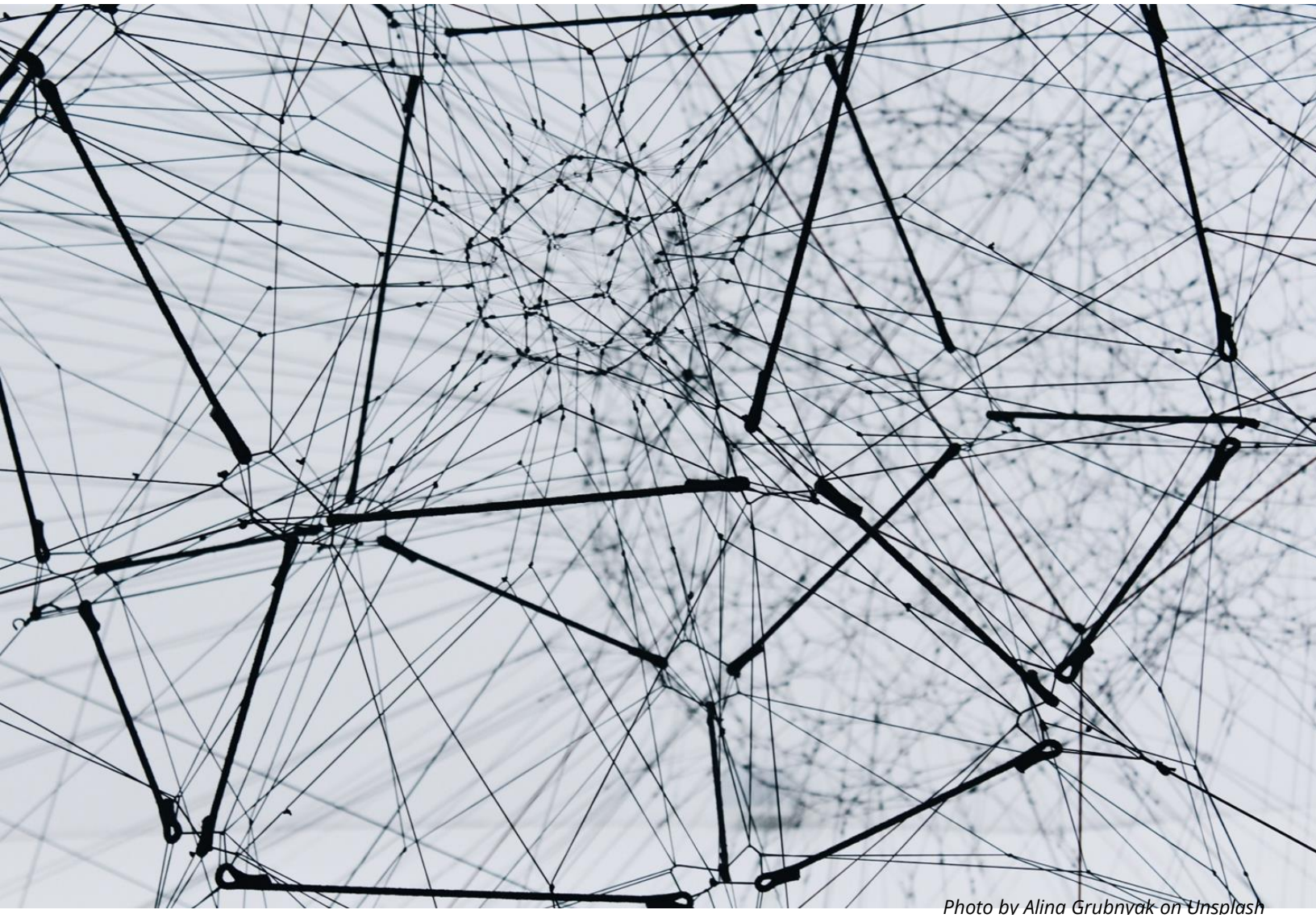


Photo by Alina Grubnyak on Unsplash

TRESCA

Trustworthy, Reliable and Engaging Scientific Communication Approaches

D7.2 Data Management Plan

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 872855.

PROJECT DESCRIPTION

Acronym:	**TRESCA**
Title:	Trustworthy, Reliable and Engaging Scientific Communication Approaches
Coordinator:	Erasmus University Rotterdam
Reference:	872855
Type:	Research and Innovation Action (HORIZON 2020)
Program:	Science with and for Society (SwafS)
Theme:	(Science) Communication
Start:	01.01.2020
Duration:	28 months
Website:	https://trescaproject.eu
E-Mail:	tresca@eur.nl

Consortium: **Erasmus University Rotterdam,** The Netherlands (EUR), Coordinator

Consejo Superior de Investigaciones Científicas, Spain (CSIC)

Observa Science in Society, Italy (OBS)

Zentrum Für Soziale Innovation GMBH, Austria (ZSI)

Science Business Publishing Limited, United Kingdom (SBIZ)

In a Nutshell, Kurzgesagt GmbH, Germany (KURZ)

DELIVERABLE DESCRIPTION

Number:	**D7.2**
Title:	**Data Management Plan**
Lead beneficiary:	**EUR**
Work package:	WP7
Dissemination level:	PU
Type	R
Due date:	30.04.2020
Submission date:	30.04.2020
Authors:	**David Arroyo**, CSIC
Contributors:	**Sara Degli Esposti**, CSIC
Reviewers:	**Jason Pridmore**, EUR
	Tessa Omen, EUR

Acknowledgement: This project has received funding from the European Union's Horizon 2020 Research and Innovation Action under Grant Agreement No 826497.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

This report presents the Data Management Plan (DMP) adopted within the TRESCA project. It includes a detailed explanation of the way personal data will be collected and processed as well as other information on how the project will make data accessible for wider research after the end of the project. The DMP may be revised and updated during the project dependent upon new issues or concerns that may arise. The purpose of this iteration of the Data Management Plan (DMP) is to provide an analysis of the main elements of the project's data management policy. The Data Management Plan (DMP) is a living document that describes data processing activities from the start until the end, that is from data collection, to data handling, open access dissemination and archiving. The document helps project partners to determine how the data can be managed efficiently and effectively and reduce the risk of data loss and conflicts. Ethical issues and data security are also briefly discussed within the document. Finally, the plan ensures consistent resource and budgetary planning for data management related costs. TRESCA participates in the extended 'Pilot on Open Research Data in Horizon 2020 ('open research data by default').

DATA MANAGEMENT PLAN

1. Introduction

The plan lists and briefly describes the types and specifications of data that is and will be collected, generated, processed or generally, used, during the project. The DMP also includes details on how the research data will be handled both during the project and afterwards. In line with Horizon 2020 guidelines, data management procedures must ensure that data are Findable, Accessible, Interoperable and Reusable (FAIR) to the extent that this is possible. Data publicity and accessibility will always be guaranteed, unless there are good reasons to keep (certain) information confidential. Accordingly, the plan indicates what datasets produced as part of research activities will be made openly available to the academic community and to stakeholders to facilitate peer-review, verification of results, and the reuse of data by other researchers.

In general, personal data during the project will be processed for the following reasons: scientific research purposes (e.g. study participants' recruitment strategy in WP2 and WP3); stakeholder engagement and marketing communications (e.g. events and conferences in WP6 or interviews with policy makers in WP1); or for offering educational services (e.g. MOOC in WP5). In compliance with the General Data Protection Regulation (GDPR), additional privacy safeguards will be adopted before data collected in WP3 are made publicly available. Therefore, "D7.5: Release of Anonymised Dataset" in month 21 will make public the confidential dataset included in "D3.2 Complete and cleaned factorial survey dataset" released in month 13 after having adopted appropriate technical procedures to guarantee the privacy of participants through anonymity.

The DMP will evolve during the lifespan of the project, particularly whenever significant changes arise within the management of the consortium and the generation of datasets. Upcoming versions of the DMP may get into more detail on particular issues such as the anonymization of the dataset of the factorial survey in WP3.

The project also complies with H2020 Open Access requirements for academic publications as explained at the end of the DMP.

2. Data summary

This project aims to develop trust in science and innovation through innovative science communication practices to foster better dialogue between scientific researchers,

journalists, policy makers and the public. In particular, the project will seek novel ways to analyze public trust in science communication through:

1. interviews with policy makers in WP1;
2. qualitative, deliberative research in WP2;
3. large scale, experimental survey research in WP3.

Datasets resulting from these activities will be either in numerical or text format. Recordings of interviews will be deleted after the conversation has been transcribed and anonymised.

Ordinary working files will be handled through EUR's educational contract for Google's G-Suite services, in particular Google Drive. The educational contract with Google does not allow for the routine scanning of documents that accompanies public and free versions of Google's services. Sensitive research files or data will be managed using CSIC's SACO cloud storage service. The resulting research data will be stored in DIGITAL.CSIC, the open access repository of the Spanish National Research Council. All datasets in DIGITAL.CSIC are accompanied with their descriptive metadata and are described according to international standards and good practices in line with the journal's data sharing policy requirements and open data mandates. The project will grant DIGITAL.CSIC repository a non-exclusive license to deposit and make available the data generated by the consortium. Through this license, the repository will also be entitled to carry out all necessary tasks to ensure digital preservation and curation.

3. Open access publications

All research outputs (publications, data, and other outputs) in line with the FAIR principles to facilitate their re-use in the future. TRESCA participates in the extended 'Pilot on Open Research Data in Horizon 2020 ('open research data by default'). All TRESCA publications, including peer-reviewed scientific publications and other possible types of scientific publications such as books, conference proceedings, and reports will either immediately be made accessible online by the publisher (Gold Open Access), or publications are available on an open access repository after the appropriate embargo period, usually 6 to 12 months (Green Open Access). Gold open access license will be acquired whenever the publisher enables that option and subject to available budget. The cost of Gold Open Access has already been included in the project budget. Alternatively, the final peer-reviewed manuscript will be archived by the researcher - or a representative - in an online repository before, after or alongside its publication.

The self-archiving 'green' open access repository Digital CSIC will be used in this case with a link to the official project webpage. DIGITAL.CSIC was created in 2008 as an Open Access repository for institutional publications and other peer reviewed outputs. Since 2010, DIGITAL.CSIC has accepted datasets as research outputs for open access dissemination and long term preservation. DIGITAL.CSIC is one top provider to OpenAire infrastructure and provides specific support to comply with the Commission's requirements for open access to publications and data. The repository guarantees long term storage, easy access and downloading facilities for open access data. Its preservation strategy is integrated into CSIC institutional preservation policy.

The project will ensure open access to all peer reviewed scientific publications resulting from the project by:

(a) depositing a copy of the published version or final peer-reviewed manuscript accepted for publication, whatever the journal policy allows for. DIGITAL.CSIC is one main contributor to OpenAire, the European Commission aggregator of Open Science;

(b) deposition and open access availability of such publications will take place within the embargo period of 12 months for SSH research as set by the European Commission.

(c) these items will be accompanied with standardized bibliographic metadata that identify the deposited resource and will also include the metadata relating to:

- the project name, acronym and grant number;
- the publication date and, if applicable, length of embargo period;
- a persistent identifier.

TRESCA team members intend to publish various articles throughout the cycle of the project. Potential journals for publication have been identified and the final list of articles will be included in D6.4 "Academic articles submitted to peer-reviewed journals". Journals where to publish scientific findings will be selected in compliance with the H2020 open access mandate.

In addition, the CSIC team will rely on DIGITAL.CSIC to make research data publicly available. All information on where to find the data and more information on the methodology adopted will be present on TRESCA project website. By research data, this refers to:

(a) the research data generated during the project, including associated metadata, that are needed to validate the project's scientific publications.

(b) information about tools and instruments necessary to validate the scientific results.

CSIC will ensure that open access publications and open access research data do not jeopardize future intellectual property and exploitation rights that may arise in the future as discussed in the Consortium Agreement, especially with regards to the exploitation of the Misinformation Widget developed in WP4.

4. Formats to ensure data are findable, accessible, interoperable and re-usable (fair)

The research data collection produced in this project will be based on 'FAIR' principles (findable, accessible, interoperable and re-usable). Applying FAIR principles implies that digital assets are: (a) findable, that is, described by sufficiently rich metadata and registered or indexed in a searchable resource through the use of Persistent Identifiers that are globally resolvable (PIDs); (b) accessible, so that they can be obtained by humans and machines upon appropriate authorisation and through a well-defined and universally implementable protocol; (c) interoperable in the sense that digital assets follow a formal, accessible, shared and broadly applicable format and language for knowledge representation; (d) re-usable, in terms of adopting rich metadata and documentation that follow relevant community standards and provide information on provenance. As pointed out in EU "Guidelines for open access to publications, data and other research outputs"[1], the FAIR principles do not impose openness on digital assets, but rather refer to the adequate provisions under which the digital assets are curated.

Datasets created within TRESCA will follow the guidelines and formats available on the RDA Metadata Standards Catalog, which is a collaborative, open directory of metadata standards applicable to research data.

The identifiability of data will be ensured through the appropriate labelling of files, which will follow widely adopted conventions and standards. All file names will be normalised as expressed in the example below.

- Experiment 1 name will be EXP1;
- Date file: YYYYMMDD;

[1] See "Horizon 2020 projects working on the 2019 coronavirus disease (COVID-19), the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and related topics: Guidelines for open access to publications, data and other research outputs", European Commission, Version 1.0, April 8, 2020.

- Type of files: DATA for any data file; RM for read me files; CB for codebook; CODE for model codes;
- Version: format: _v001;
- Extension corresponding to the compressed folder extension used: SAV for spss data; SPS for spss syntax files; XLS for Excel files; TXT for text files; MNL for multinomial logit model.

In addition, the following conventions will be observed:

- ” ” and ”-” will be used in file names to deliminate units of metadata;
- ”-” is used for words that need to be globbed together
- ” ” separates different information units;
- no punctuation;
- no special characters (e.g. $, @, % , # , & ,* , (,), !);
- a new file with the same name as an existing file will automatically replace the existing file;
- for raw data ” raw” will be added at the end of a filename and the file is made read-only;
- For data exceeding a size limit, or for an ensemble of data files, those will be shared within a compressed folder. This folder will follow the naming convention detailed above, the “extension” corresponding to the compressed folder extension used (standard, as .zip or .rar). Caution will be made to avoid name repetition, for a clear identification of data. A supporting README text file will also be provided, indicating clearly the content of each individual file.

In order to make them easily findable, the data and the corresponding metadata uploaded in DIGITAL.CSIC will be described following the basic principles of Qualified Dublin Core which is the by default metadata standard in this DSpace-based repository. Furthermore, DIGITAL.CSIC offers at item level the option to export the metadata records in DataCite format, as well as other international formats, such as RDF, MARC, MODS, csv, and bibtex.

In order to promote standard citation, DIGITAL.CSIC has produced a description template for datasets that follows recommendations set by FORCE11 guidelines (https://www.force11.org/datacitationprinciples). In addition, the recommended description in this template follows recommendations set by DataCite Metadata Working Group. This

template is available at http://digital.csic.es/handle/10261/81323. This metadata template has been further extended so as to include DataCite specific metadata properties, such as types of contributor.

DIGITAL.CSIC assigns Digital Object Identifiers (DOI) to datasets for their clear identification and citability and fulfils all the quality criteria established by Re3data directory of repositories.

DOIs will be assigned to generated datasets directly by DIGITAL.CSIC through CSIC institutional membership to DataCite once the datasets are uploaded into its platform. DOI minting through DIGITAL.CSIC is also offered to CSIC researchers for other "non traditional" research outputs, such as research software, preprints and lab notebooks. In addition, DIGITAL.CSIC gives a handle identifier to any output that is uploaded onto its platform. Like DOI, handle is another global and permanent identifier for research outputs.

As said above, the metadata schema describing the data files will follow as standard the Dublin Core Metadata Initiative (DCMI). DCMI standard satisfies international standards (ISO standard 15836:2009, ISO CEI 11179) and has an official status from the World Wide Web consortium (W3C) and ISO 23950 norm. The metadata set consists of fifteen elements as follows. This metadata template has been further extended so as to include DataCite specific metadata properties, such as type of contributors or other aspects as explained in the table below.

Tab. 1 List of metadata

Element	Commentary
1. Title	Name given to the resource
2. Creator	Entity primarily responsible for making the resource
3. Subject	Topic of the resource
4. Description	Account of the resource (including description of measurements and parameters)
5. Publisher	Entity responsible for making the resource available
6. Contributor	Entity responsible for making contribution to the resource
7. Date	Point or period of time associated with an event in the lifecycle of the resource
8. Type	Nature or genre of the resource
9. Format	File format or physical medium of the resource
10. Identifier	Unambiguous reference to the resource within a given context

11. Source	Related resource from which the described source is derived
12. Language	Language of the resource
13. Relation	Related resource
14. Coverage	Spatial or temporal topic of the resource, spatial applicability of the resource, jurisdiction under which the resource is relevant
15. Rights	Information about rights held in and over the resource

4.1 Making data openly accessible

The consortium will follow the conditions, rules and regulations of DIGITAL.CSIC repository – including the settings for accessing the dataset. DIGITAL.CSIC accepts open access, embargoed and closed access datasets however there is a strong preference for the first two options unless legitimate reasons prevent from sharing the results openly.

In any case, metadata of all items in DIGITAL.CSIC are publicly available and ready for consumption by human users and third party research infrastructures, search engines and aggregators. OpenAIRE harvests DIGITAL.CSIC contents on a regular basis and so do other Open Science aggregators such as DataCite Search, SHARE of Center of Open Science (COS), and Google Dataset Search. In addition, DIGITAL.CSIC is findable through common search engines like Google, Bing and Yahoo.

Further, metadata schema used in DIGITAL.CSIC allows to fully document details about data files access requirements and the repository's content policy permits the upload of research software associated with the datasets.

All research data resulting from the project will be made available as soon as possible after the consortium has analyzed them and published their findings. However, there will be different access levels.

For instance, research results that can reasonably be expected to commercialisation or industrial exploitation, and/or that require a special protection due to confidentiality will not be put into the open domain. Whenever such access restrictions apply the associated metadata available in DIGITAL.CSIC will contain the email contact of one member of the project consortium to ask for individual agreements on accessing the data, even if there are general restrictions.

This action is compatible with the Grants agreement obligations (Article 29.3) concerning the dissemination of results, and in compliance with the obligation to protect results (Article 27), confidentiality (Article 36) security obligations (Article 37) and obligations to protect personal data (Article 39).

Other datasets will be openly shared after an embargo period during which the project consortium will be able to examine the generated data and work on preliminary results. Metadata of such datasets will be publicly available on DIGITAL.CSIC unless they may compromise the outcomes of the ongoing investigation.

The following table clarifies which research data generated by the project will be shared in open access. It also explains why several datasets cannot be shared and what alternative solutions will be provided.

Tab. 2 List of open access research data

Deliverables affected	Type of Data	Data made openly available (yes/no)	Rationale	Alternative solution
Interviews used in D1.2, D1.5, D2.2	Transcriptions and other qualitative data from the workshops (WP2) or from the interviews (WP1)	No	Verbatim transcriptions may contain sensitive information, such as people's political views or health status that require additional data protection safeguards	Anonymized data will be made openly available at the repository after the completion of the project. Sensitive data will not be publicly available, according to data protection law. Access to such data may be granted upon request via the repository's request a private copy functionality (e.g. an external scientist)

D3.2	Factorial survey data (WP3)	No	The dataset D3.2 may contain special categories of data (e.g. ideology or political preferences) that require additional data protection safeguards	While results of the analysis will be published in D3.3, the anonymised dataset will be published as D7.5 in month 21.
D5.3	Testing, evaluation and revision of MOOC report	No	Individual data about users' feedback and satisfaction with the MOOC will not be made publicly available in respect of users' privacy.	The result of the evaluation procedures and the improvements made will be presented in D5.3 in month 24.

4.2 Making data interoperable

The research datasets as well as associated metadata and documentation will be compliant with international standards in order to guarantee easy interoperability and re-use. Open file formats will be preferred and the consortium will apply widely-used controlled vocabularies as needed.

Metadata includes information about the methodology employed in the research, the instruments used in data production, the procedures used in preparing the data, the organization of the data itself, both in its original and derived forms, and other contextual material about the research.

Description of the datasets will be in English in order to make them discoverable, accessible and understandable by any potential user. The datasets will be described according to

mainstream Social Science & Humanities digital curation guidelines and standards. The default metadata schema adopted by DIGITAL.CSIC is the Qualified Dublin Core; onsite conversion services for metadata records are offered within the DataCite schema, a global standard to describe research data and software.

TRESCA will also adopt non-proprietary standard formats for the generated data. Using non-proprietary data formats will reduce the risk of inaccessibility and obsolescence in the long term. Therefore, formats which can be opened via frequently used software, or which are not specific to a certain software, or the software used is freely accessible will be used. The file formats to be used for storage of the generated data are listed in the table below.

Tab. 3. List of preferred file formats and extensions

Data Type	Preferred Formats	Editing Software
Generic data file	• .dat	Text editors (as Notepad, Wordpad)
Text file	• .txt • .doc • .pdf	• Text editors (as Notepad, Wordpad) • MS Word • Standard PDF viewers
Spreadsheet	• .xls • .csv • .asc	Spreadsheet software (as MS Excel)
Raster image	• .jpg/;jpeg • .png • .tif/tiff	Standard image viewers
Markup language	• .cif	XML editors (as Mercury, free downloading software
Archive file format	• .zip • .rar	Standard file archives (as WinRar, WinZip)

To ensure reproducibility and possible interoperability it is also crucial that every dataset is associated with detailed documentation in the form of a readme.txt file. This documentation will be submitted to DIGITAL.CSIC as a supplementary file together with the data files and will provide the necessary insights for future researchers about methodologies, approaches, decision making and software used.

Last but not the least, it is worth mentioning that DIGITAL.CSIC is about to complete the SCHOLIX standard and DataCite Event Data implementations, which will allow automatic linkage via an URL between the project's generated datasets to be available on DIGITAL.CSIC

and associated publications. Such interoperability will enrich and maximize discoverability, understanding, citation and impact of datasets.

4.3 Enhancing data re-usability

In principle all relevant research data generated by the project will be reusable by third parties unless special restrictions apply, as indicated in the associated metadata. The project data can be of interest to other scientists in the field of SSH research especially science communication. Other disciplines and interdisciplinary research groups might also be interested, especially those working around misinformation which is a growing multi-disciplinary research field. The methodological design developed in WP2 and WP3 will ensure the quality of the data, and the validity and generalizability of results. These include validation of the sample, replication, comparison with results of similar studies and control of systematic distortion.

The research data will be made available with the publication of D7.5. Research articles will be deposited on DIGITAL.CSIC in compliance with outlet policies, in the best scenario upon acceptance for publication.

To maximise the re-usability of the data, research data will be labelled with a clear user license. The Creative Commons CC-BY 4.0 International license will be used to protect the ownership of all data deposited in DIGITAL.CSIC, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. Open source licences will be selected on Github public license selector, while for software developed in WP4 (D4.4) will follow popular software licenses as summarised on FOSSA.

Furthermore, DIGITAL.CSIC also offers the option to embargo datasets for a period of time set by the researcher team. During the embargo period restricted access for private use may be requested at no cost through the "Request a copy" functionality available at the repository. The project will make use of such an option for the data not directly associated with publications. DIGITAL.CSIC also offers a support service to CSIC researchers to fully understand scope and obligations of publishers' data policies and different options to license data, software and other research outputs.

4.4 Allocation of resources

The consortium will use DIGITAL.CSIC repository for making the resulting datasets accessible. This will ensure that the data are safely stored in this certified repository for long term preservation and curation.

Regarding costs, the DIGITAL.CSIC repository is free for CSIC researchers and its services can be extended to other consortium partners as a contribution in kind. The service is offered by the Scientific Research Information Unit of CSIC. DIGITAL.CSIC also provides DOIs for datasets, software and other non-traditional outputs. These services are free of charge for CSIC researchers.

With regards to publications, the consortium will prioritise the green route (self-archiving) and set it as the preferred option as long as it complies with the EC Open Access Mandate. For those cases in which a Gold route will be chosen, expenses will be covered with project funds and claimed back as eligible costs. The table below shows the list of journals where the consortium aims to publish the results of the project together with an estimation of Open Access publication costs, if any.

Tab. 4. List of potential journals where to publish TRESCA's results.

Journal	Type of Open Access Option	Cost (if any)
Physical Review E (American Physical Society)	Self-archiving	-
PloSOne (PloS)	Open Access Publication	USD 1,695, excluding taxes
Proceedings of the National Academy of Sciences	Free for 6 months	All articles are free within 6 months of publication. Corresponding authors from institutions with current-year site licenses will be assessed a discounted open access fee of $1,300 (compared to the regular fee of $1,700) for a CC BY-NC-ND license.

Journal of Risk Research	There is the option to publish open access in this journal via our Open Select publishing program.	Free eprints link, so that authors can quickly and easily share your work with friends and colleagues.
Social Media and Society	Social Media + Society is an Open Access publication; all articles are freely available online immediately upon publication.	The costs of peer-review and production are covered by Article Processing Charges (APC) which is currently 400 USD.
International Journal of Communication	This is an Open Access journal. Users have the right to read, download, copy, distribute, print, search, or link to the full text of articles in this journal, or use them for any other lawful purpose.	The International Journal of Communication has no article processing charges and no article submission charges.
Social Networks	To provide gold open access, this journal has a publication fee which needs to be met by the authors or their research funders for each article published open access. An author is entitled to post the accepted manuscript in their institution's repository and make this public after an embargo period (24 months)	The gold open access publication fee for this journal is USD 2150, excluding taxes.
Information, Communication and Society	You have the option to publish open access in this journal via our Open Select publishing program	You will be asked to pay an article publishing charge (APC) to make your article open access: €2495

Science Communication	If you or your funder wish your article to be freely available online to nonsubscribers immediately upon publication (gold open access), you can opt for it to be included in SAGE Choice, subject to the payment of a publication fee.	The standard article processing charge (APC) for SAGE Choice is 3,000 USD

The process of managing the resulting data will be done by the researchers participating in the project and associated costs for dataset preparation and data management during the project will be covered by the project itself. As data security manager, CSIC will be responsible for data management plan updates, backup and storage. The CSIC, as research director, and EUR, as project coordinator, will be responsible for data archiving and publication within the repository. Likewise, DIGITAL.CSIC as well as the library in CSIC Institute of Public Goods and Policies will provide assistance to describe, upload and give support to best curate datasets that are uploaded onto the repository.

In compliance with the GA (Article 31.3), the data must be preserved for at least one year after the end of the project, and this will be provided by DIGITAL.CSIC. Long term preservation by the repository will result in no additional costs for the project consortium. The research data will be stored locally at the Erasmus University Rotterdam (for non-sensitive data) and Spanish National Research Council (for sensitive-data) with redundant, robust and secure storage. No additional costs are associated with this local storage.

5. Data security

The consortium will keep the files of the raw data generated during and after the project. Several copies of such data files will be made by the partners and will be shared amongst the consortium members in a secure infrastructure, namely, EUR based Google Drive service for non-sensitive data and CSIC SACO for sensitive data.

Data storage in TRESCA will be conducted using two cloud platforms: Google Drive (part of Google's G-Suite for Education provided by EUR to staff and faculty) and Nextcloud (SACO, CSIC private Cloud Storage and Data Sharing service).

Instructions on how to manage personal and confidential data have been given to consortium partners in line with current security standards and best practices.

A working space has been set on Google Drive for Data Storage in TRESCA using EUR Google for Education account[2]. Google Drive will be used to store non-sensitive information in TRESCA. Faculty and staff at EUR have unlimited storage, but it is necessary to define a clear access and authorization policy to protect personal information used in the course of TRESCA data collection and empirical analysis. According to the Principle Of Least Privilege (POLP), there are defined two types of Shared Drives on EUR Google Drive:

- **TRESCA Management**. This folder can be only accessed by TRESCA's management team and it contains internal and management information.
- **TRESCA Project Consortium**. All members of TRESCA have access to the information included in the "TRESCA Project Consortium" shared folder. This folder contains the working documents and all the information to be used in the preparation of the different deliverables.

Google Drive allows to define the following roles along with the corresponding permissions:

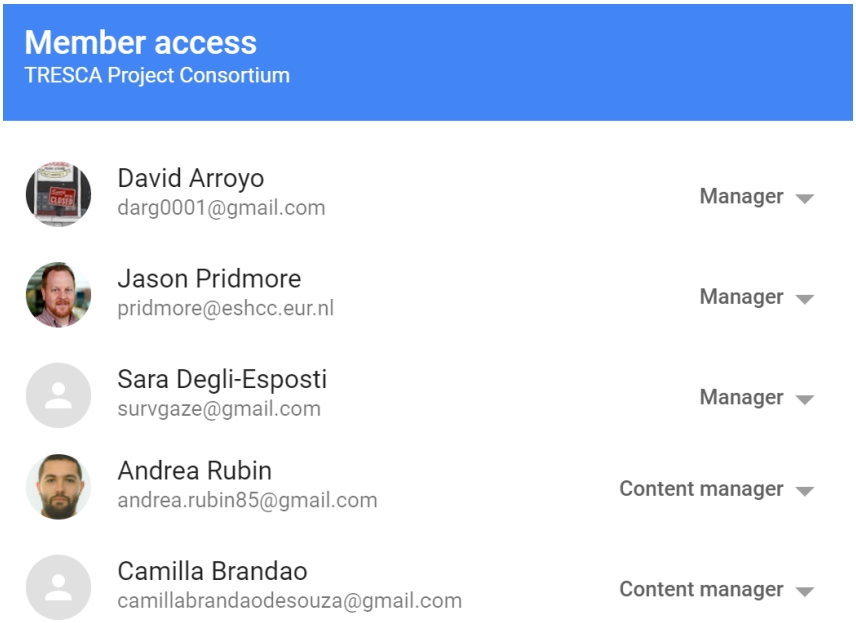
Tab. 5. Google Drive's roles and permissions.

	Permission						
Role	View	Comment	Edit	Add items	Move items	Delete items	Manage members
Manager	X	X	X	X	X	X	X
Content Manager	X	X	X	X	X	X	

[2] See https://gsuite.google.com/terms/education_terms_ie.html for Google for Education Terms of Service.

Contributor	X	X	X	X			
Commenter	X	X					
Viewer	X						

In TRESCA's shared folders only two roles are used: all the members of the management team have Manager role, whereas the rest of the team is assigned the role of Content manager.



Sensitive data is handled by CSIC. As part of its ICT infrastructure, CSIC has a cloud service deployed upon the Nextcloud technology, SACO. All staff in CSIC has access to SACO and can grant access to external collaborators through shared links associated with the SACO account of TRESCA's DPO. These shared links can be password protected and it is also possible to establish an expiration date. Since the registration of new users is something that

cannot be easily achieved on the basis of CSIC ICT security policy, TRESCA data management will be conducted applying the above protocol on shared links protected with password and expiration date. The DPO will create a folder for each of the different scenarios where data collection involves the recollection of special categories of data as established by the article 9 of the GDPR.

SACO is implemented using Nextcloud, an on-premises file sharing and collaboration platform. Nextcloud adheres to the User Data Manifesto 2.0[3].

In addition, DIGITAL.CSIC disposes of long-term storage, easy access and downloading facilities for open access data. The repository is an application under the umbrella of Central IT Services in the Spanish National Research Council and is included in the institutional preservation strategy. As a consequence, regular backups of DIGITAL.CSIC take place on a daily basis on institutional servers and copies of files are being kept in several offsite locations managed by CSIC. For its part, the Unit responsible for the repository is putting in place its own preservation system for all its library-related applications and the repository software DSpace has a number of tools to conduct regular audits in order to check the validity of formats and bitstreams and block any possible virus or undesired bot. It is worth mentioning that DIGITAL.CSIC obtained a Data Seal of Approval in 2015.

DIGITAL.CSIC offers a solid service that enables findability, accessibility and interoperability and developments towards improving the reusability of datasets are underway. In order to provide all necessary information, every deposited dataset into DIGITAL.CSIC will be accompanied by a metadata record that explains methodologies, software setups and experiment setups.

The table below provides details about storage location and backup for the datasets generated by the project:

Tab. 6. Backup policy for TRECA's datasets.

Data type	Storage medium and location	Backup location and backup frequency
Raw data & Anonymized data	The project consortium will keep copies of the raw data files in the closed shared	CSIC internal network will make backups of the files every day.

[3] https://userdatamanifesto.org/.

	environment of EUR Google Drive and CSIC SACO. Copies will also be deposited in DIGITAL.CSIC as soon as possible.	It remains possible to retrieve all stored versions for the last weeks. CSIC retention policy secures files for years.
Sensitive data	Sensitive data generated by the project will be separated as early as possible to create an anonymized dataset. Access to sensitive data is granted only for project members with clearance through non-disclosure agreements. In addition, rightful access is ensured through secure passwords and an adequate access control for the folders and files within cloud storage. Data transfer is secured via HTTPS protocol.	
Processed data	Processed data will be deposited in DIGITAL.CSIC as soon as possible. Processed data associated with publications will be deposited according to the journal's data sharing policies.	CSIC internal network will make backups of the files every day. It remains possible to retrieve all stored versions for the last weeks. CSIC retention policy secures files for years.

6. Ethical aspects

During the project, the research will generate one collection of sensitive data. In order to ensure that all ethical aspects are considered and that the project is compliant with all legal requirements and ethical issues, a general strategy has been designed by the project

coordinator (EUR) and has been discussed with all partners in the consortium. Further details for these will be detailed in D8.1, D8.2 and D8.3 which are due in Month 6 of the project.

The Guidelines will aim at offering specific guidance to all the partners of the Consortium for the performance of the different tasks and activities foreseen in WP 2 and WP 3, concerning sensitive data.

In order to ensure that all partners are compliant with the requirements related to Research Ethics and particularly, Informed consent procedures, the Guidelines include:

- a general introduction containing an explanation on the concept and meaning of Informed consent, in the context of research ethics and empirical research;
- a set of guidelines for quantitative research (online survey)
- a set of guidelines for qualitative research (interviews, focus groups, workshops);
- legal notice to be included in the online survey;
- written informed consent form;
- oral consent script.

6.1 GDPR compliance

The Guidelines will include applicable information about sensitive/personal data processing. Processing will take place when it will be strictly necessary (data minimisation) and will be accompanied by information on:

- the objectives of the data collection (purposes);
- the legal basis for processing;
- the mandatory or optional nature of data collection and the recapitulation of the categories of data processed;
- the source of the data;
- the categories of persons concerned;
- the recipients of the data;
- the duration of data conservation;
- security measures (general description);
- the possible existence of data transfers outside the European Union or automated decision-making.

Data dissemination will be conducted through aggregate analysis and according to the guideline in the Menlo Report[4]. All sensitive data will be cleansed by means of de-identification software as DEID[5], and scrubbing tools as SCRUB[6] and FAKER[7].

7. Conclusion

This document outlines the Data Management Plan (DMP) adopted within the TRESCA project. The detailed explanation is intended to be informative of the intentions of consortium members in relation to data collection and use, however in practice, the actual collection and use of data will have to be evaluated against these intentions. Data Management needs to be an active process. The normal procedures and trajectories noted here will be evaluated and reflected upon in the review process for this project. However, at present, this document represents the most up-to-date knowledge and awareness of consortium partners in relation to data that will be produced, collected and used in the TRESCA project. This plan will enable the project to be more effective and ethical in producing its deliverables but more importantly, creating ethically relevant societal impacts in our understandings of science communication.

[4] https://www.dhs.gov/sites/default/files/publications/CSD-MenloPrinciplesCORE-20120803_1.pdf.
[5] https://www.physionet.org/content/deid/1.1/.
[6] https://pypi.org/project/scrub/.
[7] https://github.com/joke2k/faker.