



Grant Agreement number: 777563
Project acronym: RI-PATHS
Project title: Research Infrastructure imPact Assessment paTHwayS
Type of action: Coordination and Support Action (CSA)

Deliverable 2.2
Title: Data Management Plan



Deliverables leader:	EFIS
Authors:	Alasdair Reid, Elina Griniece, Ildiko Ipolyi, Emily De Young Becker, Johannes Gutleber, Corinne Martin
Due date:	28 February 2018
Actual submission date:	1 August 2018
Dissemination level:	Public

Abstract: This deliverable is the first version of the RI-PATH Data Management plan (DMP). It contains an initial description of the datasets (to be) collected, processed or generated by the project and an initial plan on how sharing, archiving and preservation of these datasets will be guaranteed.





Document Revision History

Date	Version	Author/Editor/Contributor	Summary of main changes / Status
18.03.18	0.1	Alasdair Reid	Report structure and draft content
26.03.18	0.2	Elina Griniece	Reformatting the report and further content
29.03.18	0.3	Elina Griniece	Drafting content
16.04.18	0.4	Emily De Young Becker	Drafting content
25.05.18	0.5	Jean-Claude Worms	Validating/translating lawyer's edits
28.05.18	0.6	Ildiko M. Ipolyi	GDPR and MERIL related contents
29.05.18	0.7	Johannes Gutleber	Comments on the draft version
15.06.18	0.8	Corinne Martin	Comments on the draft version
27.07.18	0.9	Elina Griniece	Uptake of comments and revision of the document
01.08.18	1.0	Alasdair Reid	Final review and approval.

Disclaimer

The sole responsibility for the content of this publication lies with the authors. It does not necessarily reflect the opinion of the European Commission. The European Commission is not responsible for any use that may be made of the information contained therein.

Copyright

This document may not be copied, reproduced, or modified in whole or in part for any purpose without written permission from the RI-PATHS consortium. In addition, an acknowledgement of the authors of the document and all applicable portions of the copyright notice must be clearly referenced.

All rights reserved.

This document may change without notice.



Executive Summary

This deliverable is the first version of the RI-PATHS Data Management Plan (DMP). It contains an initial description of the datasets (to be) collected, processed or generated by the project and an initial plan on how sharing, archiving and preservation of these datasets will be guaranteed.

The DMP is intended to be a living document and thus the content of this deliverable will be updated and described in a finer level of granularity as the implementation of the project progresses in order to reflect adaptations and changes in the prepared plan.



Contents

1	INTRODUCTION	1
2	DATA SUMMARY	2
3	FAIR DATA	6
3.1	Making data findable, including provisions for metadata	6
3.2	Making data openly accessible.....	6
3.3	Making data interoperable.....	7
3.4	Increase data re-use (through clarifying licences).....	7
4	ALLOCATION OF RESOURCES	8
5	DATA SECURITY	8
6	ETHICAL ASPECTS.....	9



Acronyms

Abbreviation	Meaning
CERIF	Common European Research Information Format
CSA	Coordination and Support Action
D	Deliverable
DMP	Data Management Plan
DOI	Digital Object Identifier
GDPR	General Data Protection Regulation
GPL	General Public Licence
IA	Impact Assessment
MERIL	Mapping of the European Research Infrastructure Landscape
RI	Research Infrastructure



Short names of the consortium partners

Name of the consortium partner	Short name
European Future Innovation System Centre	EFIS
Centro Studi Industria Leggera	CSIL
Fondation Européenne de la Science	ESF
Fraunhofer Gesellschaft zur Förderung der angewandten Forschung e.V. for and on behalf of its Fraunhofer-Institut für System- und Innovationsforschung ISI	Fraunhofer ISI
European life science infrastructure for biological information	ELIXIR
European Organization for Nuclear Research	CERN
Consorcio para la Construcción, Equipamiento y Explotación del Laboratorio de Luz Sincrotrón	ALBA
Deutsches Elektronen-Synchrotron	DESY

1 Introduction

The RI Impact Pathways (RI-PATHS) project is participating in the Pilot on Open Research Data in Horizon 2020, in line with the Commission's Open Access to research data policy. The aim is to facilitate access, reuse and preservation of research data. For this purpose, the RI-PATHS project consortium has developed a Data Management Plan (DMP)¹ that outlines how research data will be handled during and after the project, describing what data will be collected, processed or generated and following what methodology and standards, whether and how this data will be shared and/or made open, and how it will be curated and preserved. The DMP describes:

- a. How the project will deposit in the institutional data repositories of our partners the project's publications and research datasets
- b. How the project coordinator will monitor, track and disseminate information about the produced research publications and datasets to the relevant channels
- c. How the project will take measures to make it possible for third parties to access, mine, exploit, reproduce and disseminate (free of charge for any user) the following:
 - the data, including associated metadata, needed to validate the results presented in scientific publications as soon as possible
 - other data, including associated metadata
- d. Descriptions, where relevant, of methods and software applications used to analyse specific datasets and validate the results.
- e. Guidelines, support material, proposed workflows and tools for both the coordinator and each project partner, translating the generic requirements of the Open Access & Open Research Data Pilot into specific practical guidelines that they can apply during the lifetime of the project. To this end, additional guidance notes will be circulated by the coordinator during the lifetime of the project to help project partners to comply with open access obligations.

The aim of the RI-PATHS project is to develop a model describing the socio-economic impact of research infrastructures (RIs) and of their related investments. The model will be developed in a modular manner adapting it to a broad range of scientific domains and types of infrastructures. The project outcomes are expected to contribute to a common approach at international level and facilitate investments in research infrastructures by funding agencies and other stakeholders.

The project will carry out a comprehensive stocktaking exercise on the existing approaches for impact assessment of RIs and map the current and future data gathering needs of the key stakeholder groups. The goal is to identify the strengths and limitations of the existing approaches, detect existing gaps and overlaps, as well as to assess the level of complexity of the developed conceptual advances. Moreover, a set of activities will be devoted to mapping the needs of various stakeholders associated with Impact Assessment (IA) of RIs. Employing participatory and systems thinking approaches, the project will develop a

¹ The DMP has been constructed based on the H2020 guidance and template :

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

modular impact assessment model that represents all major impact pathways of distinct types of research infrastructures.

The IA model will be put to practice by defining a set of reference indicators, providing guidance on the most appropriate monitoring and evaluation approaches and testing its feasibility with pilot research infrastructures. The outcomes of the validated model will be integrated with the MERIL platform (<https://portal.meril.eu/meril/>), an online source of scientific and policy relevant data on European RIs. One of the key objectives of the project is to ensure a user-friendly utilisation of the IA model by developing a dedicated web-based tool for the IA model, including KPIs, financial information, human resources, etc.

2 Data summary

What types and formats of data will the project generate/collect?

The different kinds of data that will be collected and managed by the platform are the following:

Platform content, i.e., data collected managed and stored via the project website or the related MERIL platform. In a first stage, the content will mainly concern information about the four pilot RI socio-economic impact, specific performance indicators, etc. In a second stage, the socio-economic impact assessment module developed for the MERIL platform will collect data on a broader range of RIs.

Project output, i.e., any dataset collected (e.g. via survey questionnaires and interviews) and any dataset used for documenting and providing evidence for either a dissemination report or a publication produced in the context of project activities. These datasets are collections of standard material produced by a research project, e.g. deliverables, dissemination material, training material, scientific publications, etc.

Software, i.e., datasets resulting from the software enabling the RI-PATHS website or the related MERIL platform. These datasets are mainly software artefacts and source code which can be used for various purposes including research tasks.

More specifically, the following tasks are likely to lead to data collection, analysis and storage:

- Task 2.1 Project management and coordination (responsible EFIS)
 - Data on project participants (mailing lists, etc.) - internal use only;
- Task 2.5 Advisory Board (responsible EFIS)
 - Data on Advisory Board members (mailing list, address, affiliation and bank details for reimbursement) - internal use only;
- Task 3.1 Definition of relevant RI typology (responsible CSIL)
 - The project will make use of the existing datasets collected by the participating RIs and by the MERIL-2 project. The MERIL (Mapping of the European Research Infrastructure Landscape) portal provides access to a database that stores information about openly accessible research infrastructures (RIs) in Europe, across all scientific domains. MERIL has committed to the RIs that participate in MERIL that the information provided in this section will remain internal to the MERIL-2 project and may only be used in the framework of statistical and anonymised analysis. In addition, no information indicating the relation between financial data and a specific RI may be published without prior explicit permission from the RIs.

- Task 3.2 State of play - literature review (responsible CSIL)
 - Literature review may include certain documents of a ‘grey’ (non-published) status (e.g. evaluations or impact assessment reports) collected from partners - use will be governed by the status of documents as communicated by owner;
- Task 3.3 Identification of needs - survey/interviews/consultation (responsible CSIL)
 - The survey will collect opinions and data on a broad range of research infrastructures. The data collection will concern institutional rather than personal (individual) data, however, certain key characteristics (name, position, e-mail, etc.) of respondents will also be saved for follow-up or statistical purposes. MERIL contact database will be used to disseminate the survey link. Personal data such as contact name or contact e-mail will never be communicated outside the MERIL-2 project. Any communication to the MERIL distribution list will be handled directly by ESF (MERIL-2 coordinators). The survey will be run using SurveyMonkey data collection platform. Appropriate observation of European (EU General Data Protection Regulation 2016/679) and national data privacy regulations will be ensured.
 - Interviews will be conducted with selected individuals to further develop specific survey findings or to complete information gaps. Opinions or statements of interviewees will not be quoted so as to be attributable unless express prior permission is obtained. Data use will be governed by the procedures set out in this document, in particular only metadata will be used in public reports;
- Tasks 4.1 1st round of participatory workshops - impact pathways and logical models (responsible EFIS) and 4.3 2nd round of participatory workshops - KPIs and metrics (responsible Fraunhofer ISI)
 - Data collected on the participants attending workshops will be limited to required registration data (name, organisation, position, experience with IA, etc.). Participants authorisation will be sought before listing publicly their names as attendees to a workshop. This information will be kept until the end of the project. Appropriate observation of European (EU General Data Protection Regulation 2016/679) and national data privacy regulations will be ensured. Workshop reports will avoid attributing statements or opinions to individual participants;
- Task 5.1 Draft IA model report and validation workshop (responsible Fraunhofer ISI)
 - Data collected on the participants attending the validation workshop will be limited to the required registration data (name, organisation, position, experience with IA, etc.). Participants authorisation will be sought before listing publicly their names as attendees to the workshop. This information will be kept until the end of the project. Appropriate observation of European (EU General Data Protection Regulation 2016/679) and national data privacy regulations will be ensured. The workshop report will avoid attributing statements or opinions to individual participants;
- Task 5.2 Pilot IA projects with selected RIs (responsible ESF)
 - During the pilot projects, data collected and analysed from one of the participating research infrastructures will be handled in accordance with the confidentiality status set by the RI itself as stipulated in the Attachment 1 of the

Consortium Agreement (see Table 2). Only metadata will be made public and only with the express authorisation of the RI legal representative;

- Task 5.3 Integration of model into MERIL advanced data module (responsible ESF)
 - The RI-PATHS project plans to bring modifications to the Advanced Data Module that concerns RI’s socio-economic impact indicators and collect data via the MERIL-2 portal. The activities of the task will not initially collect new data, however the new module to be integrated into MERIL platform will build on and extend the existing database and lead to the collection of new data. Any data collected via the MERIL-2 portal, even within a new set of data fields, must also be subject to the moral and contractual agreements that exist between MERIL and its participating RIs concerning data privacy. Any planned sharing of this data outside of the MERIL-2 project must be clearly communicated to RIs participating in MERIL. Data may be shared or used according to the conditions stipulated in Attachment 1 of the RI-PATHS Consortium Agreement (see Table 1). Appropriate data management and security procedures will be developed as part of a preceding Task 5.2 Pilot IA model with selected RIs and Task 5.3;
- WP6: Outreach, dissemination and exploitation (responsible EFIS)
 - The project’s dissemination and communication activities may lead to the compilation of additional mailing lists (e.g. newsletter subscriber list) and participant lists (to events). Appropriate observation of European (EU General Data Protection Regulation 2016/679) and national data privacy regulations will be ensured.

Will you re-use any existing data and how? What is the origin of the data?

Table 1: ESF background of specific limitations and/or conditions for implementation of the Grant Agreement

Background	Specific limitations and/or conditions for implementation (Article 25.2 Grant Agreement)	Specific limitations and/or conditions for exploitation (Article 25.3 Grant Agreement)
<p>ESF has background related to the MERIL-2 project, including the MERIL web portal, database, data model and data collection procedures.</p>	<p>ESF may deny any and all requests from consortium partners, or any other party, for communication of data included in the MERIL database, when the request is:</p> <ul style="list-style-type: none"> - Against any laws (national or European, including but not limited to Act N° 78-17 of 6 January 1978 on Information Technology, Data Files, and Civil Liberties and the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data) that are in effect at the date of the request; - Contrary to any commitments or obligations, moral or contractual, made by ESF, notably but not exclusively within the framework of the MERIL-2 project. 	<p>ESF will consider requests for communication of data included in the MERIL database only when the request is “fair and reasonable” as stipulated by Article 25.3 of the Grant Agreement.</p> <p>In any event, ESF may choose to only communicate information</p> <ul style="list-style-type: none"> - In the form of samples - Using aggregated and/or anonymized data - Using banded values.

Table 2: CERN background of specific limitations and/or conditions for implementation of the Grant Agreement

Background	Specific limitations and/or conditions for implementation (Article 25.2 Grant Agreement)	Specific limitations and/or conditions for exploitation (Article 25.3 Grant Agreement)
<p>Historical procurement data in anonymised, statistical form.</p> <p>Historical human resources, education and cultural goods data in anonymised, statistical form.</p> <p>Historical information about technological spillovers and Knowledge Transfer impacts.</p> <p>Historical CAPEX and OPEX information going beyond the publicly available data on CERN’s Document Server (CDS) concerning the LHC, HL-LHC and injector complex.</p> <p>Information produced in the scope of collaboration agreements with University of Milano (Italy), contract KE 3044 and Centre for Industrial Studies (Italy), contract KE 3570.</p> <p>Cost estimates for a concept of a post-LHC particle collider research infrastructure.</p> <p>Results from previous and ongoing CBA studies concerning CERN, the LHC, HL-LHC and a concept for a post-LHC particle collider research infrastructure.</p>	<p>Information will only be provided in anonymised and statistical form. Some information may not be provided due to confidentiality obligations, or provided only under confidentiality undertakings.</p> <p>Anecdotal evidence provided by CERN in the scope of this project, which is recorded in any form by RI-PATHS consortium member must be verified and released by CERN for further use in the scope of this project.</p>	<p>Use of background made available by CERN for exploitation is subject to restrictions imposed by the rights of third parties, including those of personnel or former personnel.</p>

What is the expected size of the data?

At this stage, it is difficult to quantify the exact scale of data collected. The survey will target up to 1000 potential respondents with the aim to collect at least 100 responses.

The MERIL module, which will be developed at the end of the project, should enable the collection of IA related data directly from the RIs and on a large scale, potentially covering more than 1000 RIs.

To whom might it be useful ('data utility')?

The data collected are likely to be useful to at least the following categories of users:

- Policy-makers and funders
- Academic researchers and other RI service users
- Research infrastructure managers
- Evaluators, consultants and other experts.

3 FAIR data

FAIR data efforts will take into account data privacy requirements. In particular, personal data will be treated confidentially and in compliance with the EU General Data Protection Regulation 2016/679.

3.1 Making data findable, including provisions for metadata

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

What naming conventions do you follow?

Will search keywords be provided that optimize possibilities for re-use?

Do you provide clear version numbers?

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

All of the data, documents and code underlying all RI-PATHS will be accompanied by standard general metadata and receive a DOI (Digital Object Identifier) number. Access to the project's output will be provided through the RI-PATHS project site (www.ri-paths.eu). Moreover, deliverables and project results (journal articles, videos, posters, etc.) that all partners have agreed should be made publicly available will be deposited and be made available via OpenAIRE Zenodo and/or the beneficiaries' institutional repositories. Scientific publications will be disseminated through the RI-PATHS project website as well as through scholarly communication channels, e.g., publishers/journals web sites, institutional repositories, scholarly communication networks (ResearchGate, Google Scholar).

3.2 Making data openly accessible

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

What methods or software tools are needed to access the data?

Is documentation about the software needed to access the data included?

Is it possible to include the relevant software (e.g. in open source code)?

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

Have you explored appropriate arrangements with the identified repository?

If there are restrictions on use, how will access be provided?

Is there a need for a data access committee?

Are there well described conditions for access (i.e. a machine readable license)?

How will the identity of the person accessing the data be ascertained?

The web-based tool developed under WP5: Development and implementation of modular IA model will be available as open source and made accessible through the collaboration, code review, and code management GitHub repository for open source projects (<https://github.com> and/or Zenodo). It will also be empowered by a publicly accessible issue tracker tool. Moreover, this offers the possibility to use the outcomes of the project to define the metadata and the corresponding definitions with the final aim to implement the model into the MERIL CERIF compliant model.

Although the project is a CSA, we believe that the production of technical papers, presenting the methodology for the proposed RI-PATHS impact assessment model will contribute largely to its sustainability and overall impact within the RIs community. Technical papers and policy related findings may be published in scientific journal or conference papers; either in a Gold or Green publishing mode. Green Open Access publications will be deposited at least in one repository and in a publicly available Open Access repository (e.g. Zenodo). Depending on the deliverable dissemination level, deliverables will be publicly available via the project website, while others may be restricted. Scientific papers and publications will be available/accessible via the Publishers' web site according to the associated access method. To ensure Open Access a machine-readable electronic copy of every publication will be deposited in the OpenAIRE Zenodo repository. In the case of surveys containing any kind of personal data, only anonymised versions of the datasets will be openly available; raw data will be archived in closed or restricted access.

3.3 Making data interoperable

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

Will you be using standard vocabularies for all data types present in your data set, to allow inter-disciplinary interoperability?

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies?

Interoperability of the data collected by surveys, deliverables and publications will be achieved by the standard metadata required by OpenAIRE (<https://www.openaire.eu/>). The MERIL-2 database is built on the CERIF model (a European standard for RI related data), which facilitates interoperability. The interoperability method that has been retained for MERIL-2 is the exchange of files. Two formats may be used: Microsoft Excel files and CERIF-XML files. The data exchange templates and the guidelines for filling them out are available on the MERIL portal at portal.meril.eu. Complete details on MERIL interoperability can be found in MERIL-2 Deliverable 3.4 Interoperability Guidelines.

3.4 Increase data re-use (through clarifying licences)

How will the data be licensed to permit the widest re-use possible?

When will the data be made available for re-use? If an embargo is sought to give time to publish or seek patents, specify why and how long this will apply, bearing in mind that research data should be made available as soon as possible.

Are the data produced and/or used in the project useable by third parties, in particular after the end of the project? If the re-use of some data is restricted, explain why.

How long is it intended that the data remains re-usable?

Are data quality assurance processes described?

The public content made available via the project website will be available for download and re-use with no restrictions or embargo. The content will be available under permissive licences, (CC-BY 4.0, CC-0 or comparable²) but certain conditions (e.g. Non-commercial use=NC) and/or exceptions may be applied (and will be identified in later versions of this DMP).

² <https://creativecommons.org/licenses/>

At the current stage of the project, the use of General Public Licence (GPL/A) is planned for software developed. However, this issue will be revisited in the context of the MERIL implementation of the RI-PATHS IA assessment model.

In the cases where specific service information cannot be publicly shared, the reasons will be mentioned in their metadata descriptions (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

Concerning the time for which the data will remain re-usable, the platform will be maintained, and the content will be reusable, for at least three years after the end of the project. As noted above, all material and data produced or collected by the project will also be stored on Zenodo.

The definitive description of data reuse and collection, and specific licensing as concerns the module integrated in MERIL, will be agreed by partners following finalisation of the RI-PATHS IA model and its corresponding implementation on the MERIL portal.

4 Allocation of resources

What are the costs for making data FAIR in your project?

How will these be covered? Note that costs related to open access to research data are eligible as part of the Horizon 2020 grant (if compliant with the Grant Agreement conditions).

At the current time, we do not foresee specific additional costs beyond the described processes and methods for data collection and storage we have set out in the DMP and which fall within the activities covered by the current grant. Later versions of this DMP may identify additional costs.

Who will be responsible for data management in your project?

The responsibility for managing data underlying RI-PATHS research activities will lie with the authors of the individual research studies or deliverables. The coordinator will, however, request that all authors comply with principles of this DMP and OpenAIRE procedures including the storing of data on the Zenodo repository categorised as an outcome of the RI-PATHS project. ESF can comply with the above principles under the condition that those principles ensure a sufficient level of security with regard to sensitive data in the MERIL database.

Are the resources for long term preservation discussed (costs and potential value, who decides and how what data will be kept and for how long)?

5 Data security

What provisions are in place for data security (including data recovery as well as secure storage and transfer of sensitive data)?

Is the data safely stored in certified repositories for long term preservation and curation?

Concerning data security, all data collected content will be archived in a secure manner, via regular database backups, in the project's collaborative platforms (server located in the EU) that will serve as the main backup area. Preservation and back-ups of the data are ensured by the EFIS Centre data centre preservation policies: daily backups, security updates, securely controlled administrative access. Moreover, for long term preservation, data access and reuse after the end of the project, the datasets will be available via Zenodo. Backup of the collected data will be implemented, based on processes that ensure a sufficient level of security of sensitive data in the MERIL database.

Deliverables and project material will be primarily stored on the project's collaboration server (operated by EFIS). Partners will deposit, in an OpenAIRE compliant repository, the published version or the final peer-reviewed manuscript accepted for publication. Authors

may use on their institutional repositories (if any) as well as on Zenodo. Moreover, a copy of each deliverable and open access papers will be uploaded to the project website.

6 Ethical aspects

Are there any ethical or legal issues that can have an impact on data sharing? These can also be discussed in the context of the ethics review. If relevant, include references to ethics deliverables and ethics chapter in the Description of the Action (DoA).

A priori, there are unlikely to be major ethical issues related to the data collected and shared, since personal data will not be the subject of any research carried out by the project. A further review of ethical issues related to data will be carried out in the Ethics deliverable (D1.1).

Is informed consent for data sharing and long-term preservation included in questionnaires dealing with personal data?

Ethical and legal issues will be addressed in the following way. In case of datasets containing personal data (e.g. from surveys), only anonymised data will be made publicly available. In case of datasets containing data bound by specific agreements that prohibit further data dissemination, data will not be made available (although metadata may be made available). In accordance with legal restrictions (e.g. GDPR), personal data concerning RI-PATHS website users or participants to interviews, surveys or workshops/events will not be shared in any manner.