



SOCRATIC

Social Creative Intelligence Platform



@socratic2020



www.facebook.com/socratic2020

www.socratic.eu



European
Commission

Horizon 2020
European Union funding
for Research & Innovation



About the Project

Project Title	Social Creative Intelligence Platform for achieving Global Sustainability Goals
Project Acronym	SOCRATIC
Project No	688228
Call	H2020-ICT-2015 CAPS
Type of Action	RIA- Research and Innovation Action
Thematic Priority	Social Innovation, Citizen Empowerment, Gamification, Collective Awareness Platform, Innovative Solutions, CAPS projects, Global Sustainability Challenges, Open Intelligent Collaborative Platform, Millennium Development Goals (MDGs)
Start Date of Project	01.01.2016
Duration of Project	24 Months
Project Website	www.socratic.eu
Social Networks	www.facebook.com/socractic2020 @socratic



About the Deliverable

Work Package	WP05
Task	D1.5
Deliverable lead	Farapi
Authors (Org)	Inès Dinant (Farapi)
Reviewers (Org)	María Ajuria (Farapi)
Dissemination Level:	Public
Nature of the Deliverable	Report
Due Date	01.06.2016
Submission Date	30.06.2016
Date	30.06.2016
Version	0.7
Abstract	The deliverable corresponds to the outlining of the SOCRATIC politics concerning the Open Access of the data that will be



	produced and used within the platform and by external actors and in the related publications.
Keywords	Open Data, Open Access, Licenses

Versioning and contribution history

Version	Date	Author	Notes
0.1	15.05.2016	Inès Dinant (Farapi)	
0.2	03.06.2016	Antonio Fumero, Jorge Rastrilla (CIB)	Updates
0.3	16.06.2016	Thomas Vilarinho (Sintef), Antonio Fumero, Jorge Rastrilla (CIB), Philip Reimer, Philipp Spindler, Christian Wolf (ATB)	Updates and review
0.4	17.06.2016	Inès Dinant (Farapi)	Addressing comment from review
0.5	24.06.2016	María Ajuria (Farapi)	Final review



0.6	28.06.2016	Inès Dinant (Farapi)	Addressing comments from reviewer
0.7	30.06.2016	Yolanda Rueda (CIB)	Finalization of the document

Disclaimer

The information, documentation and figures in this deliverable are written by the SOCRATIC project consortium under EC grant agreement 688228 and do not necessarily reflect the views of the European Commission. The European Commission is not liable for any use that may be made of the information contained herein.

Copyright notice

© 2016 - 2018 SOCRATIC Consortium

Acknowledgment

This report is funded under the EC H2020 CAPS project SOCRATIC, grant agreement 688228.

Table of Contents

1. Introduction.....	10
1.1. The Open Research Data Pilot.....	10
1.2. The SOCRATIC project.....	11
1.1. Deliverable structure.....	12
2. SOCRATIC Open Access Policy.....	13
2.1. What is Open Data?	13
2.2. The Legal Openness.....	14
2.3. The Technical Openness.....	15
2.4. Scientific Publications.....	16
2.6. Research Data.....	18
2.7. Metadata.....	18
2.8. Archiving and Preservation.....	20
3. Data Sets.....	22
3.1 Users Base Data.....	23
3.2. Activities Description data.....	25
3.3. Research- and Experience-based Data.....	27
3.4. Platform Performance Data.....	29
3.5. Sustaining Stage Data.....	32
3.6. Pilot Research Data.....	33
3.7 SOCRATIC Observatory Data.....	35
4. Conclusion.....	36
5. Bibliography.....	36



List of Tables

Table 1. Recommended conformant licenses.

Table 2. Metadata Scheme Elements (Zenodo)

Table 3. Users Base Data

Table 4. Activities description data

Table 5. Research -and Experience-based data

Table 6. Platform Performance data

Table 7. Sustaining Stage data

Table 8. Pilot research data

Table 9. SOCRATIC Observatory Data

Executive Summary

SOCRATIC is part of the Horizon 2020 Open Research Data Pilot; its main goal being *“to improve and maximise access to and reuse of research data generated by projects, encouraging good data management as an essential element of research best practice”*.

This deliverable provides the first version of the Data Management Plan (DMP) elaborated by the SOCRATIC consortium. This DMP will be updated during the project lifetime; hence a more detailed versions of the DMP will be submitted as additional deliverables at later stages of the project.

Different kind of data sets, metadata and related information will be generated and gathered along the validation and assessment stages of our research project.

Personal information will be anonymised, and it will be made available as consolidated statistical and socio-demographic data on the SOCRATIC platform usage.

Self-created information (interview, questionnaire, focus groups, etc.); and contributions (documents, videos, etc.) will be made accessible as research raw data to be potentially reused.

Scientific papers and publications will preferably follow an open publishing (Gold Open Access) policy; while the long term preservation and archiving of research data, and information will be guaranteed through the usage of a third party repository.

List of Terms and Definitions

APC: Author Processing Changes

API: Application Programming Interface

CC: Creative Commons

DMP: Data Management Plan

DOI: Digital Objective Identifier

DS: Data Sets

EC: European Commission

GNU GPL: GNU General Public License

IF: Impact Factor

KPI: Key Performance Indicators

OA: Open Access

OAIPMH: Open Archives Initiative Protocol for Metadata Harvesting

ODC: Open Data Commons

ODH: Open Data Handbook

ODUP: Open Data Use Plan

ORDP: Open Research Data Pilot

PDDC: Public Domain Dedication and Science

SIP: Social Innovation Process

1 Introduction

1.1. The Open Research Data Pilot

According to OpenAIRE description¹, “Open data is data that is free to access, reuse, repurpose, and redistribute. The Open Research Data Pilot (ORDP) aims to make the research data generated by selected Horizon 2020 projects accessible with as few restrictions as possible, while at the same time protecting sensitive data from inappropriate access”.

The main goal of such an ORDP is to have the results of publicly-funded research disseminated more broadly and faster, reaching a broader audience, not only within the researching community, but within the target/beneficiaries groups of the society too.

A well defined and clearly stated Open Access (OA) policy:

- Accelerates the discovery process, as well as shorten the time-to-market requirements, that directly translates into a higher return on public investment.
- Avoids the duplication of research efforts, directly producing a higher throughput from our R&D ecosystem.
- Increases the visibility of the performed research, obtaining a better impact indicators.
- Offers a higher collaboration potential for all the consortium partners, exposing their capacities, and milestones in different projects.

Under this criteria, we must ensure Open Access (OA) to all peer-reviewed scientific publications -and the “underlying data”, i.e. the research data needed to validate the results presented in such publications-, coming out of our researching efforts. Such an OA will be based on a balanced support to both ‘Green’, and ‘Gold’ OA (see below in section 2.4.), describing, at the same time, how we are going to gather and manage (share, export, and/or archive) all that underlying data.

This document is intended to be a “life” document; hence it will be updated later on. Its structure and contents are inspired in the DCC Checklist for a Data Management

¹ <https://www.openaire.eu/opendatapilot>

Plan (Digital Curation Center 2013), the guidelines given by Horizon 2020 (EC 2013)) and the DMP online platform (<https://dmponline.dcc.ac.uk/>). The DMP online platform has been a very useful tool to understand and define the kind of information to be included in this Open Data Use Plan (ODUP)

This deliverable will be the baseline for dealing and managing the data generated within the SOCRATIC platform beyond the EU funding horizon. The references regarding the process followed, roles and relationships are directly linked to our SOCRATIC concept definition, reported as deliverable D1.4.

1.2. The SOCRATIC project

According to the description found on our project's website², SOCRATIC will deliver *"a knowledge-based Internet Platform offering a set of tools and services to support the whole Social Innovation Project life cycle from problem identification and awareness and creative solution ideas to collective decision-making, design and implementation of the best ideas"*.

The data and information gathered will be related to the created challenges and the ideas that will emerge to define a solution to those challenges. For their definition, prototyping and sustaining, external information (papers, video, books, etc.) will be used by innovators, alongside with the data sets they create when observing and collecting facts from fieldwork (observation notes, questionnaires, interviews, images, etc.).

This direct relationship with the reality will be of major significance, since it represents a key element of the Social Innovation Process (SIP) that SOCRATIC platform supports. Hence, we are going to deal with a wide variety of content formats, complex relationships among different data pieces, as well as different sensitivity degree within each data set.

² <http://www.socratic.eu/about-us-3/>

1.3. Deliverable structure

The structure of this deliverable will be as follows:

Chapter 1 exposes the role played by this Open Data Use Plan (ODUP) within the SOCRATIC project's context.

Chapter 2 describes the generic Open Access (OA) policy applying to each one of elements to be considered within our DMP, i.e. Data Sets (DS), Metadata, Scientific Publications, etc.

Chapter 3 details the different DSs generated by our SOCRATIC platform according to the H2020 guidelines.

In **Chapter 4**, the conclusions of this deliverable are highlighted.

2. SOCRATIC Open Access Policy

Following the EC Guidelines on Data Management in Horizon 2020 (2013), the definition of Open Access is as follow: it *refers to the practice of providing online access to **scientific information** that is free of charge to the end-user and reusable. “Scientific” refers to all academic discipline. In the context of research and innovation, “scientific information” can mean: 1. Peer-reviewed scientific research articles (published in scholarly journals); 2. Research data (data underlying publications, curated data and/or raw data).*

As explained in the Open Data Handbook, two kind of opening have to be contemplated. On one side, the Legal Openness which means the need of the the use of an open License. On the other hand, a Technical Openness is required with the objective of making the data easily available, transmittable in bulk and in an open, machine-readable format.

2.1. What is Open Data?

The Open Data Handbook³ (ODH) defines Open Data as follows: *“Data is open if it can be freely accessed, used, modified and shared by anyone for any purpose - subject only, at most, to requirements to provide attribution and/or share-alike”.*

Following the same source, *“those have to be disponible in an easy way and have to be modifiable if possible; they have to be able to be reused and redistributed easily and anyone should have access to them”.* This openness is necessary since it permits the “interoperability”, which is important to get different actors and/or institutions working together.

As the ODH indicates for the data themselves, it is also very important since it permits to work upon existent data or mixing different existing data, without having to start from scratch.

Hence, in order to maximise this “mixing” capability, and its derived collaboration potential, it is important to define very clear definition of the openness characteristics of the data. We’ll be considering two main dimensions of such an

³ <http://opendatahandbook.org/guide/es/what-is-open-data/>

“openness” -the legal openness, and the technical one-, applying to three main realms: scientific publications, research data, and metadata.

2.2. The Legal Openness

Following the proposal of Open Definition⁴, a variety of licenses schemes that respond to the characteristics on which the Open Data is based are available. We briefly summarized in the table below -according to their domain of application, and their requirements for ‘Attribution’ (BY) and/or ‘Sharealike’ (SA)-,the different options available from two of the most popular projects so far:

- **Open Data Commons (ODC).** It was founded in the Autumn, 2007 to provide legal tools for sharing data. This project started at the University of Edinburgh, with the creation of the Public Domain Dedication & Licence (PDDL) legal tool, currently maintained by the Open Knowledge Foundation, a not-for-profit organisation promoting open knowledge.
- **Creative Commons (CC).** It was founded in 2001 with the support from the Center for the Public Domain and the William and Flora Hewlett Foundation. In December 2002, CC released its first set of copyright licenses for free to the public. Their licenses have been inspired in part by the Free Software Foundation’s GNU General Public License (GNU GPL). They are offering a Web application platform to help individuals to license their works freely for certain uses, on certain conditions; or dedicate your works to the public domain. As of 2015 there are over 1 billion CC licensed works online. Their license suite has been ported to over 50 jurisdictions.

License	Domain	BY	SA
Creative Commons CCZero (CC0)	Content, Data	N	N
Open Data Commons Public Domain Dedication and Licence (PDDL)	Data	N	N
Creative Commons Attribution 4.0 (CC-BY-4.0)	Content, Data	Y	N

⁴ <http://opendefinition.org/licenses/>

Open Data Commons Attribution License (ODC-BY)	Data	Y	N
Creative Commons Attribution Share-Alike 4.0 (CC-BY-SA-4.0)	Content, Data	Y	Y
Open Data Commons Open Database License (ODbL)	Data	Y	Y

Table 1. Recommended conformant licenses (Source: Open Definition).

We'll be considering two main types of elements to be associated with these kind of licenses: the content itself, referring to public reports and (scientific) publications; and the research data to be released as data sets (see below, chapter 3).

The variety of such data sets, and the requirements from the (scientific) publications editors will be setting the limitations imposed to our OA policy (see the following sections within this chapter below).

2.3. The Technical Openness

A good way to give access to the information is using an Application Programming Interface (API) which permits to take part of the information the users are interested in and since this data is directly linked to the platform it permits to get up to date information all the time. As exposed in the Open Data Handbook, using the API permits the independency of the user use of the data from the original data supplier.

The use of API will have to be included as a part of the platform technicity which still have to be defined. Those data will be kept in the platform's internal database and it will be maintained there until it is deleted.

The data received through this API would not contain the complete data of the database. Instead it would be an activity log. e.g. it would contain information like the following:

- User X created challenge Z
- User X posted idea Y to challenge Z

where X will be an identifier that cannot (or only with a huge effort) be traced back to the real user's name.

The actual data that will be contained in variables Y and Z is still to be defined. It could only be the information that a challenge related to a certain Innovation goal has been created - and no further information about the challenge itself.

The output of the API will be files in CSV or XML format (to be decided within an updated version of our DMP) and these files could then be made available to the research community, through the SOCRATIC website and would be accessible on request.

It's important to highlight that this is the first version of the SOCRATIC DMP which means that part of this information will be evolving and changing based on how the platform and methodological content of the SOCRATIC framework evolve.

2.4. Scientific Publications

The term "scientific publications" will primarily mean herein journal articles; while, whenever possible, SOCRATIC will provide access to other types of scientific publications such as presentations (dissemination collaterals), public deliverables (PU), etc.

One of our OA policy's goals is to manage that any peer-reviewed scientific publication made within the context of the project will be made (fully) available online to any user at no charge. The two ways EC is considering for complying with this requirement are:

- **Self-archiving (referred to as 'green' OA).** *The beneficiaries deposit the final peer-reviewed manuscript in a repository of their choice. In this case, they must ensure open access to the publication within a maximum of six months (twelve months for publications in the social sciences and humanities).*
- **Open access publishing (referred to as 'gold' OA).** *Researchers publish their results in open access journals, or in journals that sell subscriptions and also offer the possibility of making individual articles openly accessible via the payment of author processing charges (APCs) (hybrid journals). Again, open access via the chosen repository must be ensured upon publication.*

Publications arising from the SOCRATIC project will be made public preferably through the option of 'gold' OA. In other cases, the scientific publications will be deposited in a repository ('green' OA). Sometimes publishers imposed a period of restricted access (embargo period) up to 6 or 12 months (maximum).

According to the EC recommendation, whenever possible the SOCRATIC consortium will retain the ownership of the copyright for their work through the use of a 'License to Publish'. Additionally, to ensure that others can be granted further rights for the use and reuse the work, we will ask the publisher to release the work under a Creative Commons license, preferably CC-0 or CC-BY.

Beyond the generic dissemination criteria inspiring these principles, we will also consider the relevance of the journal where our research results are going to be published in terms of their "impact factor" (IF).

All the scientific publication will acknowledge the project funding. This acknowledgment must be included also into the metadata required, since it allows to maximise the discoverability of publications and to ensure the acknowledgment of EU funding.

Required metadata will include:

- The terms "European Union (EU)", and "Horizon 2020" or "H2020"
- Acronym of the action (SOCRATIC)
- GA number
- Publication date
- Length of embargo period (when applicable)
- Persistent identifier (e.g. DOI)

2.5. Research Data

According to EC's guidelines, Open Research Data Pilot (ORDP) applies to two types of (research) data:

- **Underlying data**, including associated metadata, needed to validate the results presented in scientific publications.
- **Other data**, including associated metadata, as specified and within the deadlines laid down in a Data Management Plan (DMP), to be developed by the project.

According to this requirement, the underlying data related to the scientific publications will be made publicly available (see chapter 3 below) among other data sets gathered by SOCRATIC platform. These data will include a description of the procedures followed to obtain the results supporting such publications as well as data generated following those procedures.

The data sets linked to the use of the platform will be available to the research community on a request basis. Such an access should be provided through a public API, enabling the researchers properly identified and authorised to download these raw data sets, formatted as CSV or XML files, and at least one 'readme.txt' file that is intended to be used as an established way of accounting for all the files comprising the data set and explaining how all these files that make up the data set relate to each other or what format they are in.

2.6. Metadata

Metadata usually refers to *"the relevant information that describes the data that is being published with sufficient context or instructions to be intelligible for other users"*. It should allow a proper organization, search, access and retrieval to such data, preferably via a web-based catalogue.

Within this DMP, we can consider two types of metadata: that corresponding to the publications, and that corresponding to the published research data. While the nature of the data sets generated within SOCRATIC project are quite different, we'll be taking into account a common minimum set of elements extracted from some

generic use third-party repositories, that is the case for Zenodo⁵, recommended within OpenAIRE framework.

Title	Free Text
Creator	Last, First Name
Date	DD/MM/YYYY
Contributors	Acknowledging SOCRATIC consortium
Subject	Comma separated keywords set
Description	Free Text
File Format	E.g. MP4, MP3, DOCX, PDF, PNG, TXT
Resource Type	Audio, Image, Video, Document
Persistent Identifier	DOI
Access Rights	Open Closed Restricted Embargo

Table 2. Metadata schema elements (Zenodo)

⁵ <https://zenodo.org/record/11748#.V2RRG-Z95hE>

2.7. Archiving and Preservation

A repository will be used by the project consortium to make the project results publicly available and free of charge for any user. Several options have been suggested by the EC, within its ORDP guidelines:

- *For depositing scientific publications:*
 - *Institutional repository of the research institutions*
 - *Subject-based/thematic repository*
 - *Centralised repository (e.g., Zenodo repository set up by the OpenAIRE project)*
- *For depositing generated research data:*
 - *A research data repository which allows third parties to access, mine, exploit, reproduce and disseminate free of charge*
 - *Centralised repository (e.g., Zenodo repository set up by the OpenAIRE project)*

Depending on the licensing scheme (it will vary), embargo periods, and copyright agreements imposed by editors (in the case for scientific publications), we will be prompting our partners to keep using their regular social media channels, and researching communities (e.g. ResearchGate) for disseminating SOCRATIC researching results.

While we will take advantage of our own institutional repositories, we will also opt for using Zenodo centralised repository (developed by CERN under the EU FP7 project OpenAIREplus, GA no. 283595), widening our project's reach and our dissemination scope. This repository is the one mainly recommended by the EC's OpenAIRE initiative in order to unite all the research results arising from EC funded projects.

Zenodo enables users to:

- *easily share the long tail of small data sets in a wide variety of formats, including text, spreadsheets, audio, video, and images across all fields of science*
- *display and curate research results, get credited by making the research results citable, and integrate them into existing reporting lines to funding agencies like the European Commission*



- *easily access and reuse shared research results*
- *define the different licenses and access levels that will be provided*

From Zenodo main features we can highlight the following:

- *It assigns a Digital Object Identifier (DOI) to all publicly available uploads*
- *This repository also makes use of the OAIPMH protocol (Open Archives Initiative Protocol for Metadata Harvesting) to facilitate the content search through the use of defined metadata.*
- *It follows the schema defined in INVENIO*
- *It is exported in several standard formats: MARCXML, Dublin Core, DataCite Metadata Schema, etc.*
- *It uses the same cloud infrastructure as research data from CERN's Large Hadron Collider.*

3. Data Sets

In principle, the data sets considered within this 'Open Data' Management Plan (ODPM) are the ones regarding the data underlying our research activities, and the results reflected on the scientific publications delivered to the community. Such activities are mainly limited to the two main pilot scenarios we are going to use for validating, and assessing SOCRATIC platform and methodology: the one from NTNU EiT & AppLabs programs and the one for CIB and their cybervolunteers 'on-the-field missions'.

Beyond those pilot scenarios, we are including here a few data sets (to be reviewed in future versions of this document) directly associated either to our SOCRATIC Social Innovation Process (SIP) & Conceptual Framework, or to the usage & assets from the platform itself. We consider such data sets potentially useful in terms of offering to the research community open access to a full featured collaborative platform. We are offering these data sets in a more restrictive way, anonymizing data and limiting some of them to a simple statistical dashboard visualization and/or authorized access via API or exporting tools. This will be made clear in the Terms of Service agreement of the platform.

While the licensing scheme will vary, for archiving and preservation purposes, we'll be using the repositories available from our partners, and a third party repository: Zenodo has been selected from the EC recommendation, and the OpenAIRE project outputs.

Whenever possible, the data sets standards are set to a minimum, offering the raw data in CSV or non formatted XML for downloading; while the associated metadata will follow the common requirements from the repository services selected.

Our DMP complies with strong ethics requirements; hence any data set included here, and the associated sharing mechanisms will be compliant with the relevant European (The Charter of Fundamental Rights of the EU and Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data) and national legislation.

We'll be summarizing along this chapter the seven data sets we have identified within this first version of our DMP. We are using the format outlined in the table below.

Reference	DS# - XYZ
Name	XYZ Acronym
Description	Free text
Standards & Metadata	Metadata standards or ad-hoc conventions
Sharing & Management	License scheme & Openness
Archiving & Preservation	Repository & Preservation details

3.1 Users Base Data

Reference	DS1 - UBD
Name	Users Base Data
Description	<p>As a collective, awareness platform, SOCRATIC will be based on certain degree of openness regarding the information distribution, and the platform operation and performance. This data set will gather a variety of items that will be anonymised for statistical visualization and/or raw data export.</p> <p>This data set will include socio-demographic statistics (e.g. sex, age group, location, etc.), professional (e.g. interests, skills, etc.) and activity-related (e.g. on-going projects, sustainability goals attached, target groups, etc.) figures.</p> <p>Data -shared anonymously- could underpin a scientific publication in the case a state-of-the-art research is made on the usage of collective awareness collaborative platforms for social innovation. Such a research could integrate a simple</p>

	analysis on statistical data, or even comprising a consolidated database of similar platforms.
Standards & Metadata	The data compiled within this data set will be directly associated to a few user-generated metadata items, i.e. tag-based scheme for self-organising users' skills, interests, and the SDGs tackled by the challenges the users try to cope with.
Sharing & Management	<p>License: ODbL</p> <p>Openness: the raw historical data will be available for download upon request.</p> <p>Software tools/methods: CSV formatted file for downloading.</p>
Archiving and preservation	The personal, professional and platform activities' data will be kept as long as the person remains as a platform user. At the moment that the user decides to quit the platform, while their personal data will be processed according to POPD enforced regulation, the activity-related data will be kept within SOCRATIC database as part of the platform regular operation.

Table3. Users Base Data

3.2. Activities Description data

Reference	DS2 -ADD
Name	Activities Description Data
Description	<p>The activity description is related to the challenge and idea description and development. The challenge describes the social need observed, the beneficiaries (target group) the challenge affects, and the areas of interests covered; while the idea must describe a possible solution to respond to the uncovered need exposed in the challenge definition, specifying which aspects of the challenge that idea deals with, detailing the resources needed, i.e. skills needed, partners and information.</p> <p>This information results potentially useful for anybody that is working in the same areas of interests and for researchers analyzing the social innovation process. It can provide them with new perspectives on how to tackle a societal problem.</p> <p>From a researching point of view, this information could be integrated by direct inspection through the information retrieval and searching features provided by SOCRATIC platform, or via export tool for raw database data, filtering challenges & ideas according to SDGs associated (platform-specific) tags, or user-generated ones.</p>
Standards & Metadata	<p>Challenges will be classified according both to pre-defined labels associated to the SDGs we're committed to, and the curated, user-generated, tag scheme. Such metadata will be used as an information retrieval mechanism for any external/internal platform user.</p>
Sharing &	License: ODbL

<p>Management</p>	<p>Openness: The information linked to the challenges is open data within and outside the platform. The “outside part” is important to get more people registered and involved in the platform.</p> <p>The tags will serve to connect the adequate platform users and outsiders’ interests to the challenge so that their contributions can be as accurate as possible. To do so, the tags are the same as the one the platform user provides when registering.</p> <p>The information of the idea linked to expertise and team building will be open within the platform users:</p> <ul style="list-style-type: none"> ○ Tags will serve the platform end user to identify easily the aspects/expertise areas of the challenge that the idea deals with. ○ Tags of needed skills , kind of partners and information required for the solution definition will help the possible Challenge Solver to identify possible ideas to collaborate with. <p>It’s important to outline that the link to the profiles of owners and contributors of challenges and idea will not be made available out of the platform.</p> <p>For both, the challenge and the idea information, besides the instrumentally use, previously defined, this information will have the possibility of being extracted to give a wider picture of the activity that takes place in the platform, for the coordinator and for outsider researchers.</p> <p>Software tools/methods: CSV formatted file for downloading.</p>
<p>Archiving & Preservation</p>	<p>This information will be archived and preserved in the information repository of the platform as long as this platform exists. It will be periodically backed up.</p>

Table 4. Activities Description Data

3.3. Research- and Experience-based Data

Reference	DS3 - RED
Name	Research- and Experience-based Data
Description	<p>The external and internal inputs are collected in the challenge description, idea definition and the different tests that can take place during the solution development, in the prototyping phase.</p> <p>Such <u>inputs</u> may include research collaterals (papers, pictures, videos, etc.) -mainly shared for scientifically backing the different stages within the SIP process- and experience-based material (questionnaires' answers, comments, interviews, focus groups, etc.) that are intended for contributing with challenge solvers and/or beneficiaries' feedback into such stages.</p> <p>This information can be potentially useful for any researcher and/or innovator that is working on the same topics and/or SDGs-related projects. The data items compiled here will be easily integrated as individual identifiable content items and documents to be downloaded.</p>
Standards & Metadata	Standard metadata for content assets will be used, describing file format (MP3, MP4, PDF, etc.), resource type (audio, video, text, etc.), creation time stamp, etc. Additional platform-specific, and user-generated metadata will be attached for associating all the assets attached to the same challenge-idea tupla.
Sharing & Management	License: CC-BY-SA whenever possible, following the requirements enforced by the licensing scheme attached in the case of external inputs coming from Internet publicly



	<p>available sources.</p> <p>Research collaterals (papers, pictures, videos, etc.)</p> <ul style="list-style-type: none"> • Is available to the platform users once logged in and through SOCRATIC platform. The used tags serve to connect the information with the area of interest and/or skills (depending if it's to contribute to a challenge definition or to the ideation phase) of the platform user to the challenge so that the made contribution can be as accurate as possible. The tags are the same as the one the platform user has to fill in when registering. This external generated information will be made available within the platform and will be stored in the platform repository. <p>Experience-based material (questionnaires' answers, comments, etc.)</p> <ul style="list-style-type: none"> • The comments remain only visible for other platform users involved in the same challenges and/or ideation processes. When tagged they can become open for the rest of the platform users. • In case they share information directly produced by them, they will have the choice to select to share this information through the repository (and tags) or not. <p>Software tools/methods: Information retrieval + assets download features via FTP/HTTP</p>
<p>Archiving & Preservation</p>	<p>The Research material (papers, pictures, videos) provided for the challenge remains in an external database and will be integrated into SOCRATIC Observatory (to be confirmed).</p> <p>Experience based material (questionnaires' answers, comments, ...) will be kept as raw (anonymized) data that</p>

	<p>could be reused for processes that could have similarities.</p> <p>All the assets will be archived and preserved in the information repository of the platform as long as this platform exists. They will be periodically backed up.</p>
--	---

Table 5. Research- and Experience-based Data

3.4. Platform Performance Data

Reference	DS3 - PPD
Name	Platform Performance Data
Description	<p>Platform- and Process-specific Key Performance Indicators (KPI) will be defined as part of SOCRATIC concept. They are defined along four main dimensions:</p> <ul style="list-style-type: none"> ● Organization: The organization KPIs captures the performance of the organization through obtaining the following data: <ul style="list-style-type: none"> ○ The number of project taking place in the platform ○ The number of external actors (experts, beneficiaries, other users, ...) involved in the process ○ The use of internal resources (guidelines, information repository, ...) ● Individual. The individual KPIs are used to capture how the actors engaged in SI are performing through obtaining the following data: <ul style="list-style-type: none"> ○ The number of challenges and ideation processes a platform user is involved in and his/her involvement ○ The number of challenges, ideation processes and

	<p>prototyping activities a Beneficiary</p> <ul style="list-style-type: none"> ○ The team effectiveness <ul style="list-style-type: none"> ■ Expertise diversity when needed ■ Good communication ● Technology. Technology indicators related to how the platform is being used by the performance through obtaining the following data: <ul style="list-style-type: none"> ○ The number of platform users ○ The number of institutions using the platform ○ The degree of interaction between the different actors of the platform, through the platform ○ The degree of use of the resources offered by the platform (guidelines, comments making/exchange, information repository ...) ● Process. These indicators are derived from the measurement of efficiency of the process execution through obtaining the following data: <ul style="list-style-type: none"> ○ The number of challenges identified ○ The number of ideas that goes through sustaining ○ The number of solution that goes through the scaling phase ○ The level of interaction with target population ○ The level of integration of the target population's feedback ○ The degree of satisfaction of the Challenge Owner and the Beneficiaries at the end of the process <p>The performance data are directly extracted from the platform usage; and they will be used by the platform coordinator herself. This information could also be used by external users to analyze and understand the innovation process that the SOCRATIC platform is following.</p>
--	--

	Performance data will be made available through dashboard-like visualization only for internal monitoring and coordination purposes; while they will be available for external researchers upon request.
Standards & Metadata	There are no standard metadata to identify the kind of information we are compiling here; hence the raw data will be made available attaching platform-specific tags for each one of the KPI sets, i.e. the OITP dimensions previously described.
Sharing & Management	<p>License: ODbL</p> <p>Openness: All the information related to the KPI will be managed through the creation and API shared upon request.</p> <p>Software tools/methods: CSV formatted file for downloading.</p>
Archiving & Preservation	This information will be archived and preserved in the information repository of the platform as long as this platform exists.

Table 6. Platform Performance Data

3.5. Sustaining Stage Data

Reference	DS4 - SSD
Name	Sustaining Stage Data
Description	<p>Data related with the sustaining stage of the different solutions' development may include a variety of assets, e.g. Business Model Canvas, Marketing Plan, investment and funding information, etc.</p> <p>This information can be useful for other internal users and/or external innovators, and researchers that wants to apply the experience followed in the platform in their own processes.</p> <p>Data items here could be integrated from outside the platform, both as individual assets to be downloaded, and as pre-packaged bundle identifying the solutions reaching such a sustaining stage, i.e. going-to-market projects to be potentially resumed within different platforms (e.g. crowdfunding or entrepreneurial platforms and/or communities).</p>
Standards & Metadata	There are no standard metadata to be attached to the sustaining stage data; while platform-specific tag scheme will help to identify the assets and information associated to this stage.
Sharing & Management	<p>License: CC-BY-SA whenever is possible for individual assets, and ODbL for publicly available bundles.</p> <p>Openness: Business Model specific items will be added in the form of assets, e.g. a user journey (an epic description in plain text, as a.TXT file) or a Business Model Canvas (BMC) as a .PNG file.</p>

	<p>Specific tags will help locating the information provided for fulfilling some required information items within the BMC itself, e.g. key partners, key activities, key resources, value propositions, relationships, channels, customer segments, cost structure and revenue stream.</p> <p>The same criteria applies to the description of resources needed and/or investment and funding requirements.</p> <p>Software tools/methods: data retrieval + download feature (FTP/HTTP) upon request.</p>
<p>Archiving & Preservation</p>	<p>This information will be archived and preserved in the information repository of the platform as long as this platform exists.</p>

Table 7. Sustaining Stage Data

3.6. Pilot Research Data

<p>Reference</p>	<p>DS5 - PRD</p>
<p>Name</p>	<p>Pilot Research Data</p>
<p>Description</p>	<p>This data is collected from the pilot institutions (CIB, and NTNU). It is compiled offline, via interviews and/or focus group that can be carried out to understand better their realities and the usefulness of the SIP as defined by the SOCRATIC framework.</p> <p>This data set will comprise the whole content items, individual assets, collaterals, internal and external inputs, etc. for the whole social innovation life cycle supporting both pilot scenarios.</p>

	<p>The whole bunch, or the individual packages for each of the pilot scenarios will be available for downloading upon request, to be integrated within a different researching framework; while the associated assets could still be itemized for exporting too.</p> <p>Beyond such assets, the (researching) data will compile mainly anonymised interaction patterns, activity logs, and external and internal inputs complementing and fulfilling the SOCRATIC methodology for managing the whole social innovation process.</p>
Standards & Metadata	<p>Itemized assets, content items and collaterals will follow the ad-hoc standards regarding media type, file format, timestamping, and creation metadata, while the packaged bundles for researching will be identified via source (e.g. pilot institution, project-specific tags), research objective, and methodology followed to gather the information shared (e.g. interview, online survey, form-based feedback, etc.).</p>
Sharing & Management	<p>License: ODbL, and CC-BY-SA whenever applicable to content items.</p> <p>Openness: raw bundled data will be provided upon request; while individual assets will be available for downloading as part of the regular platform operation.</p> <p>Software tools/methods: information retrieval and export features (raw data into CSV or XML formats).</p>
Archiving & Preservation	<p>This information will be archived and preserved in a repository of access only to the project participants and available on request.</p>

Table 8. Pilot Research Data

3.7 SOCRATIC Observatory Data

Reference	DS7 - SOD
Name	SOCRATIC Observatory Data
Description	<p>We'll be implementing a syndication feature for gathering information, news, and data feeds regarding the UN's SDGs we are going to cope with, i.e. #3, #4 and #8.</p> <p>We, at the same time are going to compile for researching purposes the feeds regarding CAPS-related initiatives, as well as external ideas, projects, etc.</p> <p>The SOCRATIC observatory itself is one of the project outputs; while the feeds log and assets archive would be a useful, easy to integrate, research input for different initiatives.</p>
Standards & Metadata	RSS (plain text formatted XML) and/or Atom will be used for exporting information feeds, while the available assets or external content items will include the regular file format, creation information, external link/identifier if available, and content type metadata attached.
Sharing & Management	<p>License: ODbL, and CC-BY-SA when applicable.</p> <p>Openness: a standard content syndication file will be available for exporting information feeds included within our observatory; while the individual assets included as external inputs will be available for downloading (FTP/HTTP) upon request.</p> <p>Software tools/methods: Atom/RSS/OML syndication feeds readers, and platform-specific content items export features.</p>
Archiving & Preservation	This information will be archived and preserved in a repository of access only to the project participants and available on request.

Table 9. SOCRATIC Observatory Data

4. Conclusion

This deliverable provides the first version of the Data Management Plan (DMP) elaborated by the SOCRATIC consortium. This DMP will be updated during the project lifetime; hence a more detailed versions of the DMP will be submitted as additional deliverables at later stages of the project. This is due to the fact that we might have not considered kinds of datas that will be produced in the process and that will emerge while going on the development of the SOCRATIC framework.

At this moment, seven different kind of data sets, metadata and related information has been defined based on the information gathered in the early stages of our research project.

The Open Access Policy section defines the legal openness and technical openness that can and will be followed in the data management and sharing of the SOCRATIC framework in general. The used policy and technical openness chosen for each data sets depend on the kind of data itself, the method through which they have been obtained and their foreseen use within the project and for further researches.

5. Bibliography

Digital Curation Centre (2013). Checklist for a Data Management Plan. v.4.0. Retrived (17/06/2016) from <http://www.dcc.ac.uk/resources/data-management-plans>

European Commission (EC) (2013). "Guidelines on Data Management in Horizon 2020" Retrived (17/06/2016) from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Open Data Handbook. Retrived (17/06/2016) from <http://opendatahandbook.org/>