
# D1.2: Data Management Plan

*Authors*: S. Fuglsang, N. Mejlgaard

*Project title*: Standard Operating Procedures for Research Integrity
*Project acronym*: SOPs4RI
*Grant Agreement no.*: 824481
*Lead contractor for this deliverable:* Aarhus University

The project leading to this application has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824481.

*Coordination and Support Action*
*H2020-SwafS-03-2018*

*Project full title*

"Standard Operating Procedures for Research Integrity"

*Project acronym*

# SOPs4RI

*Grant Agreement no.*

824481

# D1.2: Data Management Plan

| | |
|---|---|
| Editor: | |
| Version: | 1.0 |
| Dissemination Level[1]: | PU |
| Author(s): | Simon Fuglsang, Niels Mejlgaard |
| Due date of deliverable: | 30 June 2019 |
| Actual submission date: | 28 June 2019 |
| Start date of project: | 01 January 2019 |
| Duration: | 48 months |
| Organisation name of lead contractor for this deliverable: | Aarhus University |

[1]This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824481

PU – Public; PP - Restricted to other programme participants (including the Commission Services); RE - Restricted to a group specified by the consortium (including the Commission Services); CO - Confidential, only for members of the consortium (including the Commission Services).

# Document metadata

| *Version* | *Date* | *Modification reason* | *Modified by* |
|---|---|---|---|
| 0.1 | 20.05.2019 | First draft | Simon Fuglsang, Niels Mejlgaard |
| 0.2 | 20.06.2019 | Second Draft, modified following input from KU Leuven, LSE, NTUA | Simon Fuglsang |
| 1.0 | 27.06.2019 | Final version | Niels Mejlgaard |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

# Table of contents



 

# Introduction

This deliverable provides the first consolidated version of the Data Management Plan (DMP) for the Standard Operating Procedures for Research Integrity (SOPs4RI) project. It describes the data collected across all Work Packages (WPs), the data collection methods, actions to secure data, and actions to secure FAIR (Findable, Accessible, Interoperable, Reusable) data.

The document is based on the Horizon 2020 template for DMPs, provided in the Horizon 2020 Participant Portal. It also draws on the structure and contents of the DMPs previously developed for the EnTIRE[2] and VIRT$^2$UE[3] projects.

In accordance with HORIZON 2020 guidelines, the document is "intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur."[4] Information regarding document history is provided on page 2.

[2] http://www.entireconsortium.eu/

[3] https://cordis.europa.eu/project/rcn/214892/factsheet/en

[4] http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

# About SOPs4RI

The objective of the SOPs4RI project is to promote excellent research that aligns with the principles and norms of the European Code of Conduct for Research Integrity, and to counter research misconduct. SOPs4RI will stimulate transformational processes across European research performing organisations (RPOs) and research funding organisations (RFOs) to achieve this goal. SOPs4RI involves a mixed-methods, co-creative approach to the development and empirical validation of standard operating procedures (SOPs) and guidelines to cultivate research integrity and reduce detrimental practices. With rigorous methods, the SOPs and guidelines will be developed by experts and key stakeholders and will be offered as flexible tools for RPOs and RFOs to develop Research Integrity Promotion Plans (RIPPs). Active researchers in the EU and selected OECD countries will be surveyed as part of the validation procedure. Proposals for appropriate incentives and novel sanctions will be developed to promote adoption. The discipline-sensitive SOPs and guidelines will support the production of sound, trust-worthy knowledge, a precondition for tackling the current crisis of truth and trust in science. In turn, a strong research integrity culture and high epistemic quality of research-based knowledge will make a vital contribution to addressing the EU's current and future challenges. SOPs4RI is fully aligned with the core objectives of the Science with and for Society 2018-20 Work Programme and will draw upon and liaise with other projects in the work programme. To maximise the feasibility and impact of the project's outcomes, SOPs4RI engages with key stakeholders in the ERA and beyond in an iterative development, cost benefit assessment, and real world validation of the toolbox of SOPs and guidelines. A strategic communication plan leveraging the support of key partners will be implemented to maximise short, medium, and longer term impacts.

# 1. Data summary

This section provides information about the character of the data collected and/or generated by the project partners. As the data generation takes place in five WPs, led by different partners, data descriptions are specific to the WPs responsible for data generation.

Specific guidelines and procedures for data collection will be provided in research protocols. These protocols will be delivered by the respective WPs and are listed among the formal deliverables of the SOPs4RI project. Preliminary outlines of the protocols can be found in deliverable 8.1 submitted simultaneously with this deliverable (SOPs4RI_AU_WP8_D8.1_EthicsRequirementNo1, Version_1.0).

## 1.1 Purpose of data collection/generation

The purpose of the data collection and generation within SOPs4RI is to contribute to the curation of existing and development of new SOPs and guidelines to promote research integrity. It is the intention that the SOPs and guidelines will be applied by RPOs and RFOs. The SOPs and guidelines will be made available through an interactive Toolbox, from which RPOs and RFOs may select tools to be employed in RIPPs.

WPs 3-7 are the empirical WPs that generate (and to a lesser degree collect existing) data. The aims for each WP are summarized separately below, and additional information about the WPs can be found in the Grant Agreement[5] and on the project's Open Science Framework (OSF) site[6].

WP1. Management and coordination
- No data collection expected.

WP2. Dissemination and Communication
- No data collection expected.

WP3. [MEFST] Systematic review of practices and research cultures
- The aim of this WP is to collect evidence and identify existing practices concerning SOPs and guidelines for research integrity through a literature review and

[5] Grant Agreement-824481-SOPs4RI

[6] https://osf.io/49fbk/

consultation of leading experts. This will be achieved by conducting scoping reviews of existing evidence, conducting expert interviews, and employing a Delphi survey.

WP4. [STICHTING VUMC] Developing SOPs and guidelines

- This WP drafts, prepares, improves, tests and finalizes the SOPs and guidelines with input from the remaining WPs. This is informed by expert interviews and co-creation workshops.

WP5. [AU] Focus group interviews

- This WP provides discipline specific knowledge on SOPs and guidelines related to research integrity. It collects data through focus group interviews.

WP6. [UESSEX] Survey

- The purpose of data collection in this WP is to test the feasibility of the prototype version of SOPs and guidelines developed in previous WPs by gathering perceptions and behavioural reports from researchers using a cross-European cross-disciplinary survey.

WP7. [OEAWI] Pilot testing

- This WP tests the SOPs and guidelines developed in WP4 in eight settings among RPOs and RFOs and collects feedback on efficiency and effectiveness as well as on costs and benefits of the toolbox, in turn enabling WP4 to improve and adjust the SOPs and guidelines and make the final version of the toolbox.

WP8. [AU] Ethics requirements

- No data collection expected.

The data collection and generation within SOPs4RI forms the basis of the generation of SOPs and guidelines and the development of the toolbox. These are the main outputs of the project, and will be made publicly available in a web-based version on the project's homepage and integrated with SINAPSE.

## 1.2 Types of collected/generated data

Below is a list of the types of data collected by the project, presented by WPs. The lead beneficiary manages data generated by the respective WPs and makes data available to all partners in the consortium using a secure internal platform (Sharepoint) hosted and managed by Aarhus University.

Transcriptions (e.g. expert interviews) will largely be done in Microsoft office software (.doc or .docx), but will also be made available in non-restrictive packages (.txt). Quantitative data (e.g. survey) will be made available in non-restrictive formats as well (.csv or .txt).

WP1. [AU] Management and coordination

 

- No data collection expected.

WP2. [NTUA] Dissemination and Communication

- No data collection expected.

WP3. [MEFST] Systematic review of practices and research cultures

- Literature review gathering existing knowledge on codes of conduct, SOPs and guidelines for promoting research integrity.
- 20 expert interviews with key stakeholders. Four interviews are performed within each of five core stake holder groups: Academia, research integrity committees, funding and process organisations, policy-makers, and industry.
- Delphi Survey is performed among a panel of around 100 stakeholders from RPOs and RFOs, distinct from the expert interview group.

WP4. [STICHTING VUMC] Developing SOPs and guidelines

- Expert interviews with members of the advisory board experienced in drafting SOPs and guidelines.
- Four Co-creation workshop, each including around 25 participants.

WP5. [AU] Focus group interviews

- 32 focus group interviews, 8 from each of four research fields: humanities, social sciences, natural sciences, and medical sciences. Half will include only researchers and the other half will include relevant stakeholders as well.

WP6. [UESSEX] Survey

- Survey data collected form a sample of researchers. This will take the form of a large-scale cross-national survey sampled from researchers in all EU member states and selected OECD countries (likely to be Australia, Canada, and the US).

WP7. [OEAWI] Pilot testing

- Feedback from pilot study of the SOPs and guidelines from eight selected institutions. Target groups are key players within the research community and relevant stakeholders: public RFOs (Austrian and Norwegian research councils: FWF and RCN), private RFOs (La Caixa Foundation and Novo Nordic Foundation), along with RPOs possibly selected from the university network 'the Guild'. Finally, data could include feedback from pilots at research integrity offices selected via ENRIO and EARMA.

WP8. [AU] Ethics requirements

- No data collection

As mentioned above, it should be noted that each of the research WPs (3-7) will develop and deliver a succinct protocol for the research activities. These protocols will be made available at the SOPs4RI project's OSF site: https://osf.io/49fbk/.

## 1.3 Origin of data

This project will generate new data to a larger degree than it utilizes existing. Some WPs will however rely on data generated outside of the project. The list below specifies the origins of all data.

WP1. [AU] Management and coordination
- No data collection expected.

WP2. [NTUA] Dissemination and Communication
- No data collection expected.

WP3. [MEFST] Systematic review of practices and research cultures
- Literature identified in bibliographic databases.
- Recordings of expert interviews.
- Data collected through Delphi rounds.

WP4. [STICHTING VUMC] Developing SOPs and guidelines
- Recordings of expert interviews.
- Records of co-creation workshops.

WP5. [AU] Focus group interviews
- Recordings of focus group interviews.

WP6. [UESSEX] Survey
- Collected survey data.

WP7. [OEAWI] Pilot testing
- Records of pilot tests.

WP8. [AU] Ethics requirements
- No data collection expected.

## 1.4 Size of data

The size of data is unknown at this stage of the project. It will include multiple recordings and transcriptions as well as survey data from cross-national surveys as well as Delphi survey data. The total size in bytes is dependent on the formats in which data is stored, which is decided by the partner institutions.

## 1.5 Data utility

Collected data will be relevant for research performing- and funding organisations. It will inform the tools that are developed in the SOPs4RI project, and can serve to allow

stakeholders to delve into the analyses themselves. Data will also be available for outside actors to perform additional research.

# 2. FAIR data

SOPs4RI follows FAIR data principles[7]. All (non-sensitive) data will be available on the project's OSF site: https://osf.io/49fbk/.

## 2.1 Findable Data

Data generated by the project will be made available and measures will be taken to facilitate the findability of data by outside users. The respective WP leaders are responsible for clearly marking data, and making available necessary documentation, following naming conventions seen under 2.1.3.

### 2.1.1 Discoverability

All data will be available through the project's OSF site. Data and documentation can be discovered through common search engines, the project homepage[8] which links to the OSF site, or the OSF site itself.

Data packages will be given unique identifiers though assignments of unique DOIs for all data packages. This is facilitated by the OSF-platform[9].

Additionally, consistent naming conventions (see 2.1.3) will be followed to aid the discoverability of data packages and related documentations and/or earlier data package versions.

### 2.1.2 Identifiability

Data will be sorted by WP and labelled for easy identification (see 2.1.3 & 2.1.5). WP descriptions, research protocols, and documentation will be available alongside data, enabling outside users to identify the character of shared data.

7 Wilkinson, Mark D., et al. The FAIR Guiding Principles for scientific data management and stewardship. Scientific data, 2016, 3.

8 https://www.sops4ri.eu

9 http://help.osf.io/m/sharing/l/524208-create-dois

### 2.1.3 Naming conventions

All data and documentation files will be named in line with the following template: SOPs4RI_Partner_WP#_DX.X._Title, Version_XX or SOPs4RI_Partner_WP#_DATAX.X_Title, Version_XX. Partner is the abbreviation of the partner institution (e.g. AU), WP# indicates the work package number (e.g. WP1), DX.X indicates the deliverable number (e.g. D1.2) or DATAX.X indicates data associated with a deliverable, and title is the name given to the deliverable or data package by the responsible WP leader. Finally, all documents and data will indicate version number.

### 2.1.4 Search keywords

Data collected by SOPs4RI is found through the online platforms. These are findable through standard search engines. All data packages will be labelled with project and work package names to enable easy finding of related data.

DOIs are also provided through the ODF, allowing users to supply a direct link to data packages.

### 2.1.5 Version labelling

OSF supplies a timestamp to all documents uploaded in the system. As such, original date and changes over time are clearly identifiable. Additionally, version number is indicated by filenames (see 2.1.3), and file history can be viewed at all times in the OSF portal.

### 2.1.6 Metadata

In conjunction with the release of data on OSF, all SOPs4RI WP leaders will provide metadata on data packages generated / curated within their respective WP. These will follow a common scheme informed by the Data Documentation Initiative (DDI)[10]. Information elements will include Title, Principal Investigator, Funding, Bibliographic Citation, Summary, Subject Terms, Geographic Coverage, Time Period, Date of Collection, Unit of Observation, Universe [i.e. Population], Data Type, Sampling, Weights, Mode of Collection, Response Rates, Extent of Processing, and Restrictions.

[10] http://www.ddialliance.org/

## 2.2 Accessible data

All non-sensitive data generated by the project will be openly available for users outside of the project on the OSF platform. Available data can be downloaded by users, and will be accessible through open source software where possible.

WP leaders are responsible for submitting data to the OSF platform, and are responsible for securing that all data is properly labelled, and complies with data protection regulations as well as guidelines set out in this document.

### 2.2.1 Ethical concerns related to data accessibility

Sensitive data will not be made publicly available. Anonymised data will be available through OSF, and contact details for the data responsible partner will be available alongside all data.

Additionally, internal data protection guidelines are outlined in Section 4, and further ethical considerations are elaborated upon in deliverables D8.1 and D8.2, both submitted simultaneously with this deliverable.

### 2.2.2 Software requirements for access

Open Science Framework (OSF.io) is accessible through any internet browser, and data can be downloaded here for external usage. The partners will make all data packages accessible through open source software by providing data in .txt or .csv formats. Documentation will be uploaded in .txt or .pdf-format.

## 2.3 Making data interoperable and reusable

The reusability of the data will be secured by clearly marking data, and supplying documentation on the methods used in collection and preparation of data. Furthermore, the partners strive to limit barriers to use, by prioritizing open access and by clearly marking data-owner. The relevant WP leaders are responsible for ensuring adequate and clear documentation.

All documentation created in data collection will be stored and made available, to the extent that this does not breach data regulations. This includes protocols, codebooks, participant briefings, and questionnaires.

Data produced by the SOPs4RI project will be made accessible using the license procedure offered by OSF, making data openly available and requiring only that the source of the data

be appropriately credited. Licensing is thus managed through the Open Science Framework[11].

### 2.3.1 Increasing data re-use

Non-restricted data will be available to the public as soon as it is ready and quality-assured, and is expected to remain on the OSF platform indefinitely, or until the platform ends the hosting of data.

Providing open licensing, clear labelling, and information on sources of data will secure the ability to obtain related data, and to gain additional knowledge on the data generated by the SOPs4RI partners.

### 2.3.2 Quality assurance

Data quality is the responsibility of the respective WP leaders. Assurance of high quality is maintained by transparency of methods, and open access to generated data and documentation. Research protocols will be made public prior to data generation, allowing users to track and evaluate data quality.

The overall quality assurance mechanisms of the SOPs4RI project including training in relation to methods, data handling, and analytical approaches are described in Deliverable 1.1 “Research Integrity and Quality Assurance Plan”.

11 http://help.osf.io/m/bestpractices/l/611430-licensing

# 3. Allocation of resources

No additional costs will be associated with making the generated data FAIR. All (eligible) data will be available on the OSF site (https://osf.io/49fbk/) as a part of the project. As such, measures to make the data available and useable are incorporated into the project plan.

# 4. Data security

Each partner-organisation in SOPs4RI is responsible for complying with the general data protection regulation (GDPR)[12] when collecting data within the WPs that they participate in.

Additionally, the consortium commits to following the general guidelines:

- Collected data will be stored for a period of five years after the last publication. This includes original audio-visual files, transcriptions, signed consent forms and questionnaires. Participants must be informed of this practice in consent forms.
- Only anonymised data will be used for analysis, and will be made openly available through the Open Science Framework.
- Security and timely deletion of locally stored sensitive data is the responsibility of the relevant WP leader.

The secure Sharepoint platform constitutes the primary storage for data, including non-anonymised data. Partners will safely handle and store data, e.g. audio-files and transcriptions, locally immediately in relation to data collection / cleaning / analyses. When these are transferred to Sharepoint folders, data will be deleted from local systems. The exact timing for transferring data to Sharepoint will be decided in collaboration with the WP leader.

In case of a data breach, affected participants will be contacted and data will be temporarily removed from the compromised storage. All internal transfer of sensitive data will be done though secure pathways. Specifically, the secure Sharepoint work space established for the SOPs4RI project will be used for data transfer. This workspace is hosted by Aarhus University, and the access to the Sharepoint portal is managed by the SOPs4RI team at Aarhus University.

Long-term data preservation will comply with GDPR regulations. It is the responsibility of the WP leaders to ensure that sensitive data is secured, and is deleted in accordance with the general guidelines.

All partner-organisations confirm that they meet GDPR regulations and have taken required data protection measures. Deliverable D8.2 provides statements of compliance from the Data Protection Officers or relevant organisational representatives at each partner-organisation involved in data collection.

[12] See: http://data.europa.eu/eli/reg/2016/679/oj

## 4.1 Protection of personal data

During the project, the partner-organisations will collect and share data with the other project partners. All data including personal data will be shared and stored on a secured platform.

The partner-organisation that collects personal data for the project warrants that it has obtained the personal data with due observance of the rights of the registered persons and in compliance with all applicable laws and regulations.

The partner-organisation that contribute with personal data to the other partners warrants that it has the authority to disclose the personal data to the other partners. As each of the partners to the Consortium Agreement receives the personal data for self-standing research purposes, they will all be individually Data Controller.

When disclosing personal data, the disclosing partner-organisation must follow the rules and procedures for disclosure of personal data at the partner's own institution. All partners must register the research activities that involves personal data according to the rules and procedures at their own institution.

During the project, the partners process personal data. As each partner is data controller for their own data processing, they are responsible for their own data processing. As all partner-organisations involved in data collection are data controllers, no specific Data Protection Agreement between data controllers and data processors will be needed. Hence, the document at hand alongside deliverable D8.2 (Data Protection Officers' or organisational representatives' statements) replaces the envisaged Data Protection Agreement.

# 5. Ethical considerations

A large group of individuals will be surveyed, interviewed or recorded during the SOPs4RI data collection process. This motivates clear research protocols that specify which data might be sensitive, and clearly specifies the process of gathering consent for required information gathering.

Such actions will be taken in the preparation of data gathering and the processing and publication of data packages. The “data security” chapter describes the principles guiding this process, and the actions taken in the event of breaches of security.

The responsible WP leader will submit the WP study for ethical assessment. Procedures to ensure that ethical concerns are duly addressed are also described in the SOPs4RI Grant Agreement and in deliverable D1.1 “Research integrity and quality assurance plan”.

Specific procedures and criteria used to identify and recruit research participants are preliminarily reported in deliverable D8.1, and will be described in further detail in the protocols for each empirical WP.

AARHUS UNIVERSITY

VUmc

SVEUČILIŠTE U SPLITU
MEDICINSKI FAKULTET

University of Essex

ÖSTERREICHISCHE AGENTUR FÜR WISSENSCHAFTLICHE INTEGRITÄT

Universiteit Leiden

Health Research Board

KATHOLIEKE UNIVERSITEIT LEUVEN

LSE The London School of Economics and Political Science

EARMA

UNIVERSITAS VARSOVIENSIS

SOPs4RI

www.sops4ri.eu @sops4ri SOPs4RI Project @sops4ri

The project leading to this application has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 824481.