addition, the native 30 arc-second resolution data were aggregated to four lower resolutions (2.5 arc-minute, 15 arc-minute, 30 arc-minute, and 1 degree) to enable faster global processing and support of research communities that conduct analyses at these resolutions. All of these resolutions are available in ASCII and GeoTiff format. NetCDF files are available at all resolutions except 30 arc-second. All spatial data sets in the GPWv4 collection are stored in geographic coordinate system (latitude/longitude).

Dataset sharing: The dataset is already anonymised, thus there is no need for additional anonymization activities. The GPW data collection is licensed under the Creative Commons Attribution 4.0 International License.

⁷ <https://www.eurocontrol.int/>

Link: <http://sedac.ciesin.columbia.edu/data/collection/gpw-v4>

Archiving and preservation (including storage and backup): The data will be archived and preserved in the ICARUS platform. After the project's completion, the commercialisation scheme that the project will select will be responsible to maintain and archive the data used and created in the ICARUS platform.

4.1.2 GLEAM Simulation output

Dataset reference and name: ISI_02, GLEAM Simulation output

Dataset description: Output data concerning the time evolution of GLEAM (Global Epidemic and Mobility model) simulations: day by day number of individuals in each disease's compartment per census area, as long as new transitions between compartments and their cumulative number.

Standards and metadata: The data are semi-structured and are made available in HDF5 format.

Dataset sharing: The dataset is already anonymised, thus there is no need for additional anonymization activities. The data collections will be made available as downloadable files that are partly public and partly confidential, and will be made available free of charge on a case-by-case bilateral agreement.

Archiving and preservation (including storage and backup): The data will be archived and preserved in the ICARUS platform. After the project's completion, the commercialisation scheme that the project will select will be responsible to maintain and archive the data used and created in the ICARUS platform.

4.1.3 Virus & infections data

Dataset reference and name: ISI_03, Virus & infections data

Dataset description: Data describing the number of influenza viruses detected, total number of influenza positive/negative viruses, etc., by different countries and influenza transmission zones

Standards and metadata: The reports generated according to the user criteria can be exported in various standardised formats including XLS, CSV, XML and MHTML, in image formats (TIFF) as well as in document formats including PDF and word files.

Dataset sharing: The dataset is already anonymised, thus there is no need for additional anonymization activities. The data collection is licensed under the Creative Commons Attribution 4.0 International License.

Link: <http://apps.who.int/flumart/Default?ReportNo=12>

Archiving and preservation (including storage and backup): The data will be archived and preserved in the ICARUS platform. After the project's completion, the commercialisation scheme that the project will select will be responsible to maintain and archive the data used and created in the ICARUS platform.

4.1.4 Number of Passengers

Dataset reference and name: CELLOCK_02, Number of Passengers

Dataset description: The data asset includes the number of passengers in a flight collected/updated after a flight is completed.

Standards and metadata: The dataset can be exported in JSON format through APIs.

Dataset sharing: The dataset is already anonymised, thus there is no need for additional anonymization activities. The sharing / copyright license for the specific data collection has yet to be defined.

Archiving and preservation (including storage and backup): The data will be archived and preserved in the ICARUS platform. After the project's completion, the commercialisation scheme that the project will select will be responsible to maintain and archive the data used and created in the ICARUS platform.

4.1.5 Eurostat Air Transport Statistics

Dataset reference and name: ADS_01, Eurostat Air Transport Statistics

Dataset description: A collection of Air Transport Statistics is based on Regulation (EC) No 437/2003⁸ of the European Parliament and of the Council of 27 February 2003, on statistical returns in respect of the carriage of passengers, freight and mail by air as well as the subsequent implementing Commission Regulations 1358/2003⁹, 546/2005¹⁰ and 158/2007¹¹. Data are supplied by all Member States + EFTA countries (NO, CH and IS). Some CC countries are also participating in this data collection (TR, MK, MN and SB).

Standards and metadata: The datasets generated according to the user criteria can be exported in various standardised formats including XLS, CSV, SPSS, TSV as well as in PDF format. Metadata for these data assets are included within the assets.

Dataset sharing: The dataset is already anonymised, thus there is no need for additional anonymization activities. In line with the Community legal framework and the European Statistics Code of Practice, Eurostat disseminates European statistics on Eurostat's website respecting professional independence and in an objective, professional and transparent manner in which all users are treated equitably. The detailed arrangements are governed by the Eurostat protocol on impartial access to Eurostat data for users.

Link: <http://ec.europa.eu/eurostat/web/transport/data/database>

Archiving and preservation (including storage and backup): The data will be archived and preserved in the ICARUS platform. After the project's completion, the commercialisation scheme that the project will select will be responsible to maintain and archive the data used and created in the ICARUS platform.

4.1.6 Eurocontrol Network Operations Portal (NOP)

Dataset reference and name: ADS_03, Eurocontrol Network Operations Portal (NOP)

⁸ <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:066:0001:0008:EN:PDF>

⁹ <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2003:194:0009:0033:EN:PDF>

¹⁰ <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32005R0546&from=EN>

¹¹ <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32007R0158&from=en>

Dataset description: A centralized portal to monitor the real-time status of traffic, airspace and air traffic flow and capacity management (ATFCM) measures and to collaboratively plan pan-European operations from the strategic to the tactical phases, thus optimizing the use of available ATM capacity.

Standards and metadata: The datasets, reports and charts generated can be exported in various standardised formats including XLS, CSV, XML, in image formats (TIFF) as well as in document formats including PDF files. Metadata for these data assets are included within the assets, depending on the data asset format.

Dataset sharing: The dataset is already anonymised, thus there is no need for additional anonymization activities.

Link: <https://www.public.nm.eurocontrol.int/PUBPORTAL/gateway/spec/>

NOP public portal is accessible through Quicklinks. For accessing the restricted NOP Portal the user should visit the Network operations page through the central button. Links to Protected Applications (including NOP, Network Management Interactive Reporting (NMIR) or Central Claim Management System (CCMS) Web) are available yet require access tokens.

Archiving and preservation (including storage and backup): The data will be archived and preserved in the ICARUS platform. After the project's completion, the commercialisation scheme that the project will select will be responsible to maintain and archive the data used and created in the ICARUS platform.

4.1.7 ICAO Engine Emissions Databank

Dataset reference and name: ADS_04, ICAO Engine Emissions Databank

Dataset description: The ICAO Aircraft Engine Emissions Databank contains information on exhaust emissions of production aircraft engines, measured according to the procedures in ICAO Annex 16, Volume II¹², and where noted, certified by the States of Design of the engines according to their national regulations. The databank covers engine types which emissions are regulated, namely turbojet and turbofan engines with a static thrust greater than 26.7 kN. The information is provided by the engine manufacturers, who are solely responsible for its accuracy.

Standards and metadata: The datasets generated according to the user criteria can be exported in XLS format. Metadata for these data assets are included within the assets.

Dataset sharing: The dataset is already anonymised, thus there is no need for additional anonymization activities. The European Aviation Safety Agency (EASA) is hosting the databank on behalf of ICAO and is not responsible for the contents.

Link: <https://www.easa.europa.eu/easa-and-you/environment/icao-aircraft-engine-emissions-databank>

Archiving and preservation (including storage and backup): The data will be archived and preserved in the ICARUS platform. After the project's completion, the commercialisation scheme that the project

¹² <https://law.resource.org/pub/us/cfr/ibr/004/icao.annex.16.v2.2008.pdf>

will select will be responsible to maintain and archive the data used and created in the ICARUS platform.

4.1.8 European Aviation Environmental Report

Dataset reference and name: ADS_05, European Aviation Environmental Report

Dataset description: Information and data collected by EASA, EEA and EUROCONTROL to evaluate the environmental performance of the European aviation sector.

Standards and metadata: The datasets can be exported in XLS format. Metadata for these data assets are included within the assets.

Dataset sharing: The dataset is already anonymised, thus there is no need for additional anonymization activities. The European Aviation Safety Agency (EASA) is hosting the databank.

Link: <https://www.easa.europa.eu/eaer/downloads>

Archiving and preservation (including storage and backup): The data will be archived and preserved in the ICARUS platform. After the project’s completion, the commercialisation scheme that the project will select will be responsible to maintain and archive the data used and created in the ICARUS platform.

4.2 ICARUS project deliverables

We are considering the ICARUS Project public deliverables as part of the data management plan. The following table presents the list of public deliverables of the ICARUS project.

Deliverable	Deliverable name	Short name of lead participant	Type	Dissemination level	Delivery date
D1.1	Domain Landscape Review and Data Value Chain Definition	UCY	R	PU	M3
D1.2	The ICARUS Methodology and MVP	UCY	R	PU	M5
D1.3	Updated ICARUS Methodology and MVP	UCY	R	PU	M15
D2.1	Data Management and Value Enrichment Methods	SUITE5	R	PU	M7
D2.2	Intuitive Analytics Algorithms and Data Policy Framework	SUITE5	R	PU	M9
D2.3	Updated ICARUS Data Management, Analytics and Data Policy Methods	SUITE5	R	PU	M18
D3.1	ICARUS Architecture, APIs Specifications and Technical and User Requirements	UBITECH	OTHER	PU	M9
D3.2	Core Data Service Bundles and Value-Added Services Designs	UBITECH	OTHER	PU	M12

D3.3	Architecture, Core Data and Value-Added Services Bundles Specifications-v2.00	UBITECH	OTHER	PU	M18
D3.4	Architecture, Core Data and Value-Added Services Bundles Specifications-v3.00	UBITECH	OTHER	PU	M26
D3.5	Architecture, Core Data and Value-Added Services Bundles Specifications-v4.00	UBITECH	OTHER	PU	M32
D4.2	ICARUS Platform - Beta Version	SILO	DEM	PU	M18
D4.3	ICARUS Platform - Release 1.00	SILO	DEM	PU	M24
D4.4	ICARUS Platform - Release 1.50	SILO	DEM	PU	M30
D4.5	ICARUS Platform - Release 2.00	SILO	DEM	PU	M36
D5.1	ICARUS Demonstrators and Platform Evaluation Framework	CINECA	R	PU	M12
D5.2	Demonstrators Execution Scenarios and Readiness Documentation	SILO	R	PU	M14
D5.6	Final Evaluation, Impact Assessment and Adoption	PACE	R	PU	M36
D6.1	Plan for Dissemination, Communication and Stakeholder Engagement	CELLOCK	R	PU	M3
D6.2	Project Website and Web 2.0 Channels Setup	SUITE5	DEC	PU	M3
D6.3	Project Data Management Plan	SILO	ORDP	PU	M6
D6.4	Report on Dissemination and Communication for Period 1 and Updated Plan for Period 2	CELLOCK	R	PU	M18
D6.5	Final Report on Dissemination, Communication and Stakeholder Engagement	CELLOCK	R	PU	M36

Table 4-4. ICARUS Project Deliverables

The forthcoming sections provide an overview of the public documents prepared so far by the consortium. The template for the management of ICARUS public deliverables is provided below:

Deliverable reference and name: The name and the ID of the deliverable.

Deliverable description: A short description of the content of the deliverable.

Standards and metadata: The type of the document format and any type of metadata associated with the content of the document.

Deliverable sharing: Where the final version of the deliverable is stored.

Archiving and preservation (including storage and backup): How the deliverable will be preserved, how it will be archived and preserved.

Table 4-5. Public Deliverables Attributes Analysis Template

4.2.1 D1.1 Domain Landscape Review and Data Value Chain Definition

Data set reference and name

Domain Landscape Review and Data Value Chain Definition (Report D1.1)

Data set description

The document provides the state-of-play analysis of the big, linked and open data landscape, suggesting in each category the most promising frameworks and tools to be considered. It describes the process followed to identify stakeholders that will be potentially interested in the ICARUS outcome and can also benefit from the ICARUS data value chain. It also extracts an early indication of the needs of the prospective ICARUS users through the conduction of questionnaires with the demonstrators and other stakeholders, as well as with the analysis of the data and other information sources. D1.1 presents a large set of information/data from different sources that would feed the ICARUS data value chain and that will be used in the platform. Last but not least, it documents the regulatory data sharing framework for data protection, IPR and data sharing that has been defined in order to be used by the brokerage engine of the platform.

Standards and metadata

The document is stored in the cross-platform portable document format (PDF). Metadata is added manually and includes the title, the partner organizations, and keywords that classify this report (Big Data, Open Data, Data Collection, Data Processing, Data Analytics, Data Sources, Data Value Chain, Semantic Web, Entity Recognition Tools, Stakeholder analysis,)

Data sharing

The document will be published openly on the ICARUS webpage (following EC review and approval).

Archiving and preservation

All earlier versions of the document are archived on the project-internal ICARUS NextCloud repository. The document will be added to Zenodo for long-term preservation at the end of the project.

4.2.2 D1.2 The ICARUS Methodology and MVP

Data set reference and name

The ICARUS Methodology and MVP (Report D1.2)

Data set description

D1.2, entitled “The ICARUS Methodology and MVP”, comprises the outcome of Task 1.4 “ICARUS Methodology and MVP Definition” and its main aim is to formulate the ICARUS methodology and the Minimum Viable Product (MVP) in order to reveal how concepts interrelate and to display high-level usage scenarios of the concept. D1.2 presents the initially formulated ICARUS methodology, which

consists of different phases, which in turn comprise of discrete steps. D1.2 further describes the key challenges and the relevant methods for each step. Furthermore, D1.2 presents the defined high-level usage scenarios of ICARUS, which are based on the ICARUS methodology, the input from the ICARUS pilots and the insights provided in D1.1 “Domain Landscape Review and Data Value Chain Definition”. In addition, it presents the extracted features from the defined methodology and high-level scenarios, the initial high-level grouping of the identified implied features, the prioritization of the features and finally the formulated ICARUS MVP.

Standards and metadata

The document is stored in the cross-platform portable document format (PDF). Metadata is added manually and includes the title, the partner organizations, and keywords that classify this report (ICARUS Methodology, Minimum Viable Product, MVP)

Data sharing

The document will be published openly on the ICARUS webpage (following EC review and approval).

Archiving and preservation

All earlier versions of the document are archived on the project-internal ICARUS NextCloud repository. The document will be added to Zenodo for long-term preservation at the end of the project.

4.2.3 D6.1 Plan for Dissemination, Communication and Stakeholder Engagement

Data set reference and name

Plan for Dissemination, Communication and Stakeholder Engagement (Report D6.1)

Data set description

This document, entitled “Plan for Dissemination, Communication and Stakeholder Engagement”, provides the detailed and well-targeted dissemination plan that will uptake all possible and appropriate dissemination channels and means, also exploiting the collaboration and open deliberation capabilities provided by contemporary social networking. The scope of this document is to report the work performed mainly in the context of T6.1- Dissemination and Communication Planning.

Standards and metadata

The document is stored in the cross-platform portable document format (PDF). Metadata is added manually and includes the title, the partner organizations, and keywords that classify this report (Dissemination, Communication, Stakeholders Engagement)

Data sharing

The document will be published openly on the ICARUS webpage (following EC review and approval).

Archiving and preservation

All earlier versions of the document are archived on the project-internal ICARUS NextCloud repository. The document will be added to Zenodo for long-term preservation at the end of the project.

4.2.4 D6.2 Project Website and Web 2.0 Channels Setup

Data set reference and name

Project Website and Web 2.0 Channels Setup (Report D6.2)

Data set description

The purpose of this report is to briefly document the ICARUS website design and deployment, as well as the creation of ICARUS accounts in popular social media. The design and implementation of the ICARUS website fall under the activities of WP6 “Dissemination and Communication” and in particular of Task T6.3 “Communication Activities”. According to the DoA, the objectives of WP6 are (among others) to establish an effective online presence and to communicate the project outcomes to the intended audiences, in a way that is consistent with the project’s branding and scope.

Standards and metadata

The document is stored in the cross-platform portable document format (PDF). Metadata is added manually and includes the title, the partner organizations, and keywords that classify this report (ICARUS Website, Social Media Channels)

Data sharing

The document will be published openly on the ICARUS webpage (following E.C. review and approval).

Archiving and preservation

All earlier versions of the document are archived on the project-internal ICARUS NextCloud repository. The document will be added to Zenodo for long-term preservation at the end of the project.

4.3 ICARUS scientific publications

By the time of writing the deliverable, no scientific publications had been produced in the context of the ICARUS project. However, in the forthcoming versions of this living document, the following template will be used for the management of ICARUS scientific publications:

<p>Publication reference and name: The name and the ID of the publication (e.g. DOI).</p> <p>Publication Abstract: A short description of the content of the publication.</p> <p>Standards and metadata: The type of the document format and any type of metadata associated with the content of the document.</p> <p>Publication sharing: Where the final version of the scientific publication is stored.</p> <p>Archiving and preservation (including storage and backup): How the scientific publication will be preserved, archived and preserved.</p>
--

Table 4-6. Scientific Publications Attributes Analysis Template

5 Compliance with GDPR

The current section undertakes the initial approach towards safeguarding compliance with GDPR principles. Towards doing so, we identify initially the data subjects for which information will be collected, the personally identifiable information per data subject, the roles of the stakeholders associated with data handling activities, and the processing activities associated with the personally identifiable information per data subject and per stakeholder role. Again, we need to stretch that the current document comprises the first version of the ICARUS Data Management Plan, and thus many issues are still open and will be resolved as the project progresses. The updates will be documented in the future versions of the current document.

5.1 Data Subjects

As per the General Data Protection Regulation, a Data Subject is a physical person or a group of physical persons for which information is gathered. Towards identifying the data subjects, we need to examine the data sources which have been analytically documented in ICARUS D1.1

As per D1.1, the datasets that will be made available by the ICARUS project partners (whether proprietary or open data) and will be stored in the ICARUS platform include:

ID	Title	Owner	Privacy	Anonymisation Need
AIA_01	Time stamps and status of ground handling processes	AIA	Proprietary	N
AIA_02	Checked passengers per flight	AIA (Airlines)	Proprietary	N
AIA_03	Expected passengers per flight	Airlines	Proprietary	N
AIA_04	Connecting passengers per flight	Airlines	Proprietary	N
AIA_05	Passengers who need assistance per flight	AIA	Proprietary	Y
AIA_06	Gate open time	AIA	Proprietary	N
PACE_01	Alternative routes comparison	PACE	Proprietary	N
PACE_02	AC performance data	PACE	Proprietary	N
ISI_01	Population data	Columbia University	Public	N
ISI_02	GLEAM Simulation output	ISI	Public	N
ISI_03	Virus & infections data	WHO	Public	N
CELLOCK_01	Retail and F&B in-flight sales	Caterer	Proprietary	N
CELLOCK_02	Number of Passengers	Caterer	Public	N

ID	Title	Owner	Privacy	Anonymisation Need
CELLOCK_03	In-flight, In-Flight Entertainment Passenger data	Airline	Confidential	Y
CELLOCK_04	IFE Content data	Airline	Confidential	Y
OAG_01	Schedules	OAG (Airlines)	Proprietary	N
OAG_02	Flight Status	OAG (Airlines)	Proprietary	N
OAG_03	Flight Tracking	OAG (Airlines)	Proprietary	N
OAG_04	Carrier File	OAG (Airlines)	Proprietary	N
OAG_05	Locations File	OAG (Airlines)	Proprietary	N
OAG_06	MCT (Minimum Connection Times)	OAG (Airlines)	Proprietary	N
OAG_07	DST (Daylight Saving Time)	OAG	Proprietary	N
OAG_08	Country File	OAG	Proprietary	N
OAG_09	Connections	OAG	Proprietary	N
OAG_10	OTP (On Time Performance)	OAG	Proprietary	N

Table 5-1. ICARUS Data Assets Attributes

There are also additional public / open data assets that will be included in the platform, as well as needed data assets which the consortium will try to obtain during the course of the project, but which do not affect the compliance of the ICARUS project with GDPR since they will be anonymised prior to being stored in the ICARUS platform.

Out of these data assets, data that contain potentially personal identifiable information and are in need of anonymization include the following:

ID	Title	Owner	Privacy	Anonymisation Need
AIA_05	Passengers who need assistance per flight	AIA	Proprietary	Y
CELLOCK_03	In-flight, IFE Passenger data	Airline	Confidential	Y
CELLOCK_04	IFE Content data	Airline	Confidential	Y

Table 5-2. ICARUS Data Assets with Anonymisation Requirements

Thus, we can identify two main Data subjects in the context of ICARUS:

ID	Description
DS_01	Passengers who need assistance per flight
DS_02	Passengers (in general)

Table 5-3. ICARUS Data Subjects

5.2 Personally Identifiable Information

Based upon the data subjects identified in the previous section, we now need to identify the personally identifiable information per data subject.

However, even though the exact data schemas per data asset has not been made available by the time of writing, it has been agreed with the owners of the specific data assets (AIA for AIA_05 and CELLOCK for CELLOCK_03 and CELLOCK_04) that user-related data that may lead to single user identification will not be made available to the consortium, and will not be stored on the ICARUS platform. On the contrary, these data assets, (which are either proprietary or confidential and not public), will be anonymised by the data owners prior to being stored on the ICARUS platform.

5.3 Roles

The current section undertakes the analysis of the legal entities involved in the collection, access and processing of the Personally Identifiable Information included in the previous section, according to Article 4 of the EU GDPR, along with their mapping to the ICARUS project partners and external stakeholders.

Thus, according to Article 4 of the EU GDPR, we can identify three main legal entities: 1) the Data Controllers, 2) the Data Processors and 3) the Data Recipients.

Data Controller

Controller – “means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data”.

In the context of the ICARUS project, the Data Controllers are actually the owners of the Data Assets that will be made available to the consortium, either in a public, in a proprietary or a confidential manner. They provide the information, either generated or collected internally, or collected and/or compiled from external stakeholders (e.g. airline companies) that will be managed and analysed in the context of the project. More specifically these organisations are: PACE Aerospace Engineering and Information Technology GmbH (PACE), OAG aviation Worldwide LTD (OAG), ISTITUTO PER L'INTERSCAMBIO ISI SCIENTIFICO (ISI), CELLOCK LTD (CELLOCK) and ATHENS INTERNATIONAL AIRPORT S.A. (AIA).

Data Processor

Processor – “means a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller”.

In the context of the ICARUS project, the Data Processors are the technical partners who are developing the ICARUS Platform, who provide the technical infrastructure and the storage facilities for the data handling and analysis, and for the data storage activities respectively, and who design and implement the data access, security and privacy policies. These partners are UBITECH (UBITECH), ENGINEERING - INGEGNERIA INFORMATICA SPA (ENG), SUITE5 DATA INTELLIGENCE SOLUTIONS LIMITED (SUITE5), UNIVERSITY OF CYPRUS (UCY), CINECA CONSORZIO INTERUNIVERSITARIO (CINECA) and SingularLOGIC S.A. (SILO).

Data Recipient

Recipient – “means a natural or legal person, public authority, agency or another body, to which the personal data are disclosed, whether a third party or not”.

In the context of the ICARUS project, the Data Recipients are the (external) stakeholders to which the raw data assets along with the outcomes of the data analysis are transferred by the Data Controller or are made available by the Data Processors. Such stakeholders include airports, airlines, original equipment manufacturers, airport services providers, aviation-related service providers, health service providers, tourism service providers, security industry, public organizations (e.g. local municipalities), passengers.

The following table outlines the aforementioned legal entities:

ID	Description
Data Controllers	PACE, OAG, ISI, CELLOCK, AIA
Data Processors	UBITECH, ENG, SUITE5, UCY, CINECA, SILO
Data Recipients	Airports, airlines, original equipment manufacturers, airport services providers, aviation-related service providers, health service providers, tourism service providers, security industry, public organizations (e.g. local municipalities), passengers

Table 5-4. Data Controllers, Data Processors and Data Recipients

5.4 Processing Activities

As described analytically in ICARUS D1.2, the ICARUS methodology comprises of various phases, which in turn comprise of a set of steps that relate to data assets handling activities. These include:

- Phase I - Data Collection: Step I.1: Data Retrieval, Step I.2: Data Anonymization, Step I.3: Data Quality Check, Step I.4: Data Curation, Step I.5: Data Check-In
- Phase II - Data Enrichment: Step II.1: Semantic Enrichment and Annotation, Step II.2: Data Linking
- Phase III - Asset Storage
- Phase IV - Asset Exploration and Extraction: Step IV.1: Asset Indexing and Searching, Step IV.2: Asset Export
- Phase V - Data Analytics: Step V.1: Data Analysis, Step V.2: Data Visualization
- Phase VI - Added Value Services: Asset Sharing

Along with some additional steps and phases that are not directly related to data manipulation, rather than with the exploitation of the data analytics outcomes so as to provide additional added value services on top of them.

These processing activities will be applied to the complete set of data assets analysed in deliverable D1.1.

However, as explained in section 5.1 of the current document, out of the complete set of data assets, data that contain potentially personal identifiable information and are in need of anonymization include AIA_05 (Passengers who need assistance per flight), CELLOCK_03 (In-flight, IFE Passenger data) and CELLOCK_04 (IFE Content data).

On top of that, as explained in section 5.2 of the current document, it has been agreed with the owners of the specific data assets (AIA for AIA_05 and CELLOCK for CELLOCK_03 and CELLOCK_04) that user-related data that may lead to single user identification will not be made available to the consortium, and will not be stored on the ICARUS platform. On the contrary, these data assets, (which are either proprietary or confidential and not public), will be anonymised by the data owners prior to being stored on the ICARUS platform.

Thus, the aforementioned processing activities will not be applied in the context of ICARUS on Personally Identifiable Information, and thus these activities are totally aligned with the GDPR regulations.

However, since the current document constitutes a living document, should additional datasets be identified that contain PII and should this information be stored in ICARUS platform, the corresponding measures will be taken to safeguard compliance of the processing activities on the specific data assets with the GDPR regulations.

6 Conclusions

The current deliverable documents the preliminary approach on the definition of the ICARUS Data Management Plan. The main principles and regulations which the DMP is aligned with, as well as the Methodology deriving thereof, are described. The chosen storage solutions and datasets expected to be acquired in the project's lifecycle are also included in this document. As stated above, this document is intended to be a living document; the here presented version of the DMP is not the final and during the project may be subject to minor or major amendments and additions, if need be.

References and Applicable Documents

- [1]. Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020, http://www.gsrt.gr/EOX/files/h2020-hi-oa-data-mgt_en.pdf
- [2]. Guidelines on FAIR Data Management in Horizon 2020, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf
- [3]. How to create a DMP Plan, <https://www.openaire.eu/opendatapilot-dmp>
- [4]. How to select a repository?, <https://www.openaire.eu/opendatapilot-repository>

Annex I: FAIR template

The original FAIR Template is presented in the following table. Details on the content of the table can be found in the Guidelines on FAIR Data Management in Horizon 2020¹³.

DMP component	Issues to be addressed
1. Data summary	<p>State the purpose of the data collection/generation.</p> <p>Explain the relation to the objectives of the project.</p> <p>Specify the types and formats of data generated/collected.</p> <p>Specify if existing data is being re-used (if any).</p> <p>Specify the origin of the data.</p> <p>State the expected size of the data (if known).</p> <p>Outline the data utility: to whom will it be useful.</p>
2. FAIR Data	
2.1. Making data findable, including provisions for metadata	<p>Outline the discoverability of data (metadata provision).</p> <p>Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as DiGital Object Identifiers, URL etc.</p> <p>Outline naming conventions used.</p> <p>Outline the approach towards search keyword.</p> <p>Outline the approach for clear versioning.</p> <p>Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how.</p>
2.2 Making data openly Accessible	<p>Specify which data will be made openly available. If some data is kept closed provide rationale for doing so.</p> <p>Specify how the data will be made available.</p> <p>Specify what methods or software tools are needed to Access the data. Is documentation about the software needed to Access the data included? Is it possible to include the relevant software (e.g. in open source code)?</p> <p>Specify where the data and associated metadata, documentation and code are deposited.</p>

¹³ http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

	Specify how Access will be provided in case there are any restrictions.
2.3. Making data interoperable	<p>Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.</p> <p>Specify whether you will be using standard vocabulary for all data types present in your data set, to allow interdisciplinary interoperability. If not, will you provide mapping to more commonly used ontologies?</p>
2.4. Increase data re-use (through clarifying licences)	<p>Specify how the data will be licensed to permit the widest reuse possible.</p> <p>Specify when the data will be made available for re-use. If applicable, specify why and for what period a data embargo is needed.</p> <p>Specify whether the data produced and/or used in the project is usable by third parties, in particular after the end of the project. If the re-use of some data is restricted, explain why.</p> <p>Describe data quality assurance processes.</p> <p>Specify the length of time for which the data will remain re-usable.</p>

Table I-1. H2020 FAIR DMP Template

Annex II: Adapted FAIR template

Below is the Adapted FAIR Template, as modified for the purpose of describing the ICARUS project datasets.

1. Data summary	
Purpose	
Relation to project	
Types / Formats	
Re-use of other data	
Origin	
Size	
Utility for others	
2. FAIR Data	
2.1. Making data findable, including provisions for metadata	
Metadata provision	
Metadata standards	
Unique identifier	
Naming conventions	
Search keywords	
Version control	
2.2 Making data openly Accessible	
Classification	
Sharing and Access regimes	
Needed method/software	
Repository	
Access authorisation	

2.3. Making data interoperable	
Data/metadata vocabularies and other I/O standards	
Mapping to common ontologies	
2.4. Increase data re-use (through clarifying licences)	
Licence	
Re-use availability schedule	
Re-use by third parties	
Quality assurance	
Availability period	

Table I-1. H2020 FAIR DMP Template