Horizon2020/EURO-6-2015

Coordination and Support Actions

SOcietal Needs aNalysis and Emerging Technologies in the public Sector

Deliverable D1.2 The SONNETS Research Data Management Plan

Workpackage	WP1 Project Management
Editor(s):	Nuria Rodríguez
Responsible Partner:	ATOS
Internal Reviewer	Operations Committee (OC)
Status-Version:	V1.0
Date:	M6
EC Distribution:	PU
Abstract:	The SONNETS project participates in the Pilot on Open Research Data launched by the European Commission along with the Horizon 2020 programme. Hence, this document describes the data management policy that will be used by the consortium with regard to all the datasets that will be generated by the project.

Document Revision History

Version	Date	Modifications Introduced	
		Modification Reason	**Modified by**
V0.1	03/08/2016	First version	ATOS
V1.0	08/08/2016	Final version after internal review by OC	ATOS

SONNETS

Contents

List of Tables

Definitions, Acronyms and Abbreviations

Acronym	Title
CSA	Coordination and Support Action
DMP	Data Management Plan
H2020	Horizon 2020, EU funding Strategy for 2014 - 2020
MP3	MPEG Audio Layer III
OC	Operations Committee
PDF	Portable Document Format
WP	Workpackage

Table 1: Definitions, Acronyms and Abbreviations

Executive Summary

The SONNETS project participates in the Pilot on Open Research Data launched by the European Commission along with the Horizon 2020 programme. The main purpose of a Data Management Plan (DMP) is to describe Research Data with the metadata attached to make them discoverable, accessible, assessable, usable beyond the original purpose and exchangeable between researchers.

SONNETS general strategy for data management follows the Guidelines on Data Management in Horizon 2020 and will specify the types of data to be generated and collected along with their respective identification and classification methods, the standards and metadata to be used, the corresponding data exploitation and availability perspectives as well as the data sharing and preservation schemes to be used. The DMP will be implemented through the Task 1.3 lead by ATOS.

The SONNETS project aims at accelerating the transformation of the public sector through the identification, analysis and take-up of emerging technologies that hold the potential to transform and infuse real value to public services. In order to achieve this goal, the SONNETS consortium will carry desk-based research, workshops and interviews, make use of data discovery tools, construct a wide network of collaborators, engage a group of renowned experts and develop a knowledge repository, which will produce a huge amount of data. Hence the importance for SONNETS project to define a data management policy.

SONNETS involves primary data collection: (1) semi-structured interviews; (2) focus group discussions and stakeholder workshops; (3) Communication via SONNETS group. These sources will be the bases for the creation of seven SONNETS datasets:

- Dataset 1: Contact Users' Database
- Dataset 2: Stakeholders' Database
- Dataset 3 Stakeholders' Interviews
- Dataset 4 Stakeholders' Surveys
- Dataset 5 Societal and Public Sector Needs
- Dataset 6 Emerging Technologies
- Dataset 7 SONNETS roadmap and briefs.

These data sets will be handled as editable file (.doc) or Data will be shared in relation to publications (deliverables and papers).

SONNETS project intends to make its research data available for open access and use by other researchers and members of the public. In furtherance of this goal, SONNETS will deposit its research data into a digital repository and make the data available for access and reuse under a Creative Commons Attribution License.

Finally, the SONNETS data will be backed up regularly and stored on the ATOS Institutions server. Furthermore, for potential sensitivities (e.g. personal data of interviewees) of the data being collected, the project has established a system for protecting data while it is being processed.

1 Introduction

1.1 Purpose and scope

This document describes the Data Management Plan (DMP) with respect to the Horizon 2020 – EURO-6-2015 CSA Project SOcietal Needs aNalysis and Emerging Technologies in the public Sector (Grant Agreement No. 692868).

The mission of the SONNETS project, as described in its proposal and subsequent grant agreement, calls on providing an ever-evolving methodological framework for public sector organizations and related stakeholders to accelerate the transformation of the public sector through the identification, analysis and take-up of emerging technologies that hold the potential to transform and infuse real value to public services.

In order to tackle the above-mentioned goal, the SONNETS consortium will carry desk-based research, workshops and interviews, make use of data discovery tools, construct a wide network of collaborators, engage a group of renowned experts and develop a knowledge repository (the SONNETS LinkedIn group, https://www.linkedin.com/groups/8520112). These activities will produce a huge amount of data, hence the importance for SONNETS project to define a data management policy.

The purpose of the DMP is to provide an analysis of the main elements of the data management policy that will be used by the consortium with regard to all the datasets that will be generated by the project.

1.2 Structure

This document provides a description of the Data Management Plan (DMP) within the SONNETS project through the following sections:

- *Section 2* defines the data management strategy to be followed during the SONNETS project according to the Guidelines on Data Management in Horizon 2020 published by the EC services.
- *Section 3* details the information to be included in the SONNETS data tables, listing the different datasets that will be produced by the project, the main exploitation perspectives for each of those datasets, and the major management principles the project will implement to handle those datasets
- *Section 4* presents the main conclusions and finally the *References section* provides information about the documentation used to produce this deliverable.

2 Data Management strategy

The general strategy for data management, according to the Guidelines on Data Management in Horizon 2020[1], will be based on the identification and classification of data generated and collected, standards and metadata to be used, exploitation and availability of data as well as how the data will be shared and archiving preservation of the information.

The SONNETS DMP will cover all the data life cycle. Hence, the task *T1.3 Data Management* in WP1 will be devoted to formulate and continuously evolve the SONNETS research data management plan in accordance with the H2020 guidelines regarding Open Research Data. In Task 1.3, the metadata, procedures and file formats for note-taking, recording, transcribing, storing visual data from participatory techniques, and anonymising semi-structured interview and focus group discussion data will be developed and agreed. ATOS will be the leader for this task, though all partners are involved in the compliance of the DMP.

2.1 Types of data to be generated/collected

Data to be generated/extracted/collected will be obtained with the collaboration of researchers, the external experts (both the Experts Group and Experts Advisory Group) and other collaborators.

2.1.1 Data sources

SONNETS involves **primary data collection**: 1) semi-structured interviews; 2) focus group discussions and stakeholder workshops; 3) Communication via SONNETS group

1. *Semi-structured interviews with individuals*: The team anticipates undertaking 20-25 semi-structured interviews in WP2 and WP3. Data will be collected and stored using digital audio recording (e.g. MP3) where interviewees permit. In case they do not, interviews will be undertaken in pairs to enable detailed note-taking. Interview notes will be typed up according to agreed formats and standards. During data analysis, the data will be accessible only by certified members of the project team. The research project will remove any direct identifiers in the data before sending it to a digital repository.

2. *Focus group discussions and stakeholder workshops*: Focus groups and workshops conducted within WP2 and WP3 will involve two researchers. Whether recorded or not, the event will be transcribed or documented using agreed formats and standards for handling the issue of multiple voices, interruptions, labelling of participatory and visual activities, and so on.

   All transcripts will be in Microsoft Word. All the researchers will be reasonably fluent in both English and the main language in which interviews and focus groups will be conducted, so that transcriptions will be translated into English.

3. *Communication via SONNETS collaboration tool*: The consortium has created a dedicated Web site and a SONNETS LinkedIn Group. On one hand, the project site has been established using a content management

---

[1] Guidelines on Data Management in Horizon 2020, European Commission (2016)
https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

system (Drupal 7) so that data users can participate in adding site content over time in the Blogs section (http://www.sonnets-project.eu/blog) , making the site self-sustaining. For preservation, we will supply periodic copies of the data to ATOS own digital data repository. That repository will be the ultimate home for the data. On the other hand, the consortium agrees to use LinkedIn to manage and distribute the communication data of the SONNETS network. In addition to the research community, we expect these data will be used by practitioners and policymakers.

2.1.2 Type of data

The data to be collected can be:

- **No personal data**: this information is not affected by Data Protection legislation
- **Personal data**: data which relate to an individual who can be identified[2]
  - o (a) from those data, or
  - o (b) from those data and other information which is in the possession of, or is likely to come into the possession of, the data controller, and includes any expression of opinion about the individual and any indication of the intentions of the data controller or any other person in respect of the individual.

In that case, SONNETS follows the European directives:

- Data Protection Directive[3]
- Directive on privacy and electronic communications.[4]

The activities focussed on the ethical and privacy issues and compliance to legislation in SONNETS are carried out in WP6 and are applied in accordance to what have been defined in D6.2 PODP- Requirement No. 2.

2.2 Standards and metadata to be used

Data will be shared in relation to publications (deliverables and papers). As such, the publication will serve as the main piece of metadata for the shared data. Hence, formats to be used mainly include .doc, .pdf and .xls files, which substantially reduce the amount of metadata. Other standards do not apply to this project.

2.3 Exploitation, availability of data and re-use

SONNETS project intends to make its research data available for open access and use by other researchers and members of the public. In furtherance of this goal, SONNETS will deposit its research data into a digital repository (Zenodo[5] or similar) and make the data available for access and reuse under a Creative Commons Attribution License[6].

For sensitive/restricted data, if any, access restrictions will be enforced (e.g. by requiring specific credentials) or at least limited in the detail available, e.g. by granting an open access exclusively through aggregation, while providing the specific data to authorized users only.

---

[2] Key definitions of the Data Protection Act, ico (Information Commissioner's Office), https://ico.org.uk/

[3] http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31995L0046

[4] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:201:0037:0047:en:PDF

[5] http://zenodo.org/

6 https://creativecommons.org/licenses/

Additionally, the Grant Agreement and the SONNETS Consortium Agreement are to be referred to for further details on the ownership, management of intellectual property and access.

2.4 Archiving and preservation

Our data will need to be backed up regularly; because of likely problems with viruses and hardware in developing countries, so that up-to-date versions are stored on the ATOS Institutions server. Qualitative data will be backed up and secured by the coordinator on a regular basis and metadata will include clear labelling of versions and dates.

There are some potential sensitivities (e.g. personal data of interviewees) of the data being collected, so the project has established a system for protecting data (see deliverable D6.2 POPD - Requirement No. 2 for further details) while it is being processed, including use of passwords and safe back-up hardware.

3 SONNETS Datasets

As mentioned in the previous chapter, the DMP should address the points below on a dataset by dataset basis and should reflect the current status of reflection within the consortium about the data that will be produced.

- ✓ **Dataset reference and name:** Identifier for the data set to be produced.
- ✓ **Dataset description**: Description of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.
- ✓ **Standards and metadata**: it includes reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created.
- ✓ **Data sharing**: A detailed description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.).

   In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).
- ✓ **Archiving and preservation**: Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.

Within SONNETS, the consortium plans to create seven separate datasets. The datasets will have the structure in accordance with the guide of Horizon 2020 for the Data Management Plan.

3.1 DATASET 1: Contact users' database

1	Dataset reference and name
	DATASET 1: Contact users' database
2	**Dataset description**
	The data collected are implemented in Excel and refer to the following sections and purposes:

- Contact users' register for newsletter subscriptions, it contains name and e-mail (both mandatory). This dataset is automatically generated when visitors sign up to the newsletter form available on the project website. The register will be used in order to send issues of the project newsletters.

- Contact user's personal details with regard to messages sent to the website through the Contact form. It includes name, e-mail, message (all mandatory) and (possibly) phone. The contact details will be used to address the enquiry/request and to send

	information in the scope of the SONNETS project.
3	**Standards and metadata**
	This dataset can be imported from, and exported to a CSV, TXT or Excel file.
4	**Data sharing**
	The mailing list will be used for disseminating the project newsletter to a targeted audience. An analysis of newsletter subscribers may be performed in order to assess and improve the overall visibility of the project.
	As it implies personal data, the access to the dataset is restricted to SONNETS consortium.
5	**Archiving and preservation (including storage and backup)**
	The dataset will be preserved in Atos' servers.

<p align="center">**Table 2: DATASET 1: Contact users' database**</p>

3.2 DATASET 2: Stakeholders' database

1	**Dataset reference and name**
	DATASET 2: Stakeholders' database
2	**Dataset description**
	This dataset is related to the stakeholders that will build the SONNETS network and contains the name, surname, job function, domain field/expertise, e-mail contact and location as well as project interests or benefits, what they can contribute, and future actions & communication.
	• Contact users' register for Experts Group. • Contact users' register for Expert Advisory Board
	This dataset is used to extend and disseminate the information about the project as well as to get feedback from these collaborators.
3	**Standards and metadata**
	This dataset can be imported from, and exported to a CSV, TXT or Excel file
4	**Data sharing**
	For these sensitive/restricted data, access restrictions will be enforced (e.g. by requiring specific credentials) or at least limited in the detail available, e.g. by granting an open access exclusively through aggregation, while providing the specific data to authorized users only.
5	**Archiving and preservation (including storage and backup)**
	The dataset will be preserved in Atos' servers.

<p align="center">**Table 3: DATASET 2: Stakeholders' database**</p>

3.3 DATASET 3: Stakeholders' interviews

1	**Dataset reference and name**
	DATASET 3: Stakeholders' interviews
2	**Dataset description**
	This dataset will contain the answers of people who will participate in the SONNETS individual interviews, workshops and focus groups carried within WP2 and WP3 activities.
3	**Standards and metadata**
	Data will be collected and stored using digital audio recording (e.g. MP3) where interviewees permit. In any case, the data from these sources will be always held in transcript form in accessible .doc file format (Word).
4	**Data sharing**
	This dataset will be used to produce analytical reports on the most important societal and public sector needs as well to identify the most appropriate technologies to tackle these tasks. Results will be mainly shared through the WP2, WP3 and WP4 deliverables.
5	**Archiving and preservation (including storage and backup)**
	The dataset will be preserved in Atos' servers.

Table 4: DATASET 3 Stakeholders' interviews

3.4 DATASET 4: Stakeholders' questionnaires

1	**Dataset reference and name**
	DATASET 4: Stakeholders' questionnaires
2	**Data set description**
	This dataset will contain the answers of people who will participate in the SONNETS questionnaires carried within WP2 and WP3 activities. The surveys will be built using Limesurvey or similar and will be hosted at the project website (www.sonnets-project.eu).
3	**Standards and metadata**
	This dataset can be imported from, and exported to a CSV, TXT or Excel file.
4	**Data sharing**
	This dataset will be used to produce analytical reports on the most important societal and public sector needs as well to identify the most appropriate technologies to tackle these tasks. Results will be mainly shared through the WP2, WP3 and WP4 deliverables.
5	**Archiving and preservation (including storage and backup)**
	The dataset will be preserved in the internal project repository (Owncloud)

Table 5: DATASET 4 Stakeholders' surveys

3.5 DATASET 5: List of societal and public sector needs

1	Dataset reference and name
	DATASET 5: List of societal and public sector needs
2	Dataset description
	Dataset for analysis of societal and public sector needs is the basis of the WP2. Data will be collected from stakeholders. Data collection techniques used will be mainly stakeholder interviews, focus groups and data collection sheets filled by stakeholders. Literature data will also be used. The dataset will be useful for all users of the project outcomes. Metadata will be comprised of contextual information about the data in a text based document.
3	Standards and metadata
	This dataset is a combination of Excel/WORD/PDF documents.
4	Data sharing
	This dataset will be mainly shared through the WP2 and WP4 deliverables, which are public.
5	Archiving and preservation (including storage and backup)
	The dataset will be preserved in project internal repository (Owncloud), in the project website (http://www.sonnets-project.eu/downloads) and in a digital repository (Zenodo or similar).

Table 6: DATASET 5 societal and public sector needs

3.6 DATASET 6: List of emerging technologies

1	Dataset reference and name
	DATASET 6: List of emerging technologies
2	Dataset description
	Dataset for analysis of emerging technologies is the basis of the WP3. Data will be collected from stakeholders. Data collection techniques used will be mainly stakeholder interviews, focus groups and data collection sheets filled by stakeholders. Literature data will also be used. The dataset will be useful for all users of the project outcomes. Metadata will be comprised of contextual information about the data in a text based document.
3	Standards and metadata
	This dataset is a combination of Excel/WORD/PDF documents.
4	Data sharing
	This dataset will be mainly shared through the WP3 and WP4 deliverables, which are public.
5	Archiving and preservation (including storage and backup)
	The dataset will be preserved in project internal repository (Owncloud), in the project website (http://www.sonnets-project.eu/downloads) and in

	a digital repository (Zenodo or similar).

Table 7: DATASET 6 Emerging technologies

3.7 DATASET 7: SONNETS roadmap and briefs

1	**Dataset reference and name**
	DATASET 7: SONNETs roadmap and briefs
2	**Dataset description**
	This dataset will be developed on the analysis of the identified technologies regarding technology readiness level, relevant international research programmes and important actors and on the identification on research needs. The dataset will provide a set of recommendations (namely, the SONNETS briefs) for policy makers, researchers and representatives of the public sector. Metadata will be comprised of contextual information about the data in a text based document.
3	**Standards and metadata**
	This dataset is a combination of /WORD/PDF documents.
4	**Data sharing**
	This dataset will be mainly shared through the WP3 and WP4 deliverables, which are public.
5	**Archiving and preservation (including storage and backup)**
	The dataset will be preserved in project internal repository (Owncloud), in the project website (http://www.sonnets-project.eu/downloads) and in a digital repository (Zenodo or similar).

Table 8: DATASET 8 SONNETS roadmap and briefs

4 Conclusions

This Data Management Plan provides an overview of the data that SONNETS project will produce together with related challenges and constraints that need to be taken into consideration.

The analysis contained in this report allows anticipating the procedures and infrastructures to be implemented by SONNETS project to efficiently manage the data it will produce. Nearly all project partners will be owners or/and producers of data.

The SONNETS Research Data Management Plan will put a strong emphasis on the appropriate collection – and publication should the data be published – of metadata, storing all the information necessary for the optimal use and reuse of those datasets.

Specific attention will be given to ensuring that the data made do not break either partner IPR rules or regulations and good practices related to personal data protection. For this latter point, systematic anonymization of personal data will be made.

5 References

[1] Guidelines on Data Management in Horizon 2020, V2.0, 30 October 2015, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf.

[2] Guidelines on Open Access to Scientific Publication and Research Data in Horizon 2020, Version 2.0, 30, October 2015, http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf.

[3] SONNETS Grant Agreement No. 692868 (Grant Agreement-692868-SONNETS.pdf)

[4] SONNETS consortium agreement (SONNETS_CA_v2.0_signed.pdf).

[5] Key definitions of the Data Protection Act, ico (Information Commissioner's Office), https://ico.org.uk.

[6] http://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:31995L0046

[7] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:201:0037:0047:en:PDF.

[8] http://zenodo.org

[9] https://creativecommons.org/licenses/