D1.2 Data Management Plan Internal D1.2 Data Management Plan D1.2 Data Management Plan Internal 2/ 11

Technical Report · March 2021

DOI: 10.13140/RG.2.2.28216.67842

CITATIONS
0

READS
39

4 authors:

Krassimir Krastev
SAFRAN A.S. group
18 PUBLICATIONS **352** CITATIONS
SEE PROFILE

Nathanaël Muot
AxesSim
26 PUBLICATIONS **18** CITATIONS
SEE PROFILE

Christophe Girard
AxesSim
47 PUBLICATIONS **68** CITATIONS
SEE PROFILE

Jean-Philippe Parmantier
The French Aerospace Lab ONERA
94 PUBLICATIONS **748** CITATIONS
SEE PROFILE

Rev. 1.1

N°doc.:RT/2019/004

Config.: Analyst-PRCS-OUT-RT

N°.:AXS/NMT/2019/038

Internal

TECHNICAL REPORT
D1.2 Data Management Plan

	Signature	*Date*
Prepared by Nathanaël MUOT (AxesSim)	Nathanaël MUOT (AxesSim)	20/02/2019
Verified by Christophe GIRARD (AxesSim)	Christophe GIRARD (AxesSim)	22/02/2019
Verified by Manangement Assistant Aldo Bonsignore (IDS)	Aldo Bonsignore (IDS)	25/02/2019
Approved by Coordinator Jean-Philippe Parmantier (ONERA)		25/02/2019
Authorized by topic Manager Krasimir Krastev (SAFRAN)	*Krassimir Krastev*	27/02/2019

Strasbourg 20/02/2019

nr of annexes	0

KEYWORDS Project, management, data, plan

SUMMARY This document provides the plan for managing the data generated and collected during the project. It covers: a) the handling of research data during and after the project, b) what data will be collected, processed or generated, c) what methodology and standards will be applied, d) whether data will be shared/made open and how and e) how data will be curated and preserved.

CONCLUSIONS This Data Management Plan provides an overview of the data that ANALYST project will produce.

The analysis contained in this report allows anticipating the procedures and infrastructures to be implemented by the ANALYST project to efficiently manage the data it will produce.

Document Evolution		
Revision	*Date*	*Reason for change*
Rev. 1.0 Draft A	22/02/2019	Draft Edition
Rev. 1.0 Draft B	25/02/2019	Review by IDS
Rev. 1.0 Draft C	25/02/2019	Review by ONERA
Rev. 1.0	25/02/2019	First Edition
Rev.1.1	27/02/2019	Second Edition after adding paragraph 2.5 to answer Topic Manager's concern about non-disclosure of data

Document Change Record (Log)		
RNC	*Reference*	*Description of Changes*

CONTENTS

1. INTRODUCTION

Research and innovation projects such as ANALYST usually produce large sets of data. Depending on the discipline, the data could come for example from social science research, laboratory testing, field studies or observations. However, it often remains unclear and uncertain, what will happen with the data after they have been analysed and the project has been finished. Furthermore, a lot of data sets are potentially interesting also for other researchers, but due to the fact that they are either stored on a local server or miss crucial meta-data (or both), their potential value cannot be exploited. Hence, researcher need to think about the data that they will produce at the beginning of the research and this is exactly the purpose of the Data Management Plan (DMP). The purpose of the Data Management Plan (DMP) is to provide an analysis of the main elements of the data management policy that will be used in the ANALYST project and by the consortium with regard to the project research data. The DMP covers the complete research data life cycle. It describes the types of research data that will be generated or collected during the project, the standards that will be used, how the research data will be preserved and what parts of the datasets will be shared for verification or reuse. The DMP is a living document, which will evolve during the lifespan of the project, particularly whenever significant changes arise such as dataset updates or changes in Consortium policies.

2. GENERAL PRINCIPLES

2.1 Standard and metadata

This field will describe suitable standards that will be used to describe the data as well as the metadata of the data sets. In ANALYST, we identify three categories of data sets: (a) cables harness description (b) numerical datasets and (c) other datasets.

The first category is the input datasets of the project. This dataset is not covert by an industrial standard. However the specification of this datasets is defined at the beginning of the project and may also be completed during the project if required.

The second category concerns numeral result. For this kind of data, it is proposed to used Amelet-HDF Open Standard https://github.com/axessim/amelet-hdf and https://www.axessim.fr/docs/amelethdf/1.6.1/. For simple dataset simple CSV file format can be used with explicit headers.

Finally, in the second category comply all the data sets published in other formats (e.g. excel, csv, pdf, txt etc.). These data sets will also by accompanied, is some cases, by metadata.

2.2 Data Sharing

During the lifecycle of the ANALYST project all datasets will be shared between partners and third party with sufficient secured channels. All partners are free to use their own file sharing system.

However, for this propose AxesSim hosts an end-to-end encrypted file sharing software Lufi https://github.com/ldidry/lufi.

This service is available here at following link http://drop.axessim.eu:8080/.

2.3 Archiving and preservation

The ANALYST project database will be designed to remain operational for 5 years after the project end. By the end of the project, the final dataset will be transferred to a dedicated AxesSim FTP repository, except particular mention, which ensures sustainable archiving of the final research data.

2.4 Personal Data protection

For some of the activities to be carried out by the project, it may be necessary to collect basic personal data (e.g. full name, contact details, background), even though the project will avoid collecting such data unless deemed necessary.

Such data will be protected in compliance with the EU's Data Protection Directive 95/46/EC1 aiming at protecting personal data.

All data collection by the project will be done after giving data subjects, full details on the experiments to be conducted and after obtaining signed informed consent forms.

2.5 Non-disclosure of data

Disclosure of data generated by the project is governed by Section 10 of the implementation agreement signed by the Topic Manager and the consortium. This section is reminded hereafter:

"Section 10: Non-disclosure of information

The Recipients hereby undertake, without prejudice to any commitment of non-disclosure under their Relevant Grant Agreement, for a period of ten (10) years after the end of the ITD/IADP Activities

Not to use or disclose Confidential Information of which it is the Recipient, otherwise than for the purpose for which it was disclosed;

Not to disclose Confidential Information to any third party, including subcontractors and Linked Third Parties without the prior written consent by the Disclosing Party;

To ensure that internal distribution of Confidential Information by a Recipient shall take place on a strict need-to-know basis; and

To return to the Disclosing Party on demand all Confidential Information which has been supplied to the Recipients including all copies thereof and to delete all information stored in a machine readable form. If needed for the recording of ongoing obligations, the Recipients may however request to keep a copy for archival purposes only.

The Recipients shall be responsible for the fulfillment of the above obligations on the part of their employees, and sub-contractors and Linked Third Parties subject to prior explicit consent case by case as referred to above and shall ensure that their employees, and sub-contractors and Linked Third Parties remain so obliged, as far as legally possible, during and after the end of the ITD/IADP Activities and/or after the termination of employment or the relevant contract of engagement.

The above mentioned obligations of confidentiality shall not apply for disclosure or use of Confidential Information, if and in so far as the Recipient can show that:

The Confidential Information becomes publicly available by means other than a breach of the Recipient's confidentiality obligations;

The Disclosing Party subsequently informs the Recipient that the Confidential Information is no longer confidential;

The Confidential Information is or has been communicated to the Recipient without any obligation of confidence by a third party who is in lawful possession thereof and under no obligation of confidence to the Disclosing Party;

The Confidential Information, at any time, was developed by the Recipient completely independently of any such disclosure by the Disclosing Party; or

The Recipient shall apply the same degree of care with regard to the Confidential Information disclosed as with its own confidential and/or proprietary information, but in no case less than reasonable care.

It is understood and acknowledged by the Parties that the disclosure of Confidential Information doesn't mean any transfer of or change in the ownership to such information.

Each Party shall promptly advise the other Party in writing of any unauthorized disclosure, misappropriation or misuse by any person of Confidential Information as soon as practicable after it becomes aware of such unauthorized disclosure, misappropriation or misuse.

If any Party becomes aware that it will be required, or is likely to be required, to disclose Confidential Information in order to comply with applicable laws or regulations or with a court or administrative order, it shall, to the extent it is lawfully able to do so, prior to any such disclosure notify the Disclosing Party in sufficient time to allow the Disclosing Party to seek an order for protective relief, and comply with the Disclosing Party's reasonable instructions to protect the confidentiality of the information.

The confidentiality obligations under this Implementation Agreement and the Relevant Grant Agreements shall not prevent the communication of Confidential Information to the CSJU, provided that the Disclosing Party is informed of such a communication beforehand and is given an opportunity to disclose it by itself to the CSJU and / or take every measure to control the extent of such a communication."

3. DATA SET DESCRIPTION

In order to provide an overview of the various data sets that are currently available and will be produced in the ANALYST project, the following table shows the data type, the origin of the data, the related WP number and the format, in which the data will be presumably stored.

Table 1 Data sets overview

#	Data type	Origin	WP	Format
1	Test case definition (ANALYST ToolBox input data file)	SAFRAN	2	.xlsx + .csv
2	Numerical results (ANALYST ToolBox output data file)	Consortium	3 & 5	.csv + .ah5
3	Project deliverables	Consortium	1	.docx + .pdf
4				
5				

Table 1 describes the data set and the purpose of the data collection of data generation in relation with the objectives of the project. Additionally, it shows the data utility for clarifying to whom the data might be useful.

Legend:

- xlsx: Microsoft excel file
- csv: standard table format
- ah5: AMELEThdf5
- docx: Microsoft word format
- pdf: Adobe acrobat reader files

3.1 DataSet 1

DS1 Test case definition (ANALYST ToolBox input data file)	
Data identification	
Dataset description	This dataset contains the preliminary canonical test cases and representative tests case
Source	SAFRAN
Partners activities and responsibilities	
Partner owner of the data; copyright holder (if applicable)	SAFRAN
Partner in charge of the data collection	ONERA
Partner in charge of the data analysis	ONERA
Partner in charge of the data storage	ONERA
Related WP(s) and task(s)	2
Standards	
Info about metadata (production and storage dates, places) and documentation?	N/A
Standards, format, estimated volume of data	This data set format is specified by SAFRAN in collaboration with ANALYST partners. This file format use standard file format (.xlsx and .csv). The final file format will be describe in D2.2
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	This dataset will be used during the project as specification, requirement and assessment.
Data access policy / Dissemination level : confidential (only for members of the Consortium and the Commission Services) or Public	As it implies personal data, the access to the dataset is restricted to the ANALYST consortium.
Data sharing, re-use, distribution, publication (How?)	None
Embargo periods (if any)	None
Personal data protection: are they personal data? If so, have you gained (written) consent from data subjects to collect this information?	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	

3.2 DataSet 2

DS2 Numerical results (ANALYST ToolBox output data file)	
Data identification	
Dataset description	This dataset contains the intermediate and final numerical result used for validation and assessment.
Source	Produce during the project
Partners activities and responsibilities	
Partner owner of the data; copyright holder (if applicable)	
Partner in charge of the data collection	ONERA
Partner in charge of the data analysis	ONERA
Partner in charge of the data storage	AxesSim
Related WP(s) and task(s)	2 & 5
Standards	
Info about metadata (production and storage dates, places) and documentation?	N/A
Standards, format, estimated volume of data	This dataset can be imported from, and exported to a CSV, TXT, Excel file or Amelet-HDF.
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	This dataset is used to project's validation and assessment. It could be also used for disseminate and communication about the ANALYST results
Data access policy / Dissemination level : confidential (only for members of the Consortium and the Commission Services) or Public	As it implies personal data, the access to the dataset is restricted to the ANALYST consortium. In fact, the data use for dissemination and communication are in open access (D6.1 Analyst Dissemination, Communication and Exploitation Plan)
Data sharing, re-use, distribution, publication (How?)	This dataset could be published by indirect ways (scientific publication, public communication etc) by curve plots, results analysis, but in any way in initial raw data.
Embargo periods (if any)	None
Personal data protection: are they personal data? If so, have you gained (written) consent from data subjects to collect this information?	None
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	

3.3 DataSet 3

DS3 Project deliverables	
Data identification	
Dataset description	The deliverables of the project.
Source	Generated by WP leaders.
Partners activities and responsibilities	
Partner owner of the data; copyright holder (if applicable)	
Partner in charge of the data collection	ONERA
Partner in charge of the data analysis	ONERA
Partner in charge of the data storage	EC
Related WP(s) and task(s)	WP1
Standards	
Info about metadata (production and storage dates, places) and documentation?	N/A
Standards, format, estimated volume of data	This dataset is a combination of WORD/PDF documents.
Data exploitation and sharing	
Data exploitation (purpose/use of the data analysis)	This dataset presents the outcomes of the project.
Data access policy / Dissemination level : confidential (only for members of the Consortium and the Commission Services) or Public	This dataset does not contain confidential information. Therefore, access to the dataset is public (except the financial information).
Data sharing, re-use, distribution, publication (How?)	None
Embargo periods (if any)	None
Personal data protection: are they personal data? If so, have you gained (written) consent from data subjects to collect this information?	The dataset contains personal data: names of people included in the attendee list of the workshops.
Archiving and preservation (including storage and backup)	
Data storage (including backup): where? For how long?	SIGMA tool of the EC

4. REFERENCES

European Commission (version 3 2016), Guidelines on Data Management in Horizon 2020. Available at

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf (accessed 2019)