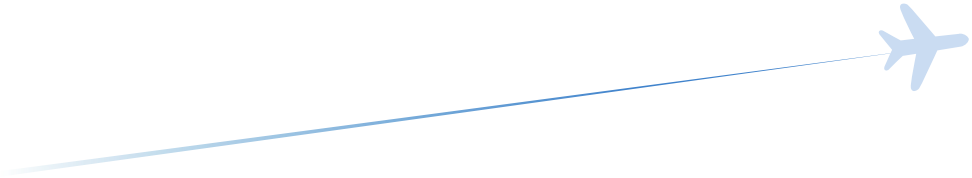

DATA MANAGEMENT PLAN

Deliverable ID:	**D1.2**
Dissemination Level:	**PU**
Project Acronym:	**START**
Grant:	**893204**
Call:	**Call: H2020-SESAR-2019-2**
Topic:	**SESAR-ER4-15-2019-Increased Levels of Automation for the ATM Network**
Consortium Coordinator:	**UC3M**
Edition date:	**15 July 2020**
Edition:	**00.02.00**
Template Edition:	**02.00.02**

Founding Members

EUROPEAN UNION EUROCONTROL

SESAR JOINT UNDERTAKING

Authoring & Approval

Authors of the document

Name/Beneficiary	Position/Title	Date
Manuel Soler (UC3M)	Project Coordinator	15/07/2020

Reviewers internal to the project

Name/Beneficiary	Position/Title	Date
Andrés Muñoz (BGD)	WP2 leader	10/09/2020
Emre Koyuncu (ITU)	WP3 leader	10/09/2020
Daniel Delahaye (ENAC)	WP4 leader	10/09/2020
Raimund Zopp (FLIGHTKEYS)	WP5 leader	10/09/2020
Alexander Kuenz (DLR)	WP6 leader	10/09/2020
Xavier Prats (UPC)	Task 2.1 and 4.1 Leader	10/09/2020

Approved for submission to the SJU By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date
Manuel Soler (UC3M)	Project Coordinator	11/09/2020

Rejected By - Representatives of beneficiaries involved in the project

Name/Beneficiary	Position/Title	Date

Document History

Edition	Date	Status	Author	Justification
00.00.01	20/06/2020	Initial Draft	Manuel Soler	New document
00.00.02	30/06/2020	Complete Draft	Manuel Soler	Sent for internal review
00.01.00	15/07/2020	Submission	Manuel Soler	Internally reviewed
00.01.01	25/08/2020	Reviewed Draft	Manuel Soler	Reviewed by SJU
00.02.00	11/09/2020	Submission	Manuel Soler	Internally reviewed

START

A STABLE AND RESILIENT ATM BY INTEGRATING ROBUST AIRLINE OPERATIONS INTO THE NETWORK

This Deliverable is part of a project that has received funding from the SESAR Joint Undertaking under grant agreement No 893204 under European Union's Horizon 2020 research and innovation programme.

Abstract

The objective of this Deliverable is to describe the Data Management Plan (DMP) of START Project. START is included in the pilot under Horizon 2020 called the Open Research Data Pilot (ORD pilot).

The Deliverable includes a data summary, including the purpose of data to fulfil START objectives, a description of types and formats of input and output data, the origin, accessibility and approximate size of the data, the re-use of data and their utility for different target audiences.

The deliverable also states a plan for continuous update of the document (with versions and content to be updated).

START will follow FAIR data principles, and, thus, the deliverable elaborates on a set of guiding principles to make data Findable, Accessible, Interoperable, and Reusable. Allocation of resources to comply with FAIR principles and the data management strategy of START is provided within the document

Last but not least, an analysis on data security, including aspects related to data sharing and data storage is covered in the Deliverable.

Table of Contents

Abstract 3
1 Introduction 6
1.1 Data Summary 6
1.2 DMP continuous update 12
1.3 Acronyms 13
2 FAIR data 15
2.1 Making data findable, including provisions for metadata 15
2.2 Making data openly accessible 18
2.3 Making data interoperable 18
2.4 Increase data re-use (through clarifying licences) 18
3 Allocation of resources 21
4 Data security 22
4.1 DATA SHARING 22
4.2 DATA Storage and Preservation 23
5 Ethical aspects 24
6 References 25

List of Tables

Table 1: START data input (description) 7
Table 2: START data input (origin, access and size) 9
Table 3: History of changes 12
Table 4: Calendar of updates 12
Table 5: Acronyms 14
Table 6: START consortium acronyms 14

List of Figures

Figure 1: DCAT summary of attributes 17
Figure 2: Creative Commons Attribution 4.0 International 20
Figure 3: GNU General Public License v3.0 20

1 Introduction

The objective of this document is to describe the Data Management Plan of START Project. This plan is based on the "Horizon 2020 DMP" template provided by European Commission (Horizon 2020) [1]. START is included in the pilot under Horizon 2020 called the Open Research Data Pilot (ORD pilot). As participants in the pilot, START is submitting this DMP as a deliverable. The DMP will be updated over the course of the project as indicated in Section 1.2. To create this DMP we are following the guidelines provided by the European Commission on FAIR Data Management in Horizon 2020 [2].

1.1 Data Summary

1.1.1 Purpose

The main goal of the Data Management Plan is to provide an analysis of the main elements of the data management policy that the START consortium will use to comply with the established goals.

Recall that the specific goals of START are:

1. To model uncertainties at the micro (trajectory) level, **assimilate observations (via ADSB/Radar) every 15 min. using advanced data science methods**, and propagate trajectory uncertainties using assimilated models and a stochastic trajectory predictor.
2. To model uncertainties at the macro (ATM network) level, **assimilate observations (satellite data for storm, and network status) every 15 min. using advanced data science methods**, and propagate ATM network uncertainties using the assimilated models.
3. To develop an Artificial Intelligence (AI) algorithm capable of generating a set of pan-European (i.e., considering the whole traffic over Europe) robust trajectories that make the European ATM system resilient when facing these relevant uncertainties.
4. To implement those algorithms as an advanced flight dispatching demo functionality for airspace users to obtain robust trajectories.
5. To validate these concepts through system-wide simulation procedures in order to evaluate their stability, assessing the benefits for both the airspace users and the network manager. Recommendations for the derivation of resilient TBO networks will be derived.

In light of this, it becomes apparent that the availability of data is essential to fulfil with Objective 1 and Objective 2, in which we intend to assimilate data. Thus, they are essential to successfully conduct the research activities of the whole project.

1.1.2 Types and formats

As input data, START will tentatively (START consortium is still discussing details[1]) make use of the data sources to model meteorological phenomena; traffic and network status; and aircraft performance.

Category	Data	Description
Traffic and Network	DDR2	EUROCONTROL, as the Network Manager (NM), provides operational stakeholders with the most accurate picture of past and future pan-European air traffic demand from several years ahead until the day before operations. This also includes environment data, analysis reports and tools (SAAM and NEST), and includes both filed flight plans and flights based on radar tracks. All this information is managed by and can be accessed via the Demand Data Repository V2 (DDR2). Access to the DDR is restricted.
	ADSB	ADS-B messages broadcasted basically by all airliners using their transponders. On the base of those messages, received by ground-based receivers, the trajectories of flight can be reconstructed.
	QAR	Quick Access Recorder (QAR) is a device used to record specific aircraft performance parameters. The purpose of an QAR is to collect and record data from a variety of aircraft sensors. In provide data to re-construct a trajectory, including the dynamics of the aircraft.
Aircraft performance	BADA	BADA (Base of Aircraft Data) is the world's leading aircraft performance model, managed by EUROCONTROL for use by the aviation community. The main application of BADA is trajectory simulation and prediction. Though detailed data on aircraft performance is confidential and commercially sensitive, BADA transforms it such that aircraft models can be made available to a wide audience.
Meteorological products	EPS forecasts	Ensemble Prediction Systems (EPS) are numerical weather prediction (NWP) systems that allow us to estimate the uncertainty in a weather forecast. They represent uncertainty in the initial conditions by creating a set of forecasts starting from slightly different states that are close, but not identical, to the best estimate of the initial state of the atmosphere.
	RDT Satellite observations	The RDT software has been developed in the context of EUMETSAT's Satellite Application Facility for Nowcasting (SAF-NWC) by the Nowcasting Department of Météo-France. The software tracks clouds, identifies those which are convective, and provides some descriptive attributes for their dynamics.

Table 1: START data input (description)

As for output data, three main sources are considered within the START Project:

1. First, the consortium members will produce documentation, mainly related to data analysis, methodologies, processes and so on. The Data produced in that way will be presented as Open

[1] Detailed description of the data, including its origin, size, access rights (if any), utility and metadata (if any) will be provided as appendixes to this deliverable in the periodical updates (see Section 1.2).

Access under free licence whenever possible.[2] All papers, reports and articles will be presented as Green-Open Access. Both OpenAire platform and the START Web-page will host this documentation. Furthermore, there will be some documents, that will be also published as Gold-Open Access. See Section 3 for the allocation of recourses to cover it

2. Second, START will produce libraries/software as main output of the project. They will follow the software management plan (see[3]), including languages and standards specified therein. Some of these libraries might be offered to the community in open source (eventually with some restrictions), while others might be kept closed for further exploitation by the partners of START. The decision of whether to disclose this libraries/software or not will be taken by the partners upon discussion and deliberation within the General Assembly (see PMP [3] for the management structure of the Project).

3. Last, the consortium will run use cases on specific days. Traffic all over Europe for a given window of 4 hours will be considered. Altogether, thousands of flights per use case. We will compute the initial flight plan for each of the flights (each initial flight plan will be an xml object), we will incorporate uncertainties at both trajectory and network level, and then run the optimization metaheuristic to calculate the "optimized flight plans" that make the system resilient. Then, we will have again thousands of "optimized flight plans", each as a xml object. All these objects will be stored and in an internal database for its common use during the project. They will not be shared.

1.1.3 Origin, access and size of the data

Category	Data	Origin	Access	Size
Traffic and Network	DDR2	Eurocontrol.	Restricted access.	~1Gb (1 day in Europe)
	ADSB	ADS-B messages.	Open access: • In-house antennas. • OpenSky network Service-based: • Flight-Radar24	~1Gb (1 hour of traffic over Europe)
	QAR	Airlines and aircraft manufacturers	Confidential access	~1Mb (1 flight)
Aircraft performance	BADA4	Eurocontrol	Restricted access	~1Mb

[2] Note that, as recognized by the Commission, participating in the ORD Pilot does not necessarily mean opening up all your research data. Rather, the ORD pilot follows the principle "as open as possible, as closed as necessary" and focuses on encouraging sound data management as an essential part of research best practice. There might good reasons to keep some or even all research data generated in a project closed. In the case of sharing data under restrictions, explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Meteorological products	EPS forecasts	Meteorological Agencies	Restricted access	~1Gb (1 forecast)
	RDT Satellite observation	Meteorological Agencies	Restricted access[3]	~100Mb (1 day)

Table 2: START data input (origin, access and size)

DDR2 data access:

EUROCONTROL, as the Network Manager (NM), provides operational stakeholders with the most accurate picture of past and future pan-European air traffic demand from several years ahead until the day before operations. See info https://www.eurocontrol.int/ddr. Access to the DDR portal is restricted to air navigation service providers (ANSPs) and airline operators (AOs) within Europe, on the basis of approval of a licence agreement and formal acceptance by EUROCONTROL. The agreement does not include any constraint regarding the publication of results from these data. Airlines are allowed to download (historical or forecast) traffic data for their own fleets only. Access was also provided to researchers, however, as of June 2019, Eurocontrol no longer provides access to the DDR portal to students/researchers. All the partners of the project (UC3M, BGD, ITU, ENAC, UPC, DLR, and FLIGHTKEYS) have had access to DDR2 data in the past (prior to June 2019) on the basis of the approval of a license agreement and formal acceptance by Eurocontrol. They have downloaded and stored the information and they are allowed to use it within the framework of START project.

ADSB data access:

In the case of Open-Sky (https://opensky-network.org/**)**, OpenSky Network's authorization to access the data grants[4] You a limited, non-exclusive, non-transferable, non-assignable, and terminable license to copy, modify, and use the data in accordance with this AGREEMENT solely for the purpose of non-profit research, non-profit education, commercial internal testing and evaluation of the data, or for government purposes. No license is granted for any other purpose and there are no implied licenses in this AGREEMENT. This AGREEMENT shall become effective as of the date of approval by OpenSky Network. This AGREEMENT is subject to modification or revocation at any time at OpenSky Network sole discretion. OpenSky Network shall have the right to use any of Your feedback received during the license period for its non-profit educational and/or research purposes.

In the **case of FlightRadar24**: following the license agreement reached with FlightRadar24 AB for the usage of ADS-B data (with data meaning the tracking of air traffic with world coverage), BGD can use and reproduce data and extracts of the data solely within its system for its internal use in research activities and for the production of documentation for internal, non-commercial use. Additionally, the agreement allows to use, reproduce and distribute research publications and results which are derived or based on the data for non-commercial use or purposes, ensuring adequate acknowledgement of Flightradar24 as source of the data. The data cannot be transferred or copied. The raw data cannot be lent to any other person or access permitted. BGD cannot permit any party to download, extract or

[3] http://www.nwcsaf.org/

[4] the general terms of use and data license agreement can be consulted at https://opensky-network.org/index.php/about/terms-of-use

transmit to any other system and/or software or to copy or otherwise reproduce all or any part of the data. In the particular case of START Project, only BGD intends to use ADS-B data.

QAR data access:

Regarding QAR data, BGD can use and reproduce data and extracts of data solely within its system for its internal use in research activities and for the production of documentation for internal, non-commercial use. The use, reproduction and distribution of research publications and results which are derived or based on the data for non-commercial use or purposes is allowed when ensuring the confidentiality of the data. The data cannot be transferred or copied. The raw data cannot be lent to any other person or access permitted. BGD cannot permit any party to download, extract or transmit to any other system and/or software or to copy or otherwise reproduce all or any part of the data.

BADA4 data access:

The use of BADA is regulated through a license agreement which stipulates the terms and conditions of use based on the contractual constraints EUROCONTROL is committed to with its data providers. The provision of access is subject to EUROCONTROL's approval. The agreement and its conditions are accessible through the BADA Users Interface. The agreement does not include any constraint regarding the publication of results from these data. All the partners of the project (UC3M, BGD, ITU, ENAC, UPC, DLR, and FLIGHTKEYS) have had access to BADA 4 data in the past. However, BADA request to ask for a specific license associated to the activities of START project. All the partners are in the process to ask for a license.

Meteorological data access

Both EPS forecasts and RDT Satellite observation can be accessed via AEMET, the Spanish met service. According to the agreement signed between the Directorate General for Scientific and Technical Research (Spanish institution) and the State Meteorological Agency (AEMET), the main researcher (in the case, referring to Manuel Soler, PC of START) must submit the Form of application for Meteorological and Climatological Benefits L2. This application form must be submitted independently to the project report the list of data corresponding to both observations and results of the execution of models, including data from analysis, reanalysis and meteorological prediction prepared by AEMET and / or the European Prediction Center Medium Term that foresees that they will be necessary to carry out the project. The provision of the information supply service shall comply with the AEMET Resolution of December 30, 2015 (BOE of January 5, 2016) which establishes the public prices that must govern the provision of meteorological and climatological services. In Annex III of said Resolution it is established that Research Bodies, officially recognized as such, for carrying out non-profit research projects up to a maximum of € 6,000.00 are excluded from the application of public prices. From this amount a 50% discount will be applied to the rates including management costs. Therefore, if it is estimated that the cost will exceed the established limit of € 6,000, the extra cost must be included as an eligible expense in the research project proposal, motivating such need. The agreement does not include any constraint regarding the publication of results from these data.

In the case of START project, which will only conduct research activities, selected dates for use cases will be limited to few of them, not going beyond the 6000€ limit. Thus, access to the data will be provided for free. Data request will be launched as soon as dates for the use cases are selected (before the first General Assembly).

1.1.4 Re-use of data

Data reuse is a concept that involves using research data for a research activity or purpose other than that for which it was originally intended. In light of this, all the data input needed for START (see Table 1**Error! Reference source not found.**) will be re-used from its original purpose

START will also foster its output data reuse by submitting data to OpenAire and by ensuring that data are described and comprehensible to any other researcher who may want to use it. When required, we will always follow metadata schemas to describing datasets so that they can be reused over time. In particular, revision of the present document (see Section 1.2) will incorporate detailed information, including metadata, about the data used for START research activities. Similarly, software libraries will follow standards, will be documented and will include test cases to facilitate its re-use.

1.1.5 Data utility

Data utility is a measure of the business value attributed to data within specific usage contexts. Though utility is typically measured in monetary units, in the context of research, data utility can be measured in terms of knowledge re-use by different target communities.

The target audiences for START output data utility will be:

1. General public.
2. Undergraduate and graduate students.
3. Research and scientific community.
4. Stakeholders.
5. Institutions.

In particular START will be making available different types of data to maximize the data utility for the different target groups. In particular:

For the General public, START will be realising a Concept Note, press-releases, and periodical updates with key results. The documentations, yet key data (e.g., videos, animations) will be released following the FAIR principles stated in Chapter 2. The goal is to ensure that the research activities are made known to society in such a way that they can be understood by non-specialists, thereby improving the public's understanding of science and technology. Establish a concern on how European collaboration and funding contributes to society.

For undergraduate and graduate students, START will participate in science weeks, researchers' nights, and incorporate some of the results in the teaching activities of the academic institutions, including Bachelor and Master Theses. Similarly, the documents and key data (e.g., videos, animations), including those used for teaching activities, will be released following the FAIR principles stated in Chapter 2. The goal is to wake up interest to follow scientific careers, especially to join START groups and follow PhD studies and develop BSc and MSc thesis.

For the research and scientific community, START main activity will be the publication of scientific papers and the development of open-source libraries of its re-use. The papers (including relevant metadata) will be published open source (as a general rule, following green-open access schemes at OpenAIRE; in some cases, following gold open-access schemes); some of the libraries will be also realises as open source using (including standardized documentation), e.g., uploading them in a github

repository. The goal is maximizing the dissemination, including other scientific disciplines; enhance excellence and scientific reputation; find follow-up ideas and collaborations.

For the Stakeholders, besides scientific papers and open-source libraries, which could be used on a free-basis to further develop commercial tools, commercial libraries/services will be also developed during the project. Fostering the commercial use of these libraries is another objective of START.

For the Institutions, START intends to show (gathering all the data to be generated targeted to the previous groups) that the research and the re-usability of data in indeed interesting for the aviation community. The goal, in the end, is to draw the attention of different institutions, e.g., European Commission and SESAR and International institutions (ICAO, EASA) and/or groups within them (ICAO's TBO group), to make them aware of START results and make them visible to their agendas. This will facilitate the allocation of more funding and the revision/modification of standards.

1.2 DMP continuous update

The DMP is intended to be a living document in which information will be made available on a finer level of granularity through updates as the implementation of START project progresses and when significant changes occur. Therefore, START's DMPs will have version numbers and will include a timetable (see Table 3) specifying the updates.

History of changes		
Version	Publication date	Change
1.0	31.07.2020	Initial version

Table 3: History of changes

We envision three phases for the DMP:

1. Initial DMP (Current version of the document), which includes the main guidelines.
2. Detailed DMP, updated over the course of the project whenever significant changes arise (updated, at least, every 6 months coinciding with the General Assembly meetings), such as (but not limited to): new data; changes in consortium policies (e.g. decision to file for a patent); changes in consortium composition and external factors.
3. Final DMP, reviewed at the final review meeting at the latest with complete information.

Version	Publication date	Change
1.0	31/07/2020	Initial version
2.0	30/11/2020	Incorporation of detailed description of data sources
3.0	30/05/2021	Incorporation of the specification of the data used in WP2 and WP3
4.0	30/11/2021	Incorporation of the specification of the data used in WP4
5.0	30/04/2021	Incorporation of the specification of the data used in WP5 and WP6

Table 4: Calendar of updates

1.3 Acronyms

Non-exhaustive list of acronyms used across the text.

Acronym	Description
ADSB	Automatic Dependant Surveillance Broadcast
AI	Artificial Intelligence
ATM	Air Traffic Management
BADA	Base of Aircraft DAta
CA	Consortium Agreement
DB	DataBase
DCAT	Data Catalog Vocabulary
DDR2	Demand Data Repository V2
DMP	Data Management Plan
EASA	European Union Aviation Safety Agency
EPS	Ensemble Probabilistic Systems
FAIR	Findable, Accessible, Interoperable and Reusable
FDR	Flight Data Recorder
GA	Grant Agreement
GAss	General Assembly
ICAO	International Civil Aviation Organization
JU	Joint Undertaking
LDAP	Lightweight Directory Access Protocol
NDA	Non-Disclosure Agreement
NM	Network Manager
NWP	Numerical Weather Prediction
ORD	Open Research Data
PC	Project Coordinator
PMP	Project Management Plan
RDT	Rapidly Developing Thunderstorms
SAF	Satellite Application Facilities
SESAR	Single European Sky ATM Research Programme
SJU	SESAR Joint Undertaking
START	a Stable and resilienT ATM by integrAting Robust airline operations into the neTwork
TBO	Trajectory-Based Operations

VPN	Virtual Private Network
WP	Work Package

Table 5: Acronyms

START Consortium

Acronym	Description
BGD	Boeing Research and Technology Germany
DLR	German Aerospace Center
ENAC	Ecole Nationale de l'Aviation Civile
FLIGHTKEYS	FlightKeys
ITU	Istambul Teknik Universitesi
UC3M	University Carlos III of Madrid
UPC	Universitat Politecnica de Catalunya

Table 6: START consortium acronyms

2 FAIR data

FAIR data refers to a set of guiding principles to make data Findable, Accessible, Interoperable, and Reusable. The term was first presented at a Lorentz Workshop in 2014[5] and the final principles were published in Nature in 2016[6]. The START Project will follow those principles as shown below.

2.1 Making data findable, including provisions for metadata

The findability of data is the ability to locate information by other users. To fulfil this principle, START will provide the necessary metadata to locate the information. As referred in the FAIR principles, the necessary metadata will include, wherever it is possible, the assignment of a unique and persisted identifier to each dataset, a data description with rich metadata, the inclusion of the identifier in the metadata and register on a searchable index the metadata.

2.1.1 Metadata standards

Sharing data resources to different target audiences, including citizens, requires the provision of metadata. This is irrespective of the data being open or not. START will rely of DCAT[7] (alternatively, on r3data.org for those documents in OpenAire) standard as a general rule.

DCAT is a vocabulary for publishing data catalogs on the Web, which was originally developed in the context of government data catalogs such as data.gov and data.gov.uk, but it is also applicable and has been used in other contexts.

DCAT provides classes and properties to allow datasets and data services to be described and included in a catalog. The use of a standard model and vocabulary facilitates the consumption and aggregation of metadata from multiple catalogs, which can:

- increase the discoverability of datasets and data services
- allow federated search for datasets across catalogs in multiple sites

[5] https://www.force11.org/group/fairgroup/fairprinciples

[6] https://www.nature.com/articles/sdata201618

[7] Data Catalog Vocabulary, W3C Recommendation 04/02/2020 https://www.w3.org/TR/vocab-dcat/

Data described in such a catalogue can come in many formats, ranging from spreadsheets, through XML and RDF to various specialized formats. DCAT is based around six main classes (see Figure 1)

- *dcat:Catalog* represents a catalogue, which is a dataset in which each individual item is a metadata record describing some resource; the scope of *dcat:Catalog* is collections of metadata about datasets or data services.
- *dcat:Resource* represents a dataset, a data service or any other resource that may be described by a metadata record in a catalogue. This class is not intended to be used directly, but is the parent class of *dcat:Dataset, dcat:DataService and dcat:Catalog*. Member items in a catalogue should be members of one of the sub-classes, or of a sub-class of these, or of a sub-class of *dcat:Resource* defined in a DCAT profile. *dcat:Resource* is effectively an extension point for defining a catalogue of any kind of resource. *dcat:Dataset* and *dcat:DataService* can be used for datasets and services which are not documented in any catalogue.
- *dcat:Dataset* represents a dataset. A dataset is a collection of data, published or curated by a single agent. Data comes in many forms including numbers, words, pixels, imagery, sound and other multi-media, and potentially other types, any of which might be collected into a dataset.
- *dcat:Distribution* represents an accessible form of a dataset such as a downloadable file.
- *dcat:DataService* represents a data service. A data service is a collection of operations accessible through an interface (API) that provide access to one or more datasets or data processing functions.
- *dcat:CatalogRecord* represents a metadata item in the catalogue, primarily concerning the registration information, such as who added the item and when.

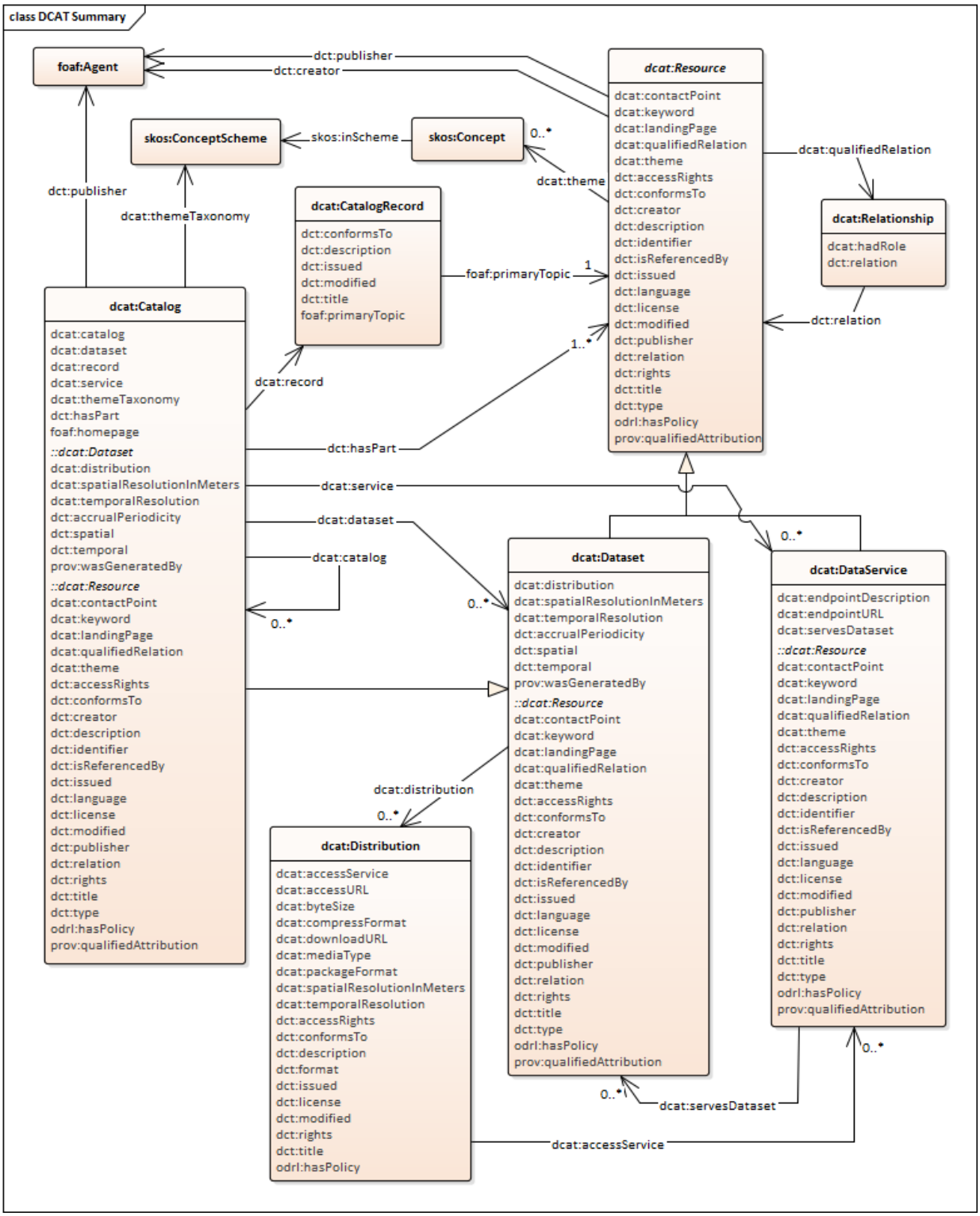

Figure 1: DCAT summary of attributes

2.2 Making data openly accessible

START, as previously stated, will produce two types of data outputs, namely: documentation (scientific and outreach) and libraries/software. Data and results will be, whenever it is possible and in accordance with the Grant Agreement Art.26 [4], published openly and provide access.

2.2.1 Protection of data

As for the decision on whether to published data or protect them, it should be stated that participating in the ORD Pilot does not necessarily mean opening up all the research data, as recognized by the Commission. Rather, START will follow the ORD pilot principle "as open as possible, as closed as necessary", which is encouraging sound data management as an essential part of research best practice. The PMP establishes the procedures for internal review of documentation, including scientific publications (which might be subject to protection) and the eventual release of a library/software as open source (conversely, protected for further exploitation). The governance structure of START (refer to [3]) includes an exploitation board and a dissemination board (both of them included in the General Assembly board). Any decision to be taken, must be previously discussed at these forum levels.

2.2.2 Accessibility of data

As for documents, there is no specific software requirement or documentation needed:

- Reports (in pdf format, including metadata) will be accessible mainly via project web-page.
- Scientific papers will be made available via OpenAIRE (in pdf format, including metadata) and https://arxiv.org/, besides the website of the project. OpenAIRE has an arrangement with UC3M (the Project Coordinator) institutional repository

As for open-source software libraries, they will be stored in GitHub, including the standardized documentation needed to use and access to the available data within it. The documentation will include software requirements (e.g., Python version and third-party libraries required) and instructions on how to install it and run available tests/examples.

2.3 Making data interoperable

START will make its data interoperable, i.e., allowing data exchange and re-use between researchers, institutions, organisations, including different countries. For that purpose, we will use standards for the metadata (namely DCAT and r3data.org for those documents in OpenAIRE, as indicated in Section 2.1.1). The libraries and software will be also developed following standards (see more details about software management in the [3]), e.g., we will follow the PEP8 standard in Python developments, including coding, documentation, and test standards to ensure interoperability.

2.4 Increase data re-use (through clarifying licences)

In order to increase data reuse, START will first follow a decision process on whether to protect certain data or make them freely available. This has been covered in Section 2.2.1.

In the case of open release, we will add an open license.

An open license is a notice that grants the recipient of a creation (can be a software, a document, or other creations) extensive rights (with certain conditions) to use, modify and redistribute that creation. These actions are usually prohibited by copyright law, but the rights-holder (usually the author) of a creation can remove these restrictions by accompanying the creation with a license which grants the recipient these rights. Using such a license is free as conferred by the copyright holder. However, the way authorship must attribute, the way of further sharing, licensing derived products, and eventually commercialized is included in the open license.

The permissions and limitations of a free license are classified according to the following subjects:

- Linking - linking of the licensed code with code licensed under a different licence (e.g. when the code is provided as a library)
- Distribution: distribution of the code to third parties
- Modification: modification of the code by a licensee
- Patent grant: protection of licensees from patent claims made by code contributors regarding their contribution, and protection of contributors from patent claims made by licensees
- Private use: whether modification to the code must be shared with the community or may be used privately (e.g. internal use by a corporation)
- Sublicensing: whether modified code may be licensed under a different licence (for example a copyright) or must retain the same licence under which it was provided
- TM grant: use of trademarks associated with the licensed code or its contributors by a licensee

Recent tools like the European Commissions' Joinup Licensing Assistant[8], makes possible the licenses selection and comparison based on more than 40 subjects or categories, with access to their SPDX identifier and full text.

Though START will be exploring the most adequate license for each case, tentatively, START:

- For documents, will rely on Creative Commons Attribution 4.0 Int. (see Fig. Figure 2)
- For software libraries, will rely on GNU General Public License v3.0 (see Figure 33)

[8] https://joinup.ec.europa.eu/solution/joinup-licensing-assistant/joinup-licensing-assistant-jla

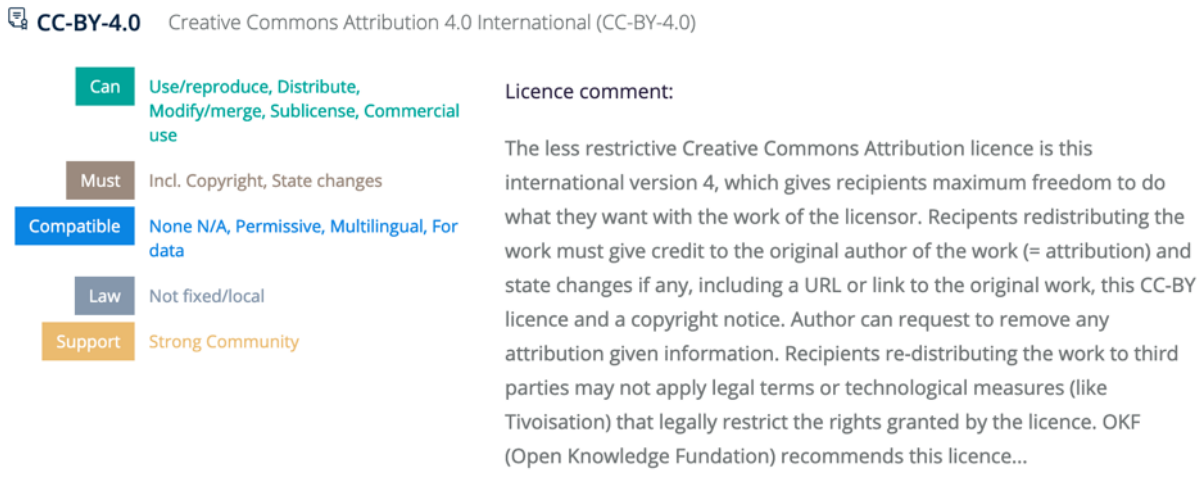

Figure 2: Creative Commons Attribution 4.0 International

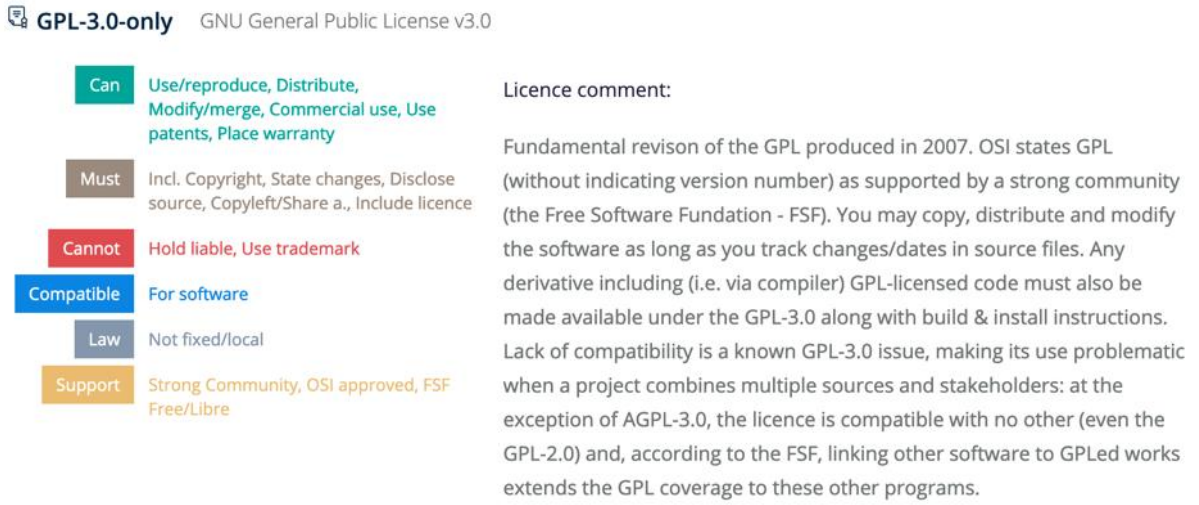

Figure 3: GNU General Public License v3.0

3 Allocation of resources

As defined in the H2020 Grant Agreement [4], in particular Article 6 and Article 6.2, costs related to open access to research data are eligible for reimbursement.

Accordingly, and in order to make START data FAIR, the project has allocated:

- 4000€ per partner to gold open access publications.

Moreover, the following activities are also covered:

- The publication of research articles in Green Access (via OpenAIRE) is covered by institutional resources. For instance, UC3M's institutional repository is directly linked to OpenAIRE.
- The publications of open-source libraries in GitHub is free, including its maintenance.
- UC3M team has a Server with capacity over 100Tb for data storage where the research data needed for the project will be allocated.

UC3M, as coordinator of the project, will be responsible for the management of data.

Preserved datasets will need to be updated and this means a data preservation policy and process will need to be defined. A central consideration for any long-term DMP is the cost of preserving that data and what will happen after the completion of the project. Preservation costs may be considerable depending on the exploitation of the project after its finalization. Examples include:

- Personnel time for data preparation, management, documentation, and preservation, Hardware and/or software needed for data management, backing up, security, documentation, and preservation,
- Costs associated with submitting the data to an archive,
- Costs of maintaining the physical backup copies (disks age and need to be replaced).

These topics will be discussed during the Project at the General Assembly level. If any relevant conclusions or recommendations are reached, they will be included in potential updates of this DMP. This will be updated in future version of the present document.

4 Data security

4.1 DATA SHARING

A number of datasets will be used during START, which will share those datasets with respect to the legal and IPR constraints. Specifically, we have to consider where, how, and to whom the data could be made available. This information will be updated in the updates of this document, together with the description of the data.

Datasets will be shared either by opening a specific API for third parties to fetch datasets from the START store, or via making such datasets downloadable as single files. In any case, the methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data.

The following general conditions for accessing and using the data provided by UC3M (as data management leader in START project) apply:

- The Dataset provided will never be the raw data obtained from operational sources, but the result of the fusion and processing performed by UC3M (or, eventually, other START partners providing the data).
- The Dataset can only be used to achieve the research purposes stated in the START Technical Description of the Action. No other use is allowed.
- The user is not allowed to create or derive new datasets from the original one.
- The Dataset will be anonymized through its “delocalization” (either in place or time) except in specific cases where the specific confidentiality agreement allows data concerning places and time to appear. Delocalization will be done to disable its traceability to operational events and situations. The user might not change or process the Dataset in any way to remove the anonymity of the data.
- The Dataset will be stored physically in the premises of UC3M. No physical copy of the Dataset will be provided. The users will not store, copy or otherwise move the Dataset (physically or logically) to databases or systems outside the premises of UC3M.
- Access to the Dataset will be granted as needed by the research members of the consortium either for implementing their own tasks under the START action or for exploiting their own results.
- Access to the Dataset will be granted under a specific confidentiality agreement that will clearly identify the Dataset users, purpose, usage timeframe and any specific clauses that might be needed.
- The Dataset access confidentiality agreement will be signed before the start of the work to set the conditions of this access. UC3M (or, eventually, other START partners providing the data) is committed to provide access in a secure way, royalty-free, in order to achieve the research goals of the project in an efficient way.
- Access to the Dataset will be provided only for the specific purposes and scheduled time windows described in each one of the user confidentiality agreements.
- For security and confidentiality reasons no permanent access to the Dataset will be granted.
- The Dataset will be available through a secure channel as specified by UC3M.

In particular, and following Table X data, the following sharing conditions will apply:

4.2 DATA Storage and Preservation

In START we aim to address issues concerning long term storage of data as well as research generated data since there is a growing interest in archiving and performing research over data.

These issues related to safety and long-term data storage are complicated when the data evolves or/and they could be merged with other data coming from other sources. So, in this project we need to consider:

- What is the volume of the data to be maintained?
- What is considered long-term (2-3 years, 10 years, etc.)?
- Identification of archive for long-term preservation of data.
- Which datasets will need to be preserved in the archive?
- What about relevant dependent datasets?

These questions will be assessed throughout the project and the planned updates of the DMP.

In terms of security, data stored in UC3M (similarly with other partners in START consortium) premises is currently accessed from the outside by using a VPN access with authentication. Retrieval of data is done via DB which requires additional LDAP authentication, and permissions are granted according to each profile allowing several levels of access to data. Data is also replicated by a factor of 2, ensuring no data loss happens. Additionally, NDAs are signed by all the users with granted access.

5 Ethical aspects

No ethical issues have been identified for START since it does not deal with any personal data. From a legal perspective, all restricted data used within START will be strictly used for START research purposes only, following a formal request.

6 References

1. Commission, E., *H2020 templates: Data management plan v1.0* 13.10.2016.
2. Innovation, E.C.D.G.f.R., *H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020. V 3.0*. 26 July 2016.
3. Soler, M., *Project Management Plan. START Project. D1.1*. 2020.
4. Commission, E., *AGA – Annotated Model Grant Agreement. Version 5.2, 26 June 2019.* 2019.

Participant No		Participant organisation name	Country
1 - BRTE		Boeing Research and Technology Germany (BRTE)	Germany
2 - DLR		German Aerospace Center (DLR)	Germany
3- ENAC		Ecole Nationale de l'Aviation Civile (ENAC)	France
4- FK		FlightKeys (FK)	Austria
5- ITU		Istambul Teknik Universitesi (ITU)	Turkey
6 – UC3M (Coordinator)		**Universidad Carlos III de Madrid (UC3M)**	**Spain**
7 - UPC		Universitat Politecnica de Catalunya (UPC)	Spain