

Project Title	Expanding FAIR solutions across EOSC
Project Acronym	FAIR-IMPACT
Grant Agreement No.	101057344
Start Date of Project	2022-06-01
Duration of Project	36 months
Project Website	fair-impact.eu

1 D1.2 - Data Management Plan

Work Package	WP 1, Project management, synchronisation and sustainability
Lead Author (Org)	Daniel Turner (KNAW-DANS), Vasso Kalaitzi (KNAW-DANS), Emilie Kraaikamp (KNAW-DANS)
Contributing Author(s) (Org)	Joy Davidson (DCC), Jessica Parland-von Essen (CSC), Sara Pittonet (TRUST-IT), Marjan Grootveld (KNAW-DANS), Clement Jonquet (INRAE).
Due Date	2022.11.30
Date	2022.11.30
Version	V1.0 DRAFT NOT YET APPROVED BY THE EUROPEAN COMMISSION

Dissemination Level

<input checked="" type="checkbox"/>	PU: Public
<input type="checkbox"/>	PP: Restricted to other programme participants (including the Commission)
<input type="checkbox"/>	RE: Restricted to a group specified by the consortium (including the Commission)
<input type="checkbox"/>	CO: Confidential, only for members of the consortium (including the Commission)



2 Versioning and contribution history

Version	Date	Author	Notes
0.1	2022.08.05	Daniel Turner (KNAW-DANS)	TOC and V0.1
0.2	2022.09.05	Marjan Grootveld (KNAW-DANS), Emilie Kraaikamp (KNAW-DANS), Jessica Parland-von Essen (CSC)	Review comments
0.3	2022.09.22	Vasso Kalaitzi (KNAW-DANS)	Review, edits and additions;
0.4	2022.09.28	Daniel Turner (KNAW-DANS)	Text revisions and comment resolution
0.5	2022.11.21	Joy Davidson (DCC), Jessica Parland-von Essen (CSC), Sara Pittonet (TRUST-IT), Marjan Grootveld (KNAW-DANS), Clement Jonquet (INRAE)	EXC input and review
0.6	2022.11.22	Vasso Kalaitzi (KNAW-DANS)	Addressing comments and incorporating input
0.7	2022.11.25	Ingrid Dillo (KNAW-DANS)	Final review
1.0	2022.11.30	Vasso Kalaitzi (KNAW-DANS)	Final version

Disclaimer

FAIR-IMPACT has received funding from the European Commission's Horizon Europe funding programme for research and innovation programme under the Grant Agreement no. 101057344. The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.

3 Table of Contents

TERMINOLOGY	4
EXECUTIVE SUMMARY	5
1 INTRODUCTION	6
2 DATA SUMMARY	7
3 FAIR DATA AND OTHER RESEARCH OUTPUTS	12
4 ALLOCATION OF RESOURCES	13
5 DATA SECURITY	13
6 ETHICS	14
7 OTHER ISSUES	15

3.1 List of Tables

TABLE 1 – LIST OF ALL DATASETS TO BE GENERATED AND PUBLISHED BY FAIR-IMPACT WITH VOLUME INDICATION	8
--	---

TERMINOLOGY

Terminology/Acronym	Description
CSA	Coordination and support action
CSO	Citizen science organisations
DI	Data infrastructures
DMP	Data management plan
DSP	Data service providers
EFC	EOSC FAIR Champions
EOSC	European Open Science Cloud
EXC	Executive Committee
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
HLAC	High Level Advisory Committee
IIS	Individuals in Science
IP	Intellectual Property
NLI	National level initiatives
PCO	Project Coordination Office
PID	Persistent identifier
RFO	Research funding organisation
RI	Research communities and infrastructures
RPO	Research-performing organisation
SF	Synchronisation Force
SRIA	Strategic research and innovation agenda
SS&A	Scientific societies and academies

Executive Summary

This deliverable represents the FAIR-IMPACT Data Management Plan (DMP) and is under the responsibility of the project's Project Coordination Office. It contains a description of how the data generated, collected and processed will be managed by the FAIR-IMPACT project. These data may derive from and/or be linked to research outputs, dissemination and communication activities, engagement activities and other means depicted in the present deliverable. The DMP includes an overview of how sharing, archiving, and preserving these data will be guaranteed in the course of the project and afterwards.

The DMP includes information on data ethics and data protection. FAIR-IMPACT contributes to Open Access, Open Science and the European Open Science Cloud (EOSC) and handles its data responsibly and according to the FAIR principles and the conditions laid out in the relevant articles of the Grant Agreement and Consortium Agreement.

The deliverable is intended to be a living document and thus its content will be updated and described in more detail as the project progresses, in order to reflect adaptations and changes with respect to its first version. The followed structure is based on the Horizon Europe Data Management Plan template (v1.0, 5 May 2021).

1 Introduction

The overall objective of FAIR-IMPACT is to realise a FAIR EOSC, that is an EOSC of FAIR data and services, by supporting the implementation of FAIR-enabling practices across scientific communities and research outputs at a European, national, and institutional level. The consortium comprises 22 partners coordinated by KNAW-DANS, with efforts laid out in the Grant Agreement No. 101057344. The project is FAIR-IMPACT identifies the optimal components for enabling FAIR (practices, policies, tools, and technical specifications) and defines the support, governance, and coordination mechanisms needed to ensure their adoption and continued utility in a FAIR EOSC. At the same time the project fosters alignment for interoperability and focuses on persistent identifiers (PIDs), metadata, ontologies, metrics, certification, and interoperability.

FAIR-IMPACT promotes and provides support to embed FAIR components and mechanisms across domains, geographic areas, and stakeholder groups at multiple levels. The project focuses on repositories, data and metadata service providers, research-performing organisations (RPOs) and infrastructures, and national-level initiatives and communities. It further ensures new stakeholders are reached and establishes wider engagement and contribution to EOSC.

FAIR-IMPACT also ensures that governance and coordination mechanisms are aligned to the Strategic Research and Innovation Agenda (SRIA) objectives by engaging with relevant EOSC Task Forces. Building on the successful Synchronisation Force (SF) approach from the FAIRsFAIR project, FAIR-IMPACT establishes dialogue for collaboration and harmonisation of efforts between FAIR-IMPACT and other EOSC and FAIR-related initiatives. This aims to reduce redundancy and ensure that solutions are more widely promoted, sustainable, and can be transferred to the relevant EOSC Partnership and current and future EOSC stakeholders.

Integrated use-cases play a significant role in defining requirements, identifying gaps and implementation opportunities across the project's focus areas. These use-cases cover four scientific areas and form the first support Tier of the project. Three more Tiers of support that are crucial for the project's implementation and impact are: Tier 2: targeted adoption and implementation funding via cascading grants, Tier 3: In-kind support programs and provision of assessment frameworks and examples, and Tier 4: Supporting EOSC-readiness through harmonisation.

Despite the focus on FAIR research data, the project itself, being a Coordination and Support Action (CSA), does not intend to generate large amounts of research data or any types of research outputs, as that is not part of its methodology. Examples of research data that

FAIR-IMPACT may generate include data service tools, interviews, standards, semantic artefacts, and policy documents. This Data Management Plan describes how the project will handle such data.

At the same time, the project will collect a substantial amount of administrative data that is needed for running the project or for communication with external partners like research communities and research funders. Administrative data is included in the data described in this DMP only when it concerns personal data. The handling of such data is discussed in the Section “Ethical aspects”.

2 Data summary

In order to support the expansion of FAIR solutions, EOSC projects build a web of FAIR data and related services for open science together with the scientific community and relevant stakeholder groups that have been identified by respondents to the SRIA consultation. To support this, FAIR-IMPACT will collect and manage different types of data. Datasets will be collected through desk research, interviews, workshops, and use case studies, as well as any other procedures linked to the four support tiers, which will be used for analyses, recommendations, and reports. These datasets are collections of standard material produced by a research project, e.g., project deliverables, dissemination material, and training materials.

The data collected in FAIR-IMPACT are intended to be useful for (meta)data service providers (DSP) including repository and PID services, national-level initiatives (NLI), research communities and infrastructures (RI), research-performing organisations (RPOs), individuals in science (IIS), research-funding organisations (RFOs), policymakers, publishers, data infrastructures beyond research (DI), the EOSC ecosystem, the FAIR ecosystem, citizen science organisations (CSOs), and scientific societies and academies (SS&A).

More specifically, the following activities will lead to data collection, analysis, storage, and publication:

- Deployment of a range of governance, coordination, and collaboration mechanisms necessary for a functioning FAIR research ecosystem, including the establishment of relevant bodies, such as the High level Advisory Committee (HLAC) and EOSC FAIR Champions (EFC).
- Creation of FAIR implementation action plans and organisation of workshops to coordinate and support the adoption of FAIR-enabling components and mechanisms at multiple levels among a diverse group of stakeholders.

- Implementation of a PID support programme and guidance for the provision of PID services in EOSC, including mechanisms and components to facilitate adoption and implementation of the EOSC PID Policy and alignment with PID practices.
- Designing a semantic framework for the governance, creation, mapping, sharing, reuse, FAIRness assessment, and interoperability of semantic artefacts for EOSC. This includes interoperable and harmonised semantic artefact catalogues, crosswalks, and mappings to improve data findability and accessibility within and across disciplines.
- Creation of guidelines and prototype(s) which support both enabling FAIR and assessing FAIRness and trustworthiness of different types of digital objects in multiple domains.
- Making available components and services for increased legal, organisational, semantic, and technical interoperability in EOSC within and across disciplines.
- Organising workshops, including the Synchronisation Force workshops, aiming to establish and maintain the dialogue for collaboration and harmonisation of efforts between FAIR-IMPACT and other EOSC and FAIR related initiatives.
- Creation and dissemination of FAIR implementation stories providing examples, guidance, workflows, and inspiration for research communities and stakeholders.
- Registration through the FAIR-IMPACT website for events and other participatory activities.

The estimated volume of the data that will be archived and published is less than 50 GB. This consists of the following:

Table 1 – List of research outputs to be generated by FAIR-IMPACT

Content /Origin	Retention	Format	Personal information	Dissemination level	Comments
Surveys run on the website	FAIR-IMPACT website database	MySQL database, html	Available to data processors only	Anonymised survey results will be made public	
Surveys on third party tools (mentimeter, Sli.do or similar) at workshops	The results are downloaded, shared as csv/excel with data processors and published as pdfs on Zenodo	csv/excel, pdf	In principle, no personal information will be collected. In cases personal information is included,	Anonymised survey results will be made public	

Content /Origin	Retention	Format	Personal information	Dissemination level	Comments
			the results will be anonymised before publication		
Results of the Open Calls	<p>The name of the organisation for successful applicants will be related and shared through project reports and the website.</p> <p>Personal Data processed for Applications will be kept for up to 5 years, according to FAIR-IMPACT privacy policy.</p>	MySQL database;	Applicants' name, role, organisation and email address will be collected when applying to the open calls. For successful applicants, organisational bank information will be needed to provide payment to support participation. Successful applicant's organisations will be made public via the website	The name of the organisation for successful applicants will be related and shared publicly through project reports and the website	Contact details may only need to be retained for as long as the project finances need to be retained in case of audit. The organisational information will be made public through our reports.
Project results (deliverables, milestone reports, FAIR Implementation Stories, FAIR action plans, specifications, etc.)	In the case of public deliverables, the final-non approved version is published on Zenodo with relevant disclaimer. It is replaced by the final version once approved by the EC. The milestones are marked as achieved in the European Commission's Funding &	pdf/a	Author information, (name, affiliation)	All deliverables, apart from one, will be made public on Zenodo and FAIR-IMPACT website. D1.1 is marked as Sensitive and will be available for the project consortium, reviewers and European Commission. Some of the Milestone	Concept DOI (one less than the reserved DOI on ZENODO) used to refer to all versions of the record, use: [list semantic artefacts]

Content /Origin	Retention	Format	Personal information	Dissemination level	Comments
	Tenders portal and they are accompanied by short reports for internal use.			reports created for internal use will be considered to be publicly shared on Zenodo and FAIR-IMPACT website.	
Scientific publications	OpenAIRE ORE Other venues		Authors	All scientific publications of the project will be in open access.	All open access scientific publications deriving from the project will be reported in the EC Funding & Tenders portal as per the Grant Agreement.
Recording of webinars, workshops, zoom recordings	YouTube + Zenodo	mp4	Users will give consent for publication of recordings in the registration form	Recordings will be publicly available on YouTube and Zenodo upon participants' and speakers' consent.	Cross-link using PIDs and include slides in Zenodo record
Workshop and online webinar materials other than recordings	Zenodo, website	pdf/a, csv	Users will give consent for publication of photos in the registration form	Presentations, photos and other materials will be publicly available on Zenodo and FAIR-IMPACT website upon participants' and speakers' consent	Make sure PPTx are converted to PDF and that extra materials (e.g. sample codes) are included in Zenodo upload.
FAIR Implementation Framework data	Tabular data that will be captured through desk research and via a public webform to capture details of currently available tools/methods/approaches for implementing the FAIR principles.	csv; html	For data submitted via the public webform, the submitter's name, organisation and email address will be required.	The FAIR Implementation Framework will not contain sensitive personal information. The name, organisation and email addresses	

Content /Origin	Retention	Format	Personal information	Dissemination level	Comments
	Snapshots of the data collected will be uploaded to the FAIR-IMPACT Zenodo community and the data will also be presented through the FAIR-IMPACT website.			submitted using the webform will not be shared publicly and not retained beyond the life of the project.	
New software/tools and Research Software	GitHub with Zenodo archiving ¹ for PIDs. Also submit a repository url for used dependencies and source code to Software Heritage. ²		Authorship of developers. Ensure test data is not sensitive and open license.	Edit Zenodo record and CodeMeta to include property funding attribution. A software page on the FAIR-IMPACT website will be considered to list software records that are useful for FAIR-IMPACT or developed by FAIR-IMPACT.	Follow T4.3 FAIR Research Software recommendations, CodeMeta, ³ SPDX. Permissive ⁴ Open Source licence e.g. MIT, Apache-2.0
Modified/extended software/platforms	Existing repository, encourage Zenodo archiving and attribution of FAIR-IMPACT authors		Authorship of developers. Ensure test data is not sensitive.	Need "contributed to" webpage.	Encourage T4.3 recommendations. Ensure existing open source license is declared.

The file formats are open and comply with good practices and sustainability preferences of the repositories where the data are published. Proper funding attribution will be provided at all times.

¹ <https://docs.github.com/en/repositories/archiving-a-github-repository/referencing-and-citing-content>

² <https://archive.softwareheritage.org/save/>

³ <https://codemeta.github.io/>

⁴ https://en.wikipedia.org/wiki/Permissive_software_license

3 FAIR data and other research outputs

As per its scope, the FAIR-IMPACT project is working towards realising an EOSC of FAIR data and services, by supporting the implementation of FAIR-enabling practices across scientific communities and research outputs at a European, national, and institutional level. This section focuses on how the project makes its data and other research outputs, such as software, Findable, Accessible, Interoperable and Reusable.

All outputs (except administrative items containing legal, financial, or sensitive information) will be deposited under a CC-BY 4.0 licence, using recommended file formats and metadata. A provision on software is included under Table 1. Where exploitable IP is developed, IP management will follow the Consortium Agreement. Information specifically marked as confidential will not be disclosed or used other than for the purpose for which it was collected. Anonymised survey data and software code (with supporting documentation) will be published in certified digital repositories and/or ZENODO. Data, including personal data, collected for the purposes of administering the project activities will be held securely and according to legislation. The project follows the principle of “as open as possible, as closed as necessary”.

File formats used are open and allow straightforward reuse. Where possible, metadata follows the widely accepted Dublin Core and DataCite specifications. The project will make extensive use of PIDs as well as utilise reference metadata, i.e. semantic artefacts when providing discovery metadata. In the ZENODO upload interface all required fields will be completed including ‘Upload type’ (for e.g. ‘Publication/Project deliverable’ for deliverables, ‘Publication/Project Milestone’ for milestones, ‘Dataset’ for underlying data), full list of ‘Authors’ with Contributing Authors listed in order at the discretion of the Lead Author (with ORCIDs if available), ‘Description’, and ‘Version’. The following recommended/optional fields will also be completed: ‘Funding’ (European Commission with value 101057344), ‘Related/alternate identifiers’, ‘Contributors’ and ‘Subjects’. All FAIR-IMPACT outputs will have the keywords ‘FAIR’, ‘FAIR data’, and ‘FAIR-IMPACT’. A comprehensive list of common terms based on efforts initially coming from the FAIRsFAIR project will be created for the purposes of the FAIR-IMPACT project.

FAIR-IMPACT commits to early and open sharing of all its results. All project outputs - documents and data of all types - will be open for comment, deposited in ZENODO, and be published on the project web platform as soon as they are available. To ensure Open Access, a machine-readable electronic copy of the published version or the final peer-reviewed manuscript accepted for publication, is deposited in a trusted repository for scientific publications at the time of publication at the latest. FAIR-IMPACT follows Art 17 of the Grant Agreement with regards to metadata of publications. Anonymised survey data and

software code (with supporting documentation) will be published in certified digital repositories and/or ZENODO. Source code will be shared via GitHub or alike and eventually preserved on Software Heritage. At the end of the project the Executive Committee (EXC) and Project Coordination Office (PCO) will undertake a selection process to determine which project outputs need to be preserved for the long term. The selected assets will be deposited in the CoreTrustSeal certified trustworthy data repository of DANS for long-term preservation.

4 Allocation of resources

Delivering data that is as FAIR as possible is integral to the project, thus no separate step or budget for making data FAIR has been planned. The amount of research data envisaged for the project is modest and does not require a cost-benefit analysis for the long-term storage of each component. The costs for long-term, indefinite in principle, preservation in a trustworthy archive are covered.

Each WP leader is responsible for data management within the WP, including updates and implementation according to the rules and deadlines set out in this DMP. The PCO assumes responsibility for overall data management in the context of the FAIR-IMPACT project and for evaluating the implementation of the DMP. Evaluations and updates of this DMP will take place in M14, M26 and M36, and in alignment with the FAIR-IMPACT management and progress reports.

5 Data security

At the end of the project, the selected research data and code will be published and safely preserved in the CoreTrustSeal certified trustworthy data repository of DANS with transparent data security processes in place. The CoreTrustSeal certification requirements include provisions for the protection of data and users woven into the technical infrastructure of the repository.⁵

During the project the research data are stored in the FAIR-IMPACT Google Shared Drive, with folders per WP. This allows for file sharing across partners and the tracking of revisions. A Google Workspace Enterprise Plus account is used, providing advanced security and

⁵CoreTrustSeal Requirements2023-2025_v01.00 <https://doi.org/10.5281/zenodo.7051011> Still R16

compliance controls for productivity and collaboration with unlimited storage. The data region for storing covered data is Europe.⁶

Access to the FAIR-IMPACT Shared Drive is managed by the PCO. Google's network is protected from external attacks. Data belonging to G Suite customers is stored at rest in two types of systems: disks and backup media. Google also stores data on offline backup media to help ensure recovery from any catastrophic error or natural disaster at one of their data centres. G Suite offers a data loss prevention policy to protect sensitive information within Google Drive. With respect to data protection, Google is committed to complying with the GDPR for G Suite.⁷

6 Ethics

This section of the DMP includes ethical aspects, values and information on data protection in the context of the project that might affect data sharing.

FAIR-IMPACT reaches out to many organisations, individuals, and other projects, and organises interviews, workshops, and other activities, involving people from within and outside the project. In doing so FAIR-IMPACT ensures proper handling of ethical aspects, values, and data protection in accordance with articles 14 and 15 of the Grant Agreement. While each partner is responsible for their own actions, KNAW-DANS as coordinator of the FAIR-IMPACT project directs and supports partners in acting according to the DMP.

FAIR-IMPACT has a Privacy Policy and Terms of Use Statement⁸ for services provided via the project website, which addresses personal data (processing, data subject's rights, opt-out, cookies used on the website and in social media, et cetera). KNAW-DANS is the Data Controller regarding all personal data processing carried out through the website, while all project partners have been identified as Data processors for General Data Protection Regulation (GDPR) purposes and therefore allowed to access personal data provided via the website managed by Trust-IT. For example, Name, contact details and other Personal Data (email, sensitive information regarding gender or dietary requirements needed for the organisation of events) processed for applications collected via the FAIR-IMPACT website will be kept by KNAW-DANS as Data Controller for up to 5 years, to allow for possible external audits, as requested by contractual provisions the Data Controller is subjected to. The

⁶ More information on the account can be found here: <https://support.google.com/a/answer/7284269?hl=en> (general), <https://workspace.google.com/pricing.html> (general), <https://support.google.com/a/answer/7630496?hl=en> (data regions)

⁷ https://cloud.google.com/security/infrastructure/design/resources/google_infrastructure_whitepaper_fa.pdf and <https://cloud.google.com/security/compliance/gdpr/>

⁸ <https://fair-impact.eu/privacy-policy-full>

FAIR-IMPACT website will also be kept live for 5 years after the end of the project and maintained by Trust-IT to support sustainability of the project outputs.

The project only collects personal data that is necessary to fulfil its information needs, respecting the principle of data minimisation, and will not intentionally collect “special categories of data” in terms of the GDPR.⁹ All personal data collected for the project will be processed in accordance with the GDPR.

A large part of the personal data is collected through the FAIR-IMPACT website which offers a privacy statement¹⁰. Through the website a so-called data subject can register for the newsletter, mailing lists and events. These processes are based on informed consent that is gathered through the website. The other main means of gathering information is through interviews. Processing this information is again based on informed consent, where data subjects are properly informed on, among other things, the purposes of the data collection, how the data will be used, and with whom it may be shared. For interviews the project utilises specific templates for informed consent forms and information sheets (in language and terms intelligible to the participants).

Personal data of the project partners for purposes of administering the project activities will not be shared externally unless there is a sufficiently substantiated reason to do so. The FAIR-IMPACT website and its contacts’ database are managed by the FAIR-IMPACT WP leader for Dissemination, Exploitation and Communication, Trust-IT. For the management of the project FAIR-IMPACT uses a Google Drive (see details under section 5: data security). Google is a processor of KNAW-DANS. FAIR-IMPACT uses ZOOM for video communication and online events. ZOOM is a processor of KNAW-DANS.

FAIR-IMPACT has recruited an Ethics advisor, Sabina Leonelli, Professor of Philosophy and History of Science, University of Exeter, UK, who will be a member of its High Level Advisory Committee (HLAC).

7 Other issues

FAIR-IMPACT partners agree with the principles and good practices in the European Code of Conduct for Research Integrity (ALLEA, 2017).¹¹ In addition, project partners may have their own national and/or institutional data management policies for data in their charge. The majority of the project partners are based in the EU, while the consortium also includes 1

⁹ <https://gdpr-info.eu/art-9-gdpr/>

¹⁰ <https://fair-impact.eu/privacy-policy-full>

¹¹ <https://allea.org/code-of-conduct/>

beneficiary (CESSDA) and its affiliated entity (NSD) based in Norway and 4 associated partners based in the UK (UEDIN/DCC, UESSEX/UKDS, UNIMAN and UKRI-STFC).