

SPARK

D6.1

DATA MANAGEMENT
PLAN

Approval Status

	NAME AND SURNAME	ROLE IN THE PROJECT	PARTNER
AUTHOR(S)	Niccolò Becattini Giandomenico Caruso Daniela De Lucia Aurelien Leclerc Frédérique Segond	Assistant PC Researcher Assistant for Coordination Responsible for DMP Deputy PC	PoliMI PoliMI PoliMI Viseo Viseo
REVIEWED BY	Giandomenico Caruso Gaetano Cascini	Researcher PC	PoliMI PoliMI
APPROVED BY	Gaetano Cascini	PC	PoliMI

History of Changes

VERSION	DATE	DESCRIPTION OF CHANGES	By
01	06.03.2016	Initial Draft of the document	Daniela De Lucia
02	20.04.2016	Implementation of the analysis of the data to be made open access	Niccolò Becattini
03	27.04.2016	Revision of the initial approach to the DMP	Giandomenico Caruso Gaetano Cascini
04	19.05.2016	Implementation of the policy	Niccolò Becattini Giandomenico Caruso Gaetano Cascini
05	01.06.2016	Revision of some sections of the document	Niccolò Becattini Giandomenico Caruso Daniela De Lucia
06	24.06.2016	Revision of the document, some table title added	Aurelien Leclerc
07	26.06.2016	Revision of the content	Gaetano Cascini



08	28.06.2016	Implementation of the description of the general approach	Giandomenico Caruso
09	29.06.2016	Editing, formatting and writing up of the executive summary and of the conclusions	Daniela De Lucia
10	29.06.2016	Adding "projected data size"	Aurelien Leclerc
11	30.06.2016	Added Codendi for Private data management	Aurelien Leclerc Frédérique Segond
12	30.06.2016	Final review and approval	Gaetano Cascini

Document Details

DISSEMINATION LEVEL	PUBLIC
DUE DATE	30.06.2016
ISSUE DATE	30.06.2016
CONTRACT NUMBER	H2020-ICT/2015-688417
ELECTRONIC FILE LOCATION	www.spark-project.net/deliverables
FILE NAME	D6.1_WP6_Data Management Plan_v1



TABLE OF CONTENTS

1. Executive Summary	5
2. Introduction	5
2.1. Scope of the activities and of the deliverable	5
3. the data In spark	5
4. Focus on each WP	10
4.1. WP1: CHARACTERIZATION OF USERS' NEEDS AND EXPECTATIONS	11
4.2. WP2 - DEVELOPMENT OF SPARK MODULES	13
4.3. WP3 – DEVELOPMENT AND TESTS OF SPARK PLATFORM	15
4.4. WP4 – TEST AND VALIDATION IN RELEVANT ENVIRONMENT	17
4.5. WP5 – VALIDATION AND DEMONSTRATIONS IN REAL OPERATIONAL ENVIRONMENT	20
5. Archiving and Storage of the Data	24
5.1. Public data	24
5.2. Private data	24
5.3. Format of the data to be archived and published in ZENODO.	24
6. PROPOSED POLICY	24
7. Conclusion	25

1. EXECUTIVE SUMMARY

The present deliverable is the result of the initial consultation among the partners for what concerns the Data Management Plan. The activities started soon at the beginning of the project and the issues related to storage and sharing of the project data (both generated, processed and collected) have been carefully analysed. The plan that will be presented in the following gives a general overview for the project as a whole and per each WP.

The Consortium knows that as the project will evolve, also the policy to apply for the data management will have to be updated and/or revised.

2. INTRODUCTION

2.1. SCOPE OF THE ACTIVITIES AND OF THE DELIVERABLE

This document is an initial overview on the approach the SPARK Consortium will use for the Management of the data.

It is important to highlight here that this approach will be further detailed and improved along the project evolution and two updates of the deliverable are foreseen at M21 and at M36.

The deliverable falls under the headings of the WP6 and in particular within the context of T6.1. The aim is to manage all the issues related to the availability of the data collected and generated and processed during the SPARK project lifecycle.

3. THE DATA IN SPARK

SPARK will deal with two main types of data:

- (Raw) Gathered Data;
- Processed/Generated Data.

Moreover, two main perspectives have been considered in order to define a plan for the management of data. On the one hand, data can be characterized in terms of sensitivity whenever they deal with personal data (of a person or an external entity with respect to the



consortium – e.g. a End User’s customer participating a creative design session). On the other hand, data can be also classified consistently with the overall purpose they play along the project activities. These two complementary classifications are not mutually exclusive and provide additional details to understand how to deal with the data generated along the project. This organization allows defining a data management policy that appropriately balances between the need of keeping confidential what is critical for the development of the SPARK platform and the publication of data in adequate repositories (e.g.: it supports decisions concerning the embargo duration before making data publicly available).

The two above-mentioned additional classes can be detailed as:

- **Sensitivity**
 - data could be sensitive because of confidential issues that goes beyond the consortium (because of 3rd parties rights);
 - data that are sensitive because of the ethical issues;
 - data that are not sensitive.
- **Project strategy**
 - data that are relevant for the business/market exploitation of the SPARK platform;
 - data that are relevant for what concerns the scientific objectives of the project. In particular:
 - Analysis of the dynamics of co-creative processes of teams dealing with digital and physical prototypes;
 - Development of the SAR-based responsive ICT platform, i.e. the SPARK platform;
 - Study and analysis of how and to what extent the SAR technology can stimulate and enhance design creativity through a comparison against pre-defined metrics in real operational design environments.

In the following, a matrix is provided to show how the WP leader will have to cluster the different data gathered or processed in the relevant WP, once agreed by the consortium.

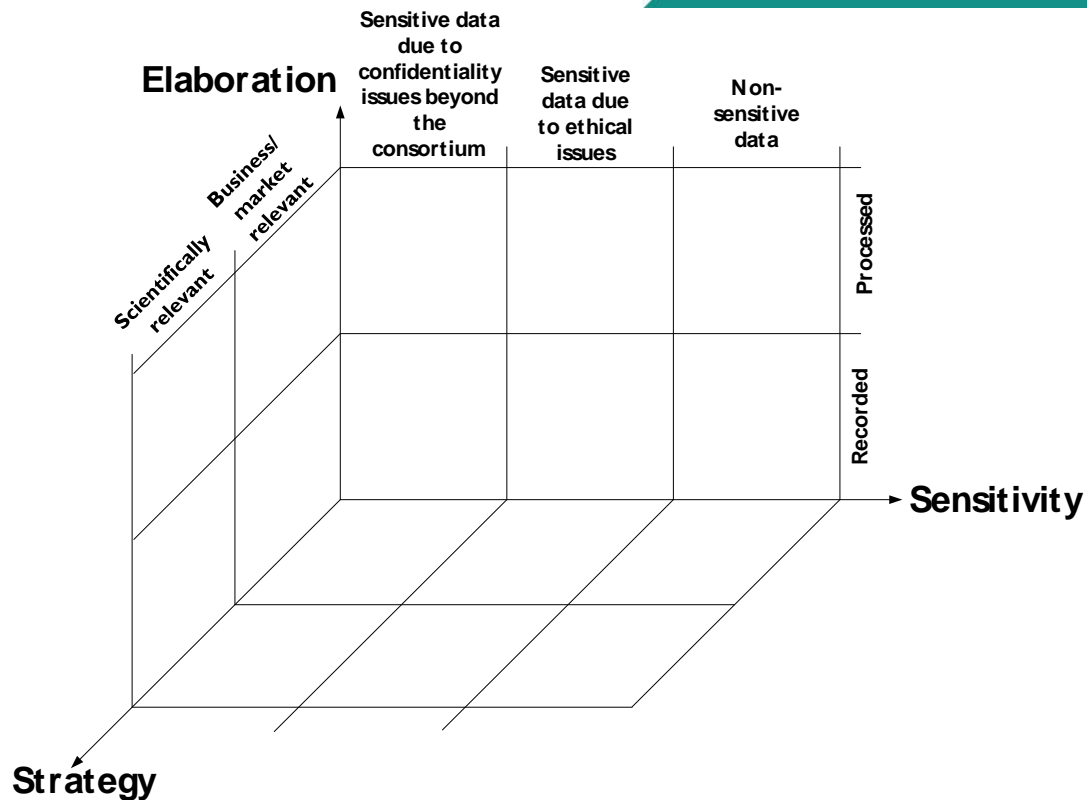


Figure 1. The three dimensions to determine the management of data

In general, whatever will be the specific classification of data according to the above presented categories, in the end they could be classified as:

- Confidential;
- Temporarily confidential (due to the expiration of the NDAs or before anonymization);
- Not confidential.

A suggested policy for the dissemination plan has been initially defined in this document for each outcome of the WP activities. This policy suggests how the publication of the data will be managed during the project.

Other descriptors should finally have identified to characterise the data from the technical point of view. These descriptors are:

- Format: type and corresponding extension that identify the file containing the data
- Medium of data: physical / virtual
- Projected volume: hypothesis made on the number of each data
- Data reading: the tool used to read the data in a specific format
- Metadata: information that further identifies the data and that can be used to manage (storing and searching) the data within a database.

Table 1 provides examples of the possible data readings while Table 2 provides the structure of the metadata (Id and Format) and some application examples related to the file used so far. Data reading and metadata should be the same for all work packages, since they are independent from the origin of the data.

Table 1: Examples of possible data readings

DATA READING		
FORMAT	MOST USED TOOL(S)	SIZE RANGE
Text / .doc	MS Word, Open Office Writer	0..10 Mo
Spreadsheet / .xls	MS Excel, Open Office Calc	0..2 Mo
Recordings	Player: VLC	0..N Go
Translation / .srt	Subtitles editor: notepad++, notepad	0..2 Mo

Table 2: Structure of the metadata and examples

DATA FORMAT	METADATA			
	ID	FORMAT	EXAMPLE	COMMENT
Text / .doc & Spreadsheet / .xls	File / Title	SPARK_[wkp]_[doc]_[version]	SPARK_WKP3_SPARK-Platform-Architecture_v1.2	-
	Subject	Text	Spark Platform Architecture	-
	Author	[company]_name_surname	WISEO_DURAND_Pierre	-
	Responsible	[company]_name_surname	WISEO_MARTIN_Arthur	Actor in charge of the document production
	Company	Text	WISEO	-
	Diffusion	yyyy-mm-dd	2016-06-23	Aimed date for data diffusion
Recordings	File	SPARK_[wkp]_[doc]_[yyyy-mm-dd]	SPARK_WKP1_Co-creative-design_2016-07-15	-
Translation / .srt	File	SPARK_[wkp]_[doc]_[yyyy-mm-dd]	-	Same as aimed record

Finally, Figure 2 shows a diagram that summarises the expected outcomes deriving from the all the task of the project. The diagram makes easier the identification and the clustering of the type of data for each WP.

SPARK Activities concurring to the achievement of the SPARK objectives (#1; #2; #3; #4)

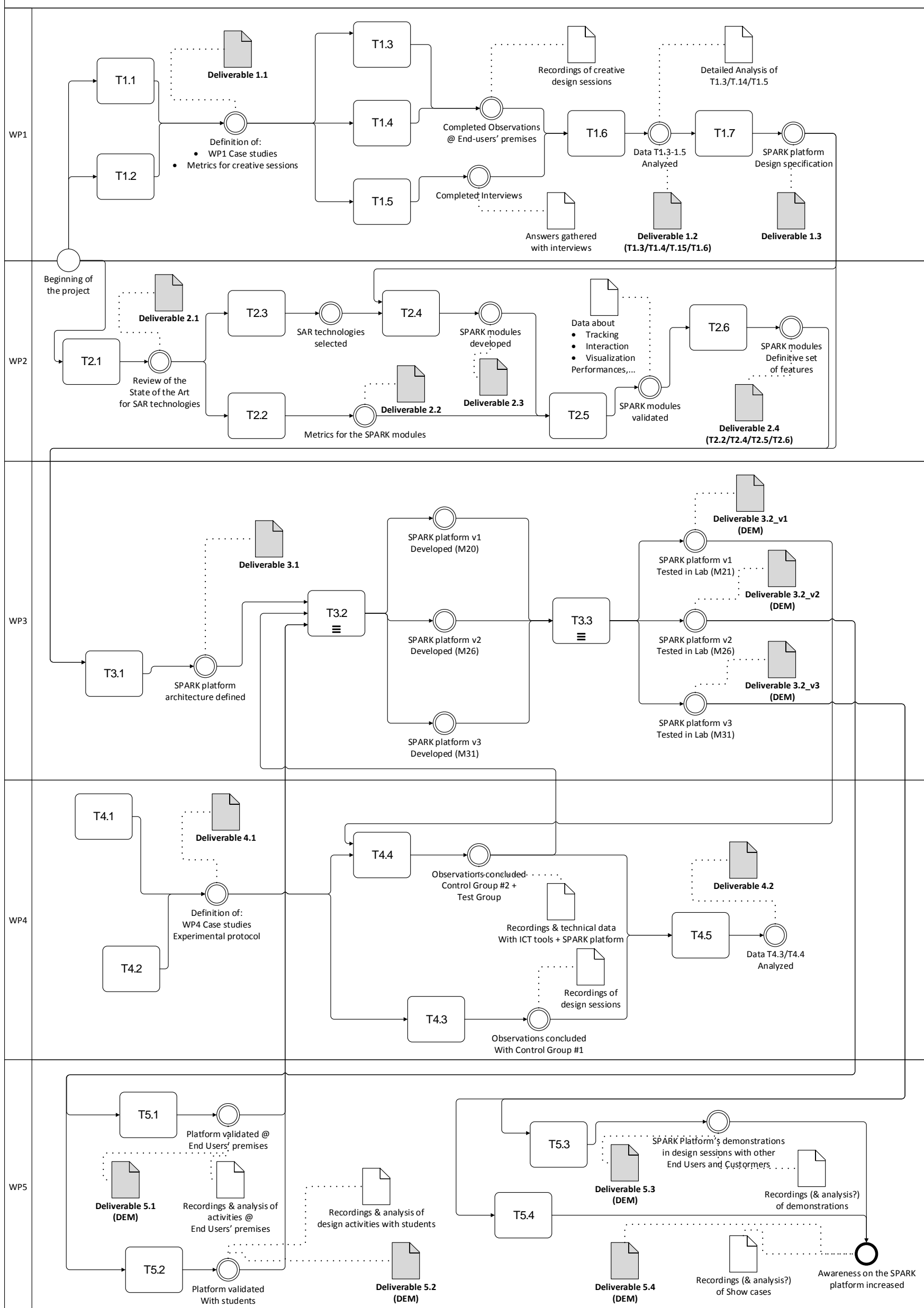


Figure 2: Summary of the outcomes for each Task (WP1-WP5).

4. FOCUS ON EACH WP

This section proposes the characterisation of the data for each WP. In particular, the following paragraphs report the part of the diagram related to a specific WP, as shown in Figure X and a table where the data are classified according to the criteria described in the previous section i.e.:

- Origin of data (type specified)
 - Gathered data
 - Processed / generated data
- Sensitivity
 - Because of 3rd parties' rights
 - Because of ethical issues
 - Non sensitive data
- Project strategy
 - Business / Commercialization
 - Scientifically relevant
- Suggested policy
- Description
 - Format
 - Medium of data
 - Projected volume
 - Data reading (see below)
 - Metadata (see below)



4.1.WP1: CHARACTERIZATION OF USERS' NEEDS AND EXPECTATIONS

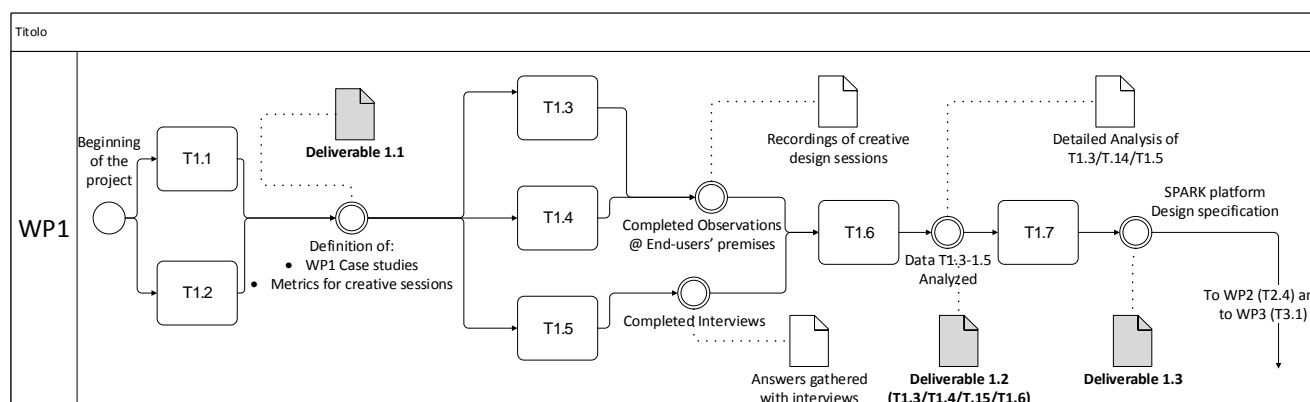


Figure 3: Expected outcomes of WP1

Table 3: Type of data and classification for WP1

	Origin of data (Type specified)		Sensitivity			Project strategy		Suggested Policy	Description		
	Gathered data	Processed / Generated data	Because of 3rd parties rights	Because of ethical issues	Non sensitive data	Business / commercialization	Scientifically relevant		Format	Medium of data	Projected volume
WP1											
Task 1.1		Case studies	X					Data to be made public according to the indication of responsible partners	Text / .doc	Virtual	1 per case study
Task 1.2	Existing metrics				X		X	To be published after the first paper	Spreadsheet / .xls	Virtual	1
Task 1.2		Co-creative design metrics			X		X	To be published after the first paper	Spreadsheet / .xls	Virtual	1



Task 1.3	Recordings		X	X			X	Data to be made public according to the indication of responsible partners	Video files	Virtual	1 per case study
Task 1.3		Transcripts/translation	X	X			X	To be published open access after anonymization, after the first paper	Text / .doc / .srt	Virtual	1 per case study
Task 1.3		Coded design protocols	X	X			X	To be published after the first paper	Text / .doc	Virtual	1
Task 1.3		Analysis of design protocols			X		X	To be published after the first paper	Text / .doc	Virtual	1
Task 1.4	Recordings		X	X			X	Data to be made public according to the indication of responsible partners	Video files	Virtual	1 per case study
Task 1.4		Transcripts/translation	X	X			X	To be published open access after anonymization, after the first paper	Text / .doc / .srt	Virtual	1 per case study
Task 1.4		Coded design protocols	X	X			X	To be published after the first paper	Text / .doc	Virtual	1
Task 1.4		Analysis of design protocols			X		X	To be published after the first paper	Text / .doc	Virtual	1
Task 1.5	Interviews		X	X			X	To be published after the first paper	Text / .doc	Virtual	1 per interview
Task 1.5		Summary of the results			X		X	To be published after the first paper	Text / .doc	Virtual	1
Task 1.6		Combined analysis of needs and expectations of End Users			X	X	X	To be published after the first paper (the deliverable D1.2 is, however, public)	Text / .doc	Virtual	1
Task 1.7		SPARK design specification			X	X	X	Ready for Open Access being the deliverable D1.3 public	Text / .doc	Virtual	1



4.2.WP2 - DEVELOPMENT OF SPARK MODULES

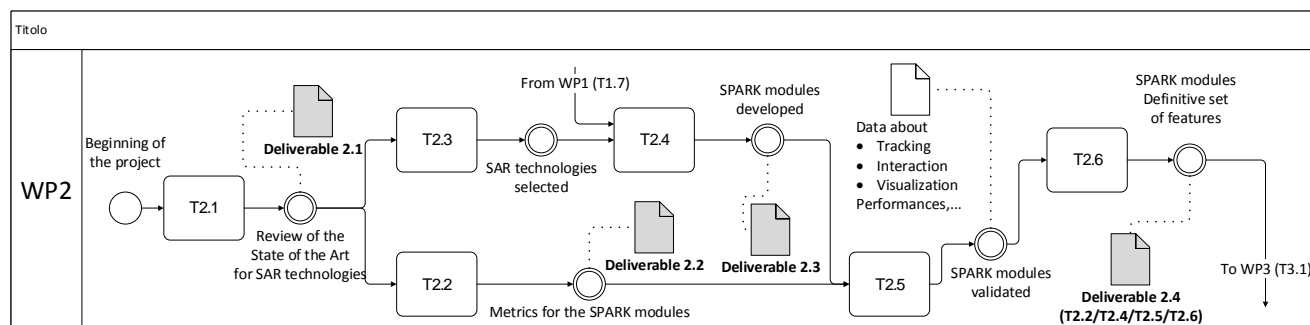


Figure 4: Expected outcomes of WP2

Table 5: Type of data and classification for WP2

	Origin of data (Type specified)		Sensitivity			Project strategy		Suggested Policy	Description		
	Gathered data	Processed / Generated data	Because of 3rd parties rights	Because of ethical issues	Non sensitive data	Business / commercialization	Scientifically relevant		Format	Medium of data	Projected volume
WP2											
Task 2.1	Literature review				X		X	Ready for Open Access being the deliverable D2.1 public	Text / .doc	Virtual	1..N
Task 2.1	Hardware specifications				X		X	Ready for Open Access being the deliverable D2.1 public	Text / .doc	Virtual	1..N

Task 2.2		Modules evaluation metrics			X		X	To be published after the first paper related to the tests in T2.5	Spreadsheet / .xls	Virtual	1
Task 2.3		Hardware selection			X	X		To be published at the beginning of the T5.3/5.4	Text / .doc	Virtual	1
Task 2.4		Modules prototype description			X	X	X	To be published after the first publication or after the exploitation strategy benefits of them, whatever comes later	Text / .doc	Virtual	1
Task 2.5	Set of answers to questionnaires			X			X	To be published after anonymization after the first paper	Spreadsheet / .xls	Virtual	1
Task 2.5	Modules evaluation data				X		X	To be published after the first paper	Spreadsheet / .xls	Virtual	1
Task 2.5		Technological benchmark			X	X	X	To be published after the first publication or after the exploitation strategy benefits of them, whatever comes later	Text / .doc	Virtual	1
Task 2.6		SPARK platform features			X	X		To be published after the first release of the integrated SPARK platform	Text / .doc	Virtual	1



4.3.WP3 – DEVELOPMENT AND TESTS OF SPARK PLATFORM

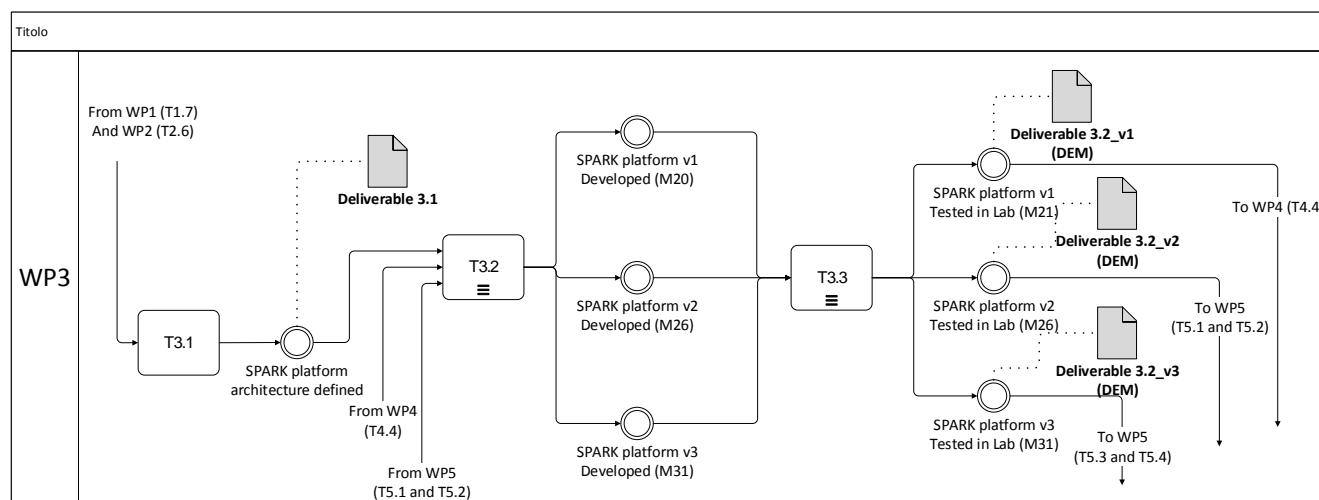


Figure 5: Expected outcomes of WP3

Table 6: Type of data and classification for WP3

	Origin of data (Type specified)		Sensitivity			Project strategy		Suggested Policy	Description		
	Gathered data	Processed / Generated data	Because of 3rd parties rights	Because of ethical issues	Non sensitive data	Business / commercialization	Scientifically relevant		Format	Medium of data	Projected volume
WP3											
Task 3.1		SPARK platform architecture			X	X		To be published after the release of the Deliverable D3.1	Text / .doc	Virtual	1
Task 3.2		Description of SPARK platform versions			X	X	X	To be published after the end of the project	Source code	Virtual	1..N
Task 3.3	Set of answers to questionnaires			X			X	To be published after the first paper	Spreadsheet / .xls	Virtual	1
Task 3.3	SPARK platform evaluation data				X	X	X	To be published after the first paper	Text / .doc and / or Spreadsheet / .xls	Virtual	1



4.4. WP4 – TEST AND VALIDATION IN RELEVANT ENVIRONMENT

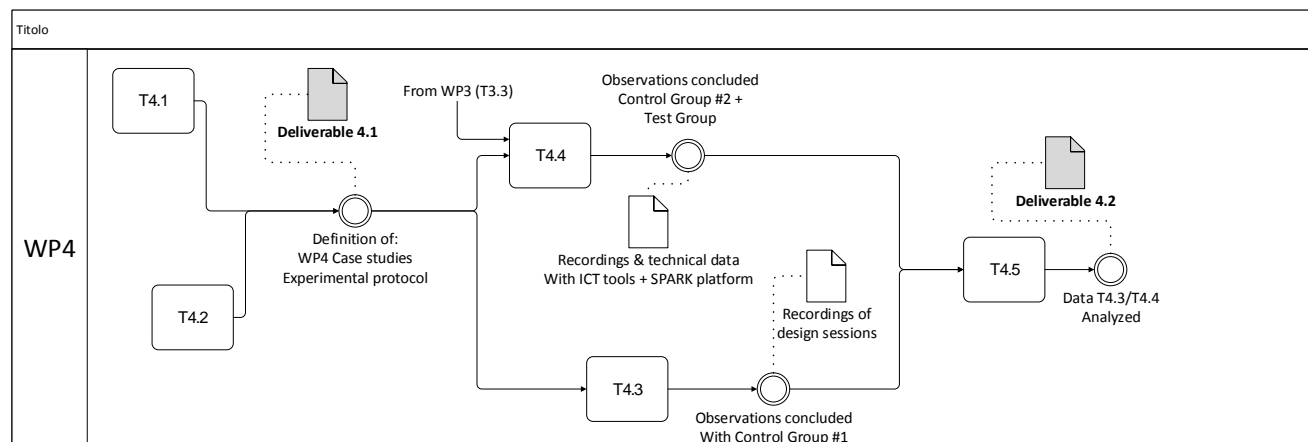


Figure 6: Expected outcomes of WP4

Table 7: Type of data and classification for WP4

	Origin of data (Type specified)		Sensitivity			Project strategy		Suggested Policy	Description		
	Gathered data	Processed / Generated data	Because of 3rd parties rights	Because of ethical issues	Non sensitive data	Business / commercialization	Scientificall y relevant		Format	Medium of data	Projected volume
WP4											
Task 4.1		Case studies	X			X		To be published after the end of the embargo period defined in the NDAs	Text / .doc	Virtual	1
Task 4.2		Testing protocol			X		X	To be published after the first paper	Text / .doc	Virtual	1



Task 4.3	Recordings		X	X			X	Data to be made public according to the indication of responsible partners	Video files	Virtual	1 per case study
Task 4.3		Transcripts	X	X			X	To be published open access after anonymization, after the first paper	Text / .doc / .srt	Virtual	1 per case study
Task 4.3		Coded design protocols	X	X			X	To be published after the first paper	Text / .doc	Virtual	1
Task 4.3		Analysis of design protocols			X		X	To be published after the first paper	Text / .doc	Virtual	1
Task 4.4	Recordings		X	X			X	Data to be made public according to the indication of responsible partners	Video files	Virtual	1 per case study
Task 4.4		Transcripts	X	X			X	To be published open access after anonymization, after the first paper	Text / .doc / .srt	Virtual	1 per case study
Task 4.4		Coded design protocols	X	X			X	To be published open access after anonymization, after the first paper	Text / .doc	Virtual	1
Task 4.4		Analysis of design protocols			X		X	To be published after the first paper	Text / .doc	Virtual	1



Task 4.4		SAR module performances of the SPARK platform			X	X	X	To be published after the first publication or after the exploitation strategy benefits of them, whatever comes later	Spreadsheet / .xls	Virtual	1
Task 4.5		Information MGMT system performances of the SPARK platform			X	X	X	To be published after the first publication or after the exploitation strategy benefits of them, whatever comes later	Spreadsheet / .xls	Virtual	1
Task 4.5		Comparison of outcomes of design sessions run with and without the SPARK platform			X	X	X	To be published after the first publication or after the exploitation strategy benefits of them, whatever comes later	Text / .doc	Virtual	1



4.5.WP5 – VALIDATION AND DEMONSTRATIONS IN REAL OPERATIONAL ENVIRONMENT

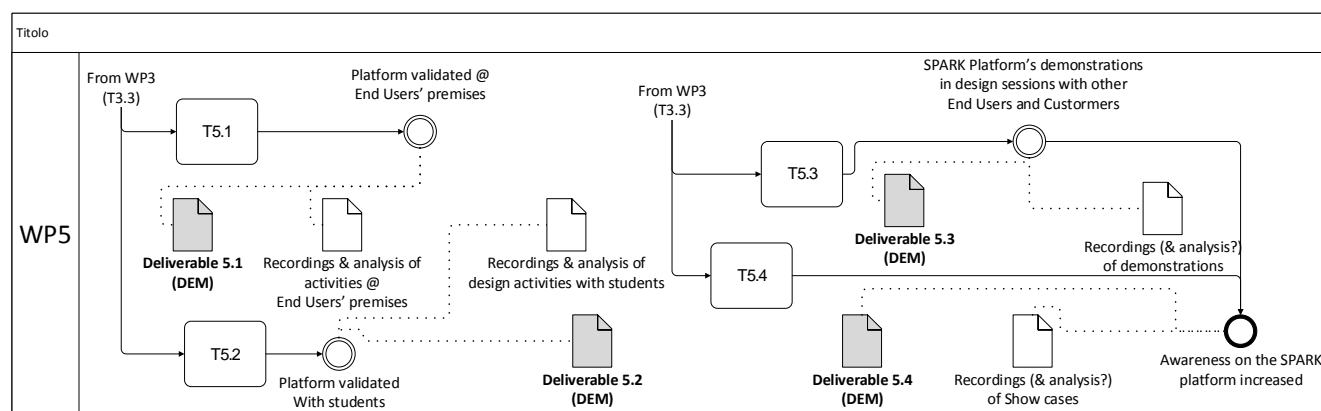


Figure 8: Expected outcomes of WP5

Table 9: Type of data and classification for WP5

	Origin of data (Type specified)		Sensitivity			Project strategy		Suggested Policy	Description		
	Gathered data	Processed / Generated data	Because of 3rd parties rights	Because of ethical issues	Non sensitive data	Business / commercialization	Scientifically relevant		Format	Medium of data	Projected volume
WP5											
Task 5.1	Recordings		X	X			X	To be published after the end of the embargo period defined in the NDAs	Video files	Virtual	1 per case study



Task 5.1		Transcripts	X	X			X	To be published open access after anonymization, after the first paper	Text / .doc / .srt	Virtual	1 per case study
Task 5.1		Coded design protocols	X	X			X	To be published after the first paper	Text / .doc	Virtual	1
Task 5.1		Analysis of design protocols	X	X			X	To be published after the first paper	Text / .doc	Virtual	1
Task 5.1		SAR module performances of the SPARK platform			X	X	X	To be published after the first publication or after the exploitation strategy benefits of them, whatever comes later	Spreadsheet / .xls	Virtual	1
Task 5.1		Information MGMT system performances of the SPARK platform			X	X	X	To be published after the first publication or after the exploitation strategy benefits of them, whatever comes later	Spreadsheet / .xls	Virtual	1
Task 5.2	Recordings			X			X	To be published open access after anonymization, after the first paper	Video files	Virtual	1 per case study
Task 5.2		Transcripts	X	X			X	To be published open access after anonymization, after the first paper	Text / .doc / .srt	Virtual	1 per case study
Task 5.2		Coded design protocols	X	X			X	To be published after the first paper	Text / .doc	Virtual	1



Task 5.2		Analysis of design protocols			X		X	To be published after the first paper	Text / .doc	Virtual	1
Task 5.2		SAR module performances of the SPARK platform			X	X	X	To be published after the first publication or after the exploitation strategy benefits of them, whatever comes later	Spreadsheet / .xls	Virtual	1
Task 5.2		Information MGMT system performances of the SPARK platform			X	X	X	To be published after the first publication or after the exploitation strategy benefits of them, whatever comes later	Spreadsheet / .xls	Virtual	1
Task 5.3	Recordings		X	X			X	Data to be made public according to the indication of responsible partners	Video files	Virtual	1 per case study
Task 5.3		Transcripts	X	X			X	To be published open access after anonymization, after the first paper has been published	Text / .doc / .srt	Virtual	1 per case study
Task 5.3		Coded design protocols	X	X			X	To be published after the first paper	Text / .doc	Virtual	1
Task 5.3		Analysis of design protocols			X		X	To be published after the first paper	Text / .doc	Virtual	1
Task 5.3		SAR module performances of the SPARK platform			X	X	X	To be published after the first publication or after the exploitation strategy benefits of them, whatever comes later	Spreadsheet / .xls	Virtual	1



Task 5.3		Information MGMT system performances of the SPARK platform			X	X	X	To be published after the first publication or after the exploitation strategy benefits of them, whatever comes later	Spreadsheet / .xls	Virtual	1
Task 5.4	Recordings		X	X			X	Data to be made public according to the indication of responsible partners	Video files	Virtual	1 per case study
Task 5.4		Transcripts	X	X			X	To be published open access after anonymization, after the first paper	Text / .doc / .srt	Virtual	1 per case study
Task 5.4		Coded design protocols	X	X			X	To be published after the first paper	Text / .doc	Virtual	1
Task 5.4		Analysis of design protocols			X		X	To be published after the first paper	Text / .doc	Virtual	1
Task 5.4		SAR module performances of the SPARK platform			X	X	X	To be published after the first publication or after the exploitation strategy benefits of them, whatever comes later	Spreadsheet / .xls	Virtual	1
Task 5.4		Information MGMT system performances of the SPARK platform			X	X	X	To be published after the first publication or after the exploitation strategy benefits of them, whatever comes later	Spreadsheet / .xls	Virtual	1



5. ARCHIVING AND STORAGE OF THE DATA

5.1. PUBLIC DATA

The public data will be made available through the ZENODO platform, which will ensure the publication of such data also on OPENAIRE (<http://zenodo.org/about>).

5.2. PRIVATE DATA

All the project gathered, generated and processed data and documents will be archived on a web-repository managed by Viseo: CODENDI, which will ensure the secure and safe storage of the data.

Codendi is an ALM (Application Lifecycle Management) which permits a fine tuning of access rights.

However, video recordings of the design sessions are likely to imply very large files (several TBs per session) that can hardly be managed via a standard Internet connection. Therefore, all the original video recordings will be stored in appropriate data storage devices and at least two copies of each file will be stored in at least two different locations (at PoliMI and at GINP) in order to reduce the risks of losing data. All the devices shall be kept with by the involved organizations with the maximum care they dedicate to the storage of sensible data.

All private data will be stored for at least 5 years.

5.3. FORMAT OF THE DATA TO BE ARCHIVED AND PUBLISHED IN ZENODO.

The managed data will concern recordings of design sessions (with downsized image resolution for reducing the size of the files), documents (Text files; Spreadsheets; Presentations) as well as data acquired to assess the technical performance of the system. All these data will be archived using standard file format, i.e. files that the most diffused editors, readers, office suites can handle.

6. PROPOSED POLICY

All the data that are relevant and strategic from a scientific point of view will be made available after the accomplishment of the analysis and the issue of the first scientific publications.

For what concerns the data gathered during the testing activities, they will be initially anonymised, so that also the ethical issues could be respected.

Data that are not confidential and that are not strategic for the business and market will be made available at latest after 3 years from the collection or processing.

As the project cannot be considered as a fixed element, and therefore changes and evolutions can occur during the whole project lifecycle, the Steering Committee can introduce modifications to the current classification as soon as these modifications are necessary in order to avoid problems related to confidential issues. The modifications will be registered in the minutes of the meeting and will be integrated in the next versions of the DMP.

7. CONCLUSION

This document is the first version of the Data Management Plan as it has been conceived by the SPARK Consortium, so as to have a path that will be used for the Management of the data that will be collected, generated and/or processed during the project lifecycle. This deliverable is the result of the activities of task 6.1 (WP6) that have carefully analysed all the possibilities related to the confidentiality of the data and the importance of these data from a strategic point of view. In addition, ethical issues related to the sensitivity of data collected during the testing activities have been taken into account.

Eventually, the modalities for the data archiving, storage and safeguard of the data have been addressed.

All the data that can be made openly accessible will be uploaded by means of Zenodo.

Finally, the approach described in this deliverable will be further detailed and improved along the project evolution and specifically at M21 and at M36.

