



Next generation data-driven reference European models and methods towards silent and green aircraft operations around airports

Horizon Europe | HORIZON-CL5-2022-D5-01-12

Data Management Plan



This project receives funding from the European Union's Horizon Europe research and innovation programme under grant agreement no. 101069508 (HighSpin).

This publication reflects only the author's view and the European Climate, Infrastructure and Environment Executive Agency (CINEA) is not responsible for any use that may be made of the information it contains.

REVISION HISTORY

Deliverable No.	D7.2	
Deliverable Title	Data Management Plan	
Deliverable Type	Report	
Dissemination level	Public	
Written By	Michele DE GENNARO	11-07-2023
Checked by	Alessandro ZANON	18-07-2023
Approved by	Alessandro ZANON	18-07-2023
Status	Final	18-07-2023

Version	Date	Who	Change
1.0	11.07.2023	Michele DE GENNARO	Initial Version

TABLE OF CONTENTS

Revision History	2
Project Abstract.....	4
List of Abbreviations.....	5
Executive Summary	6
1 Introduction	7
2. Data Summary	8
3. FAIR data.....	10
3.1. Making data findable, including provisions for metadata	10
3.2. Making data accessible	10
3.3. Making data interoperable	12
3.4. Increase data re-use.....	12
4. Other Aspects.....	14
4.1. Other research outputs.....	14
4.2. Allocation of resources	14
4.3. Data security.....	15
4.4. Ethics.....	15
4.5. Other issues.....	15
5. Conclusions.....	16
6. References.....	17

PROJECT ABSTRACT

NEEDED responds to the second and third bullets of the “expected outcome” of the HORIZON-CL5-2022-D5-01-12 topic, delivering the next generation data-driven reference European models and methods to estimate present and future aircraft emissions (pollutants and noise), achieving TRL 4 at the end of the project. To do so, NEEDED will advance the state of the art by:

- improving the accuracy of the reconstruction of aircraft operations by using real-world ADS-B data,
- advancing emission inventories for current and future aircraft technologies, while delivering more accurate pollution dispersion models,
- extending the applicability of the ECAC Doc. 29 noise model towards future aircraft technologies,
- performing more accurate estimation of the number of people affected by local air transport operations by using dynamic population maps.

These activities are complemented by (i) local air quality (LAQ) and experimental noise measurements performed at Rotterdam Airport, (ii) validation of the NEEDED tool chain in a 30-week pilot study involving three airports, and (iii) delivery of a methodology to optimise the flight patterns for minimum detrimental impact on the population in present and future scenarios. The project aims to function as a technology enabler, laying the methodological groundwork for facilitating the entry into service of transformative aircraft technologies while capitalising on the potential of ADS-B data. NEEDED ensures its impact on the next generation of Air Traffic Management (ATM) regulation and policies through the direct involvement of EUROCONTROL.

The consortium combines a wide portfolio of competences from 10 partners from 7 different EU member states (Austria, Belgium, Italy, Sweden, The Netherlands, France, and Spain) plus 1 non-EU Country and it is coordinated by AIT Austrian Institute of Technology. NEEDED is scheduled to run from January 1st, 2023, to December 31st, 2026, for a total duration of 48 months and has received funding from the European Union’s Horizon Europe research and innovation programme under Grant Agreement no. 101095754. A full list of partners and funding can be found at: <https://cordis.europa.eu/project/id/101095754>.

LIST OF ABBREVIATIONS

Acronym / Short Name	Meaning
DMP	Data Management Plan

EXECUTIVE SUMMARY

The objective of the Data Management Plans (DMP) consists of stating what data will be created and how, and outline the plans for sharing and preservation, noting what is appropriate given the nature of the data and any restrictions that may need to be applied“ [1].

Good data management can facilitate the sharing of results and knowledge and promote the EU’s open science objectives [2]. Such research data management is often done according to the FAIR principles [3].

The present document provides a first, indicative plan for what kinds of data is expected to be generated in the NEEDED project and how they will be managed, following an established DMP template.

1 INTRODUCTION

The NEEDED project comprises a wide range of activities to achieve its goals (see project abstract, above), including, but not limited, to key activities such as flight path trajectories reconstruction, emission inventories for future aircraft technology, advancement of pollutants’ dispersion and noise models, and estimation of the population adversely affected by civil air traffic. Various types of data will be managed and generated during these activities.

The present document is the first version of the NEEDED data management plan (DMP). It is an indicative plan as to what kind of data the project beneficiaries expect to generate during the project, and how these data will be managed. In establishing the DMP, we have followed a standard DMP template [4].

Section 2 Data Summary and section 3 FAIR data follow the structure of the DMP template. Sections 4-4 of the template are summarised as sub-headings in section 4 Other Aspects, as shown in Table 1. Blue text represents the verbatim questions/prompts from the DMP template whereas the black text underneath is the direct response to these questions.

Table 1: Mapping of key DMP sections between this DMP and the reference template.

This document	DMP template
1. Introduction	-
2. Data summary	1. Data summary
3. FAIR data	2. FAIR data
4.1 Other research outputs	3. Other research outputs
4.2 Allocation of resources	4. Allocation of resources
4.3 Data security	5. Data security
4.4 Ethics	6. Ethics
4.5 Other issues	7. Other issues

2. DATA SUMMARY

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existing data has been considered but discarded.

NEEDED will re-use know-how from previous projects indicated as reference activities from the consortium partners. The most prominent background projects are: ARTEM (H2020 GA 769350), ANIMA (H2020 GA 769627), RAPTOR (H2020 GA 863969), IMOTHEP (H2020 GA 875006), ACACIA (H2020 GA 875036), OLGA (H2020 GA 101036871), TULIPS (H2020 GA 101036996). However, NEEDED will harvest, generate, and manage its own data, specifically the areas of:

- real-world ADS-B data for flight path reconstruction,
- emission inventories for current and future aircraft propulsion technologies and pollution data as produced by dispersion models,
- noise data as produced by the application of the ECAC Doc. 29 model and its updates,
- data on number of people affected by local air transport operations.

The reason for reuse is that NEEDED intentionally aims to re-use and build on previously developed elements (where applicable and useful) to avoid spending time and resources developing them again.

What types and formats of data will the project generate or re-use?

A wide range of data types and formats are expected to be generated in the project. While these cannot and need not be exhaustively enumerated ex ante, some general remarks can be made. The data types will vary according to the source application in which they are generated. Generally, the consortium will endeavour to utilise common data formats. While Table 2 summarises the project work packages, Table 3 gives a summary of the expected types of data expected to be generated in the project.

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

The purpose of the data generation is to support the activities of the project in the pursuit of project goals. Data about flight paths, emissions, noise and population is an integral part of the project activities and needed to determine whether the project has succeeded in its aims, as well as to support dissemination activities.

What is the expected size of the data that you intend to generate or re-use?

The expected size of the data generated is largely depending on its source: data is expected of the order of several tens of gigabytes.

What is the origin/provenance of the data, either generated or re-used?

Most of the data in the project will be generated in-project. For re-use, see further above.

To whom might your data be useful ('data utility'), outside your project?

The consortium expects some of the data to be of interest to researchers in the domains of emission and noise modelling from air transport activities.

Table 2: Work packages in NEEDED

WP	Name
WP1	Coordination with EUROCONTROL, ICAO CAEP and other international working groups and projects
WP2	ADS-B data-driven aircraft operations reconstruction
WP3	Airport LAQ models, methods, and measurement technologies
WP4	Reference noise models and measurement technologies
WP5	Dynamic population exposure maps, future scenarios and minimisation of people exposed to LAQ and noise
WP6	Dissemination, communication, exploitation and IPR management
WP7	Project management

Table 3: Research data expected to be generated or re-used in NEEDED

	Type of data	Data formats and file extensions
WP1	Requirements for air traffic data, emissions inventories, and noise estimation.	mostly reports (.pdf)
WP2	ADS-B data and flight trajectories.	Raw data and datasets (.dat), excel tables or lists (.xls, .csv), any ad-hoc format applicable to ADS-B as per consortium partners.
WP3	Local air quality data, modelling data, emission inventories.	Data and reports (.xlsx, .pdf, .csv), images and graphs (.jpg, .png, .pdf.), simulation data set and analysis files (.dta, .sta).
WP4	Noise data and noise maps over airport, airfields and populated regions.	Data and reports (.xlsx, .pdf, .csv), images and graphs (.jpg, .png, .pdf.), simulation data set and analysis files (.dta, .sta).
WP5	Exposure maps, population maps.	Data and reports (.xlsx, .pdf, .csv), images and graphs (.jpg, .png, .pdf.), simulation data set and analysis files (.dta, .sta).

Table 4: Non-research data expected to be generated or re-used in NEEDED

	Type of data	Data formats
WP6	Website traffic; audience engagement statistics from social media; photos and/or videos of project participants	Exports from analytics tool for website and social media – probably .xlsx and/or .pdf; images (.jpg) and videos (.mp4)
WP7	Project management data (e.g. financial, resources)	Excel lists (.xlsx) and reports (.pdf)

3. FAIR DATA

3.1. MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA

Will data be identified by a persistent identifier?

Unless any project beneficiary applies persistent identifiers for their own use, no persistent data identifiers are envisaged for the purposes of information sharing between partners in the project. However, any datasets made available as open data will leverage the possibilities for persistent identifiers as provided by the host platform (e.g. Zenodo).

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

The consortium is aware of ongoing efforts to define a common language in the area of estimation and quantification of the air traffic operation environmental impact, however, not of any ready-to-use taxonomy or similar specifically tailored for this specific subject. NEEDED consortium will leverage whatever possibilities are offered by the repository in which any open data from the project is hosted. Absent a framework for populating metadata, the beneficiaries providing the data will provide metadata such that interested parties will gain an understanding of the provenance and meaning of the data that is sufficient to interpret and reuse it. Some of the main types of metadata that are envisaged include:

- Description of the test setup from which the data was generated
- Meaning of the different fields/variable in the data files
- Reference to the project in which the data was generated, for further information and context.

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-use?

To be determined.

Will metadata be offered in such a way that it can be harvested and indexed?

If the host repository supports this, then yes (provided this can be done with reasonable effort on the part of the consortium).

3.2. MAKING DATA ACCESSIBLE

Repository: Most probably zenodo.org

Will the data be deposited in a trusted repository?

Yes, see above (if data is published in open access).

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

No, we expect that (a) user account(s) will suffice to publish the data.

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a digital object?

To be determined.

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate interests or other constraints as per the Grant Agreement.

Data are subject to a case-per-case evaluation to determine whether they should be made openly available, in order to protect industrially/commercially sensitive information. Only selected data will be shared after internal consultation and approval in the consortium.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

No ex-ante policy on this; to be determined on a case-by-case basis if needed, in conjunction also with activities under *T6.3 Exploitation and IPR management* and *T7.3 Quality, data and risk management*.

Will the data be accessible through a free and standardized access protocol?

Open data will be published in common formats where this is possible and makes sense. The access protocol will most likely simply be https (we expect to use zenodo.org)

If there are restrictions on use, how will access be provided to the data, both during and after the end of the project?

To be determined on a case-by-case basis.

How will the identity of the person accessing the data be ascertained?

For open data, no tracking of people accessing the data is planned from NEEDED consortium side. From host platform side, identification will take place presumably via the authentication protocols and server logs of the platform itself.

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data)?

No personal or sensitive research data is planned to be generated, much less shared, in the NEEDED project. For non-research personal data, beneficiaries in the project are requested to fill out a consent form regarding the use of photo or video material featuring still or animated images of individuals – in conformity with data protection regulation.

Metadata:

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

To be determined on a case-by-case basis. Generally, where data is made available, it will contain the information needed in order to access it (we expect this will be simply via zenodo.org).

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

This depends on the policies of the host platform. Once approved for publication as open data, the consortium does not envisage an expiry date. Rather, potential users can determine whether the data is of interest based on its content and age.

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

Code generation is not required or expected as part of the project. We expect any data that is generated to be readable with generally available tools (word processors, PDF and image viewers, spreadsheet software)

3.3. MAKING DATA INTEROPERABLE

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

As per section 3.1.

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

No generation of ontologies is planned in this project. Any open data shared will be accompanied by adequate metadata to enable interpretation and reuse of the data.

Will your data include qualified references ¹ to other data (e.g. other data from your project, or datasets from previous research)?

To be determined.

3.4. INCREASE DATA RE-USE

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

For any open data, this will be provided along with the data (probably as part of the metadata) on the chosen repository.

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

This will be decided on a case-by-case basis. Generally, where the consortium does decide to make data available in open access, this will be in the public domain with the most permissive possible reuse licences (while respecting project beneficiary IPRs).

Will the data produced in the project be useable by third parties, in particular after the end of the project?

Yes, for any data published as open data since the purpose of such publication is the use and reuse by other parties. For any other data, that is regulated in the project's Consortium Agreement ("Results").

Will the provenance of the data be thoroughly documented using the appropriate standards?

Yes, where standards exist. Where useable standards are absent, adequate documentation with reasonable effort will be produced such that third parties can interpret and re use the data.

¹ A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: <https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>)

Describe all relevant data quality assurance processes.

Data quality assurance generally rests with the project beneficiary generating the data. For open data, this will be reviewed prior to a publication by a project body to be determined (e.g., WP leaders, individual partners, or as otherwise appropriate)

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

More information on research outputs can be found in NEEDED deliverables *D6.1 Dissemination and communication plan* and *D7.1 Project handbook*. Both are public deliverables and should be automatically available at <https://cordis.europa.eu/project/id/101095754> once they are approved and released by the project's funding agency.

4. OTHER ASPECTS

4.1. OTHER RESEARCH OUTPUTS

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

The management of flight path data, emission inventories and noise and population mapping is a central goal of the project. The management of such outputs is covered through the intellectual property management, regulated in the project's Consortium Agreement and supported and further managed as part of the project's exploitation and IPR management (*T6.3 Exploitation and IPR management*), and documented in the project's technical deliverables.

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

A very large proportion of the project's deliverables are classified as "public" with respect to their dissemination level. They will be available at <https://cordis.europa.eu/project/id/101095754> once they are approved and released by the project's funding agency, as well as on the project's website.

4.2. ALLOCATION OF RESOURCES

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.)?

To be determined. Generally, the generation of data will be a part of the project activities, therefore no specific/additional resourcing is allocated for this. Where curation and preparation of open data is concerned, this will be done using the resources for the beneficiary generating that data, with support from WP leader and/or project coordinator as needed.

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

Who will be responsible for data management in your project?

Each beneficiary manages their own data. Where data is shared to facilitate the project activities, it is the responsibility of each WP leader to keep a good order ("housekeeping") with regards to the data needed for the activities of the WP for which they are responsible. For open data, once it is published to the repository of choice, we will depend on the data management procedures of the repository.

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

Internally with beneficiaries – as per their own policies. For open data, according to the policy of the repository.

4.3. DATA SECURITY

What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

Data security is as per the policies and procedures of each project beneficiary, and particularly those generating or owning the data. Generally, state-of-the-art data security is assumed to be in place. For open data, the consortium relies on the data security provision of the host repository.

Will the data be safely stored in trusted repositories for long term preservation and curation?

Yes, open data, if any, will be stored in a trusted repository (probably zenodo.org).

4.4. ETHICS

Are there, or could there be, any ethics or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

No; ethics screening was done at the project proposal stage and in the grant agreement preparation of the project. No noteworthy ethics issues (including dual-use of project results) are identified.

Will informed consent for data sharing and long term preservation be included in questionnaires dealing with personal data?

Informed consent for personal data where it regards photos or videos is handled through the consent form as part of WP7. For preservation of these data, where project beneficiaries are the data controllers, these are handled in conformity with the data protection statement (as referenced in the consent form).

4.5. OTHER ISSUES

Do you, or will you, make use of other national/funder/sectorial/departmental procedures for data management? If yes, which ones (please list and briefly describe them)?

Not beyond what has been described above on a project level, unless specific beneficiaries wish to do so.

5. CONCLUSIONS

This DMP has disclosed the plans for data management according to best available information at project month M6, following a structured DMP template.

While beneficiaries are encouraged to look for ways to share data with the research community, they are under no obligation to disclose data if this goes against their interests.

6. REFERENCES

- [1] Dr. Tomasz Miksa, 'Data Management Plans'. SBA Research & TU Wien. [Online]. Available: https://www.ffg.at/sites/default/files/dmp_ffg_miksa.pdf
- [2] 'Open Science', Feb. 10, 2023. https://research-and-innovation.ec.europa.eu/strategy/strategy-2020-2024/our-digital-future/open-science_en (accessed July 2023).
- [3] 'FAIR Principles', *GO FAIR*. <https://www.go-fair.org/fair-principles/> (accessed July 2023).
- [4] 'Horizon Europe - Data Management Plan Template'. May 05, 2021. Accessed: July 2023. [Online]. Available: <https://enspire.science/wp-content/uploads/2021/09/Horizon-Europe-Data-Management-Plan-Template.pdf>