

D1.2

Data Management Plan

Submission Date 2023.07.31	Version 2.0 – Approved by EC		
Due Date 2023.06.30	PU	Public	x
	SEN	Sensitive	
	R-UE/UE-R	EU classified	
Deliverable Title	Data Management Plan		
Deliverable No.	D1.2		
Lead beneficiary	IBL PAN		
Contributing WP	WP1		
Type	DMP – Data Management Plan		
	HORIZON-INFRA-2022-EOSC-01 Grant Agreement: 101094397		

Project Full Title	Creating a Robust Accessible Federated Technology for Open Access
Project Acronym	CRAFT-OA
Project No.	101094397
Start Date	2023.01.01
End Date	2025.12.31
Duration	36 Months
Project Website	https://craft-oa.eu
Author	Nikodem Wołczuk (https://orcid.org/0000-0002-4303-2016)
Abstract	The CRAFT-OA Data Management Plan provides information about data reused or produced in the project. The document presents the origin of the data, the methods of their processing, use and storage, as well as the actions taken to make them comply with FAIR principles.

Version and Revision History

Version	Date	Author/Reviewer/Contributors	Comments
0.1	2023.07.06	Author: Nikodem Wołczuk (IBL PAN)	Initial draft
0.2	2023.07.13	Author: Nikodem Wołczuk (IBL PAN) Contributor: Tabea Klaus (UGOE)	Version submitted to 1st review
0.3	2023.07.18	Reviewer: Birgit Schmidt (UGOE)	1st review
0.4	2023.07.19	Author: Nikodem Wołczuk (IBL PAN)	Comments from 1st review integrated, Version submitted to 2nd review
0.5	2023.07.25	Reviewer: Dulip Withanage (TIB)	2nd review
0.6	2023.07.26	Author: Nikodem Wołczuk (IBL PAN)	Comments from 2nd review integrated, submitted for final formal check
0.7	2023.07.31	Contributors: Tabea Klaus (UGOE), Theresa Waldmann (UGOE)	Final formal check
1.0	2023.07.31		Submitted version
1.1	2024.07.18	Contributors: Theresa Waldmann (UGOE), Lisa Müller (UGOE)	Correction of formal mistakes, e.g. spelling mistakes
2.0	2024.07.18		Approved by EC

Disclaimer



CRAFT-OA is funded by the European Union under Grant Agreement no. 101094397. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.



This deliverable is licensed under a Creative Commons Attribution 4.0 International License.

List of Acronyms

D	Deliverable
DDH	Diamond Discovery Hub
DMP	Data Management Plan
DOI	Digital Object Identifier
DPO	Data Protection Officer
EC	European Commission
EOSC	European Open Science Cloud
FAIR	Findable, Accessible, Interoperable, Reusable
GDPR	General Data Protection Regulation
GNU	GNU ist nicht gleich Unix
IBL PAN	Instytut Badan Literackich Polskiej Akademii Nauk
IPSPs	Institutional Publishing Service Providers
IPTPs	Institutional Publishing Technology Providers
JSON	JavaScript Object Notation
KPI	Key Performance Indicator
M	Milestone

maDMP	machine actionable Data Management Plan
OA	Open Access
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OJS	Open Journal Systems
OpenAIRE	Open Access Infrastructure for Research in Europe
OPERAS	Open Access in the European Research Area Through Scholarly Communication
ORCID	Open Researcher and Contributor ID
PDF	Portable Document Format
PID	Persistent Identifier
PKP	Public Knowledge Project
PMB	Project Management Board
ROR	Research Organization Registry
T	Task
TIB	Technische Informationsbibliothek
UBern	University of Bern
UGOE	Georg-August-Universität Stiftung öffentlichen Rechts

WP	Work Package
XML	Extensible Markup Language

Table of Content

1 Executive Summary	10
2 Method	11
2.1 Data Management Plan requirements	11
2.2 Machine actionable Data Management Plan	11
2.3 Tool	12
2.4 Workflow.....	12
3 General data description.....	13
4 CRAFT-OA DMP datasets.....	14
4.1 WP1: Personal Data for management and communication	14
4.2 WP2: T2.3 CRAFT-OA Community Building.....	21
4.3 WP3: Documentation and outputs of WP3	28
4.4 WP4: Technical documentation for development tasks T4.1, T4.2 and T4.3.....	33
4.5 WP4: Software components developed in tasks T4.1, T4.2 and T4.3	37
4.6 WP4: Survey data in T4.4	41
4.7 WP4: Deployment and upgrading toolkits T4.4.....	44
4.8 WP4: Workshop materials T4.4.....	47
4.9 WP5: T5.1 Documentation on aggregators technical requirements.....	50
4.10 WP5: T5.2 Visibility Pathfinder	56
4.11 WP5: T5.3 Diamond Discovery Hub.....	60
4.12 WP6: T6.1: EOSC onboarding	63
4.13 WP7: UBern Dataset.....	64
4.14 OpenAIRE Graph Dump	71
5 References	75



5.1 List of References 75

5.2 List of Websites 75

Appendix 1. Argos Horizon Europe template – list of questions 76

1 EXECUTIVE SUMMARY

Deliverable 1.2 “Data Management Plan” (DMP) focuses on describing the data generated and reused during project activities of the CRAFT-OA project, according to good scientific practice and efforts in support of Open Science and transparency inherent in any research project. Because of the variety, complexity and variability of the data that can be associated with the project, this document will be crucial for maintaining the correct handling of resources throughout the lifespan of the project.

The Deliverable (D) is divided into three sections, which together provide a complete overview of data management in the CRAFT-OA project:

- The first section contains information on what methods and tools have been chosen to prepare descriptions of the datasets provided in each work package (WP).
- The second section provides a general overview of the project data, which is the basis of understanding the specifics of the project and an introduction to the detailed part.
- The third section is the main element of the DMP. This is the content exported from the used online tool and contains detailed information for each dataset in a format based on the guidelines of the European Commission (EC).

By its nature, the DMP should be treated as a living document that is subject to constant review, complement and change. This document is one of the versions produced during the project and at an early stage of the project may be characterised by a certain level of incompleteness or generality that will be improved as the project proceeds. All versions of DMP will be made publicly available in the Argos¹ platform and Zenodo² repository.

¹ <https://argos.openaire.eu/splash/about/how-it-works.html>

² <https://zenodo.org/>



2 METHOD

In large scale projects, such as CRAFT-OA, it is important to adopt an adequate strategy when preparing the Data Management Plan (DMP). With a large amount of data, this document can become hard to create, and the need to control variability by introducing subsequent versions can cause further difficulties. This section presents the solutions that have been selected to overcome these difficulties and produce a qualitative DMP in line with the European Commission (EC) guidelines³.

2.1 Data Management Plan requirements

It is important to specify what scope the DMP should cover. Any research data that is generated or reused during the project will be included in the plan. This includes all research data, survey responses, workshop records, project documentation, written project outcomes (such as Deliverables (D)), personal data, software developed, etc. The described output data is not limited to formats, so it can be text, numerical, graphical, video or audio.

The main rule during the creation of the DMP was to achieve compliance with the Findable, Accessible, Interoperable, Reusable (FAIR) data principles⁴ and the guidelines of the EC. The last section of this document contains detailed information on how data management in the project corresponds to the assumptions that data should be Findable, Accessible, Interoperable and Reusable as much as possible. The data descriptions also include information about data security and potential ethical issues (if necessary).

2.2 Machine actionable Data Management Plan

In order to simplify the process of creating a DMP and standardize the information produced in this way, it was decided to look for a tool that would enable it. The answer to this type of needs is machine actionable DMP (maDMP) which is an extension of the standard DMP approach⁵.

The advantage of the tools supporting this type of solution is the adaptation of this previously only human-readable document to the technological requirements of modern data exchange practices. Adapting the data description structure in such a way that it is possible to create files that can be easily processed by computers significantly improves interoperability, sharing

³ Information on the structure and content of the DMP is available, for example, in the Data management plan (HE) file: European Commission. (2021). *Horizon Europe Data Management Plan Template*. Version 1.0. 05.05.2021. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/template/report/data-management-plan_he_en.docx provided in the Horizon Europe (HORIZON) Reference Documents <https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/reference-documents>

⁴ FAIR stands for Findable, Accessible, Interoperable and Reusable: <https://www.go-fair.org/fair-principles/>

⁵ More information: Miksa et al. (2020), Simms et al. (2017).

and reuse of information contained in the document. Relying on maDMPs allows integration of standardised information, e.g. by means of persistent identifiers, and ensures the unification of at least part of the information.

2.3 Tool

The Argos platform was selected for the creation of the CRAFT-OA DMP, an online service developed by Open Access Infrastructure for Research in Europe (OpenAIRE)⁶ and EUDAT⁷ that can be used to create a maDMP and manage its subsequent versions. It allows the simultaneous collaboration of many users while working on one DMP. Working with Argos consists in adding descriptions of datasets using simple questionnaires based on templates adapted to the requirements of individual projects. The tool also allows exporting the finished DMP in Portable Document Format (PDF), DOC, Extensible Markup Language (XML) and RDA JavaScript Object Notation (JSON)⁸ formats. It is possible to quickly deposit a document in the Zenodo repository and get a Digital Object Identifier (DOI). Argos itself is a platform that allows users to search and view DMPs marked as public.

2.4 Workflow

The preparation of the data descriptions was based on the participation of the leaders of the individual work package (WPs). Leaders entered dataset information into Argos using available questionnaires using the Horizon Europe template (based on EC guidelines) as the basis. The knowledge of the tasks (T) envisaged by the WPs they lead was used by them to present the details of the data as reliably as possible. After preparing the datasets, the descriptions were additionally verified. The result is information that clearly represents the management of data on the current project stage within individual WPs.

As far as possible, data of similar type, structure, origin and characterized by similar use and storage were grouped within single datasets to maintain a reasonable level of fragmentation of information.

Not all possible fields of the questionnaires were filled in in all cases, depending on the specificity of the described resources, some fields were omitted if they were not required. As the project activities progress, datasets will be added and made more specific. All changes will be incorporated into future versions of this document and the DMP in Argos.

⁶ <https://www.openaire.eu>

⁷ <https://www.eudat.eu>

⁸ RDA JSON is a format designed according to Research Data Alliance (<https://www.rd-alliance.org/>) guidelines to facilitate the exchange of information by systems working with machine-actionable data management plans. For more information go to <https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/tree/master>

3 GENERAL DATA DESCRIPTION

A whole range of data of different types is expected to be produced during the project. Although each WP has individual specificities of the tasks assigned in the project, some types of material are common in relation to the outputs that WPs generate. The data envisaged at an early stage of the project can be divided into certain groups based on their type, origin or expected handling during the project. The groups specified at this point are:

- Coordination and reporting materials – collaborative produced materials for coordinating activities, such as notes from project meetings, are mainly stored in Google Documents format and saved and stored on institutional servers e.g. as Word file or similar. For all reports and deliverables, the target format is PDF. Documents whose content permits this will be made public under Open Access (OA) rules.
- Survey and training data – a large part of WPs plan to carry out user needs research, which may involve the need to obtain personal information. Any such datasets will be anonymised where necessary and treated in accordance with the General Data Protection Regulation (GDPR) principles. All training data (e.g. presentations, videos) will be prepared in such a way as to maximise their content value and ensure accessibility and reusability.
- Software data – tasks which create or extend the functionalities for the publishing platforms will produce output in the form of software (e.g. Open Journal Systems (OJS) plugins) with a possible high variability in the technologies used. Source code will be treated with due regard to the rules of the open source community and stored in one of the leading repositories (e.g. GitHub).
- Metadata of scholarly records – some of the tasks, such as developing the Diamond Discovery Hub (DDH), involve processing a large amount of metadata records for enrichment purposes. These data will be transferred using open protocols (e.g. the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH)) and presented in standardised formats (e.g. DublinCore).

In all the groups mentioned above, where possible, efforts will be made to fully integrate with the FAIR principles. All personal data processed in the course of the project will be handled in compliance with GDPR. Details of the datasets produced by each WPs are provided in the next section (4 CRAFT-OA DMP datasets).

4 CRAFT-OA DMP DATASETS

This section contains dataset descriptions exported from Argos. Each of the subsections (4.x) corresponds to one dataset. Within one dataset, below the words “Dataset Description”, there are numbered questions and answers corresponding to the individual questions in the Argos questionnaire template.

The datasets below represent Version 0 of the CRAFT-OA DMP in Argos. In the early stages of a project, not all data can be described in a complete way. In some cases, more comprehensive information may only be obtained as the project proceeds and further decisions within the tasks are made. For this reason, there may be discontinuities in the numbering of successive sections of the questionnaire because not all questions have been answered (due to the early stage of the project or the specificity of the data). Some datasets (or questionnaire parts) have been described only sketchily to indicate their presence and to elaborate on their description in subsequent versions of the DMP as the project develops. The information on these datasets may change significantly in the future. The template used to create dataset descriptions in Argos is Horizon Europe template⁹. For all the questions provided by the template read Appendix 1. Argos Horizon Europe template - list of questions.

4.1 WP1: Personal Data for management and communication

Personal data (e.g. mail addresses, names, affiliations, post addresses, sociodemographic) collected for project management or communication purposes such as mailing lists, meeting minutes, event registration, sign-ups for newsletters, outputs of the projects (deliverables, reports etc.) as well as other data created/collected for internal and external management or communication purposes such as Fact Sheets, chat logs, meeting recordings, device IDs, location, IP addresses.

Dataset Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Other

1.1.2 Is it physical or digital?

Digital

⁹ The Horizon Europe template that was used for the CRAFT-OA DMP is a questionnaire template prepared by the Argos platform administrators. It is intended to comply as widely as possible with the EC guidelines for DMPs of projects participating in the Horizon Europe programme.

1.1.3 Are you generating or re-using it?

New

Data used for management purposes were filled in and shared with the Coordinator by the responsible project or work package leader for the project members from their partner institutions in the beginning of the project via an Admin Sheet and will be shared for project members who joined later usually by mail. For communication purposes the individuals signing up for events or newsletters themselves, re-use is not an option in this case.

1.1.5 What is its format?

Google Document Link File

.csv, .doc/.docx, .pdf, .jpeg/.jpg/.jpg2, .png, .wav, .mp3, .mp4, .html, .css,

1.1.6 What is its expected size?

10-12 GB

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To keep on record
- To make informed decisions

Personal data is collected to manage the project and will be stored according to the guidelines from the European Commission only as long as necessary to enable the Coordinator and the consortium members to document the project processes and results e. g. in the context of possible audits. Additionally the data will be used to inform e.g. the project members and external persons about project results, meetings and events. According to GDPR it is possible at any time project for project members and externals to request removal of data from storage at any time. The personal data will be stored on institutional, access-protected servers with a backup routine, to which only a limited number of people have access.

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

Other

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.2 Datasets

2.2.1 Does the described output use or support any published dataset?

No

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset/output?

- Data identifiers
- Researchers identifiers
- Organizations identifiers

DOI

ORCID

Other

In case a recording of an event or any communication output will be published, all outputs will appear in the CRAFT-OA Zenodo Community and thereby have a DOI, the authors of and contributors to the project results like deliverables will identify themselves as far as possible via ORCID.

3.1.1.2 Will you provide metadata for the described dataset/output?

Yes

yes, such as: authors (possibly ORCID), title, DOI, description, date, keywords

3.1.1.3 What type(s) of metadata?

- Descriptive
- Reference

3.1.1.4 Do the metadata use standardised vocabularies?

No

3.1.1.6 Are the metadata searchable?

No

3.1.1.8 Are keywords provided in the metadata?

Yes

3.1.1.9 Are metadata harvestable?

Yes

At least for Zenodo publications.

3.2.1 Repository

3.2.1.1 In which repository will the dataset/output be deposited?

Zenodo

zenodo.org

not all data will be published in this repository but a part of it

3.2.1.5 Does the repository(ies) assign datasets/outputs with persistent identifiers?

Yes

3.2.1.6 Does the repository(ies) resolve the identifiers to a digital object?

DOI

3.2.1.7 Does the repository support versioning?

Yes

3.2.2 Data

3.2.2.2 How is the dataset/output shared?

Closed

Mailing lists etc. for management purposes will only be shared in the coordination and communication task force responsible. Recordings will be made public, with the approval of the participants.

3.2.2.3 What is the reason of limiting access to the dataset/output?

Personal data is under special protection and will only be shared publicly with the consent of the project members or after external participants have given their approval. Some project outputs (deliverables) are intended for internal purposes and shared only with the consortium and the European Commission.

3.2.2.9 Please specify how the dataset/output will be accessed during and after the project ends

The personal data of the project will only be stored (as far as the consent of the person is not withdrawn before) as long as necessary in accordance with the guidelines of the EC and the requirements of the partner institutions.

To the personal data for management purposes only colleagues with access given from the coordination team can access the data, which is stored on institutional servers with backup routines, with their individual institutional login..

3.2.2.10 Please specify how long after the project has ended the dataset/output will be made accessible for

The storage of basic personal data for logins to platforms or services will be necessary by the organizations that enable the sustainable availability of the content and services after the project has ended. Users have the possibility to delete their data at any time. Project results, like deliverables and the data contained will be long-term preserved from the Zenodo repository. The personal data of the project will only be stored (as far as the consent of the person is not withdrawn before) as long as necessary in accordance with the guidelines of the EC and the requirements of the partner institutions to enable possible audits after the project has ended.

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset/output can not be openly shared?

Yes

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No



3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset/output?

The results will be shared as open as possible and as restricted as necessary with licence corresponding to the content, e.g. CC-BY for public documents or community adequate software licences.

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes

For mailing lists, lists of names: no. For recordings such as recordings of webinars, trainings and the deliverables, yes.

3.4.6 What documented procedures for quality assurance do you have in place?

The personal data for management processes is under current control and will be cleaned, (re-)structured and supplemented if necessary from the coordination team.

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

0 Euro

Storage

Indirect cost

no additional costs will be involved as Zenodo is used, for management purposes the available servers and services from the institution will be used.

4.1.2 How will this cost be covered?

Collaboration with other Projects

see above

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

The DPO for CRAFT-OA is Prof. Dr. Andreas Wiebe, DPO from the University of Göttingen.

5.1 Data Security

5.1.1 What security measures are followed?

Other

Personal data will only be used and available project-internal. Sociodemographic data concerning the website will be collected via a Matomo plug-in, meaning they are GDPR-conformant. In case events are recorded, participants receive a legal disclaimer informing them about their rights and are able to choose to attend the event without being recorded. Twitter Analytics, a third-party service, is used to collect data for KPIs from Twitter. Personal data will not be shared via Twitter except with consent of the person. All mailing lists are hosted via the partner coordinating the project, the University of Göttingen and the website is hosted by the University of Torino. The personal data for management purposes will be stored on institutional servers with restricted access and with backup routines, to avoid loss of data.

5.1.3 How will you preserve the described dataset/output in the long term?

Personal data for management purposes will only be stored as long as necessary according to the guidelines of the European Commission and the institutional regulations, e.g. of the University of Göttingen as coordinator, to enable a transparent documentation of the project within possible audits during and after the project.

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset/output?

yes

Personal data is collected.

Additional note about the craft-oa.eu website: I asked our partners from the University of Torino, the servers the website is hosted on, are their own. Therefore, we don't need an additional declaration from a service provider running the servers. (just as info in case reviewers mention this, does not necessarily need to go into the DMP).

6.1.2 Does the described dataset/output contain sensitive information?

Unknown

6.1.3 Does the described dataset/output contain personal data?

Yes

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

4.2 WP2: T2.3 CRAFT-OA Community Building

CRAFT-OA will build up communities of Institutional Publishing Service Providers (IPSPs) and Institutional Publishing Technology Providers (IPTPs). This will include collection of contact data and affiliation data of people, groups and initiatives. Collected data needs to be protected enough to comply with GDPR rules and yet FAIR enough to support community building.

Dataset Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Other

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

This specific data collection will serve a new usecase (community building around Diamond Open Access IPSPs and IPTPs), however it will include re-used data as consortium partners will contribute data from their local, regional, national or disciplinary networks.

1.1.4 What is the type of the described dataset?

Derived or compiled

1.1.6 What is its expected size?

1200 KB

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To keep on record
- To combine with other data

1.1.8 What is its origin/provenance?

personal networks, public networks, public information

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

- Research communities
- Other

Building up communities of practice requires to know these communities, network them and enable them to enter into monitoring on their terms.

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.2 Datasets

2.2.1 Does the described output use or support any published dataset?

No

2.3 Software

2.3.1 Does the described output use or support any software?

Yes

<https://pkp.sfu.ca/software/ojs/download/> and <https://github.com/OpenEdition/lodel> and <https://github.com/birkbeckctp/>

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset/output?

- Data identifiers
- Researchers identifiers
- Organizations identifiers



- National identifiers

DOI

ORCID

ROR

Finalised data sets might also be disseminated with handles or other data IDs. Is to be decided.

3.1.1.2 Will you provide metadata for the described dataset/output?

Yes

We will use generic metadata schemes such as Dublin Core to describe the datasets that we can make public.

3.1.1.3 What type(s) of metadata?

- Descriptive
- Administrative
- Structural
- Legal

3.1.1.4 Do the metadata use standardised vocabularies?

No

3.1.1.6 Are the metadata searchable?

Yes

3.1.1.7 How are searchable metadata provided?

Registry/Catalogue

Maybe there are controlled vocabularies for community-building data. We don't know of them yet, but maybe in one of the planned updates of the DMP.

3.1.1.8 Are keywords provided in the metadata?

Yes

3.1.1.9 Are metadata harvestable?

Yes

If we are able to place the data in a useful repository for the Diamond OA capacity centre they will be harvestable, f.e. via OAI-PMH. If possible the data should also be included in other aggregations such as the EOSC or OpenAIRE research graph. In that case we might need different data models and ontologies.

3.2.1 Repository

3.2.1.1 In which repository will the dataset/output be deposited?

not decided yet, either an OPERAS or OpenAIRE service

3.2.1.2 Is the selected repository a trusted source?

Yes

- Has certification
- Follows repository standards
- Has an open access content policy
- Follows metadata standards
- Uses non-proprietary formats
- Follows curation processes
- Supports authentication and authorization of users
- Has data security mechanisms in place

3.2.2 Data

3.2.2.1 What is the described dataset/output title?

something like "data on IPSPs and IPTPs communities"

3.2.2.2 How is the dataset/output shared?

Shared

3.2.2.3 What is the reason of limiting access to the dataset/output?

The data might contain personal data that shouldn't be made public. These decisions are part of the project and will become clearer during the two planned DMP updates.

3.2.2.5 Are there any methods or tools required to access the dataset/output?

No

3.2.2.8 Is the described dataset/output supported by a data access committee?

Yes

These decisions will be made by the project's PMB and General assembly.

3.2.2.9 Please specify how the dataset/output will be accessed during and after the project ends

Based on the above mentioned decisions the collected data should be made as FAIR as possible and as GDPR-compliant as needed.

3.2.2.10 Please specify how long after the project has ended the dataset/output will be made accessible for

Hopefully forever? As long as there is Diamond Open Access around the data collection should live on.

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset/output can not be openly shared?

Yes

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

3.2.3.3 Do metadata provide information about how to access the described dataset/output?

Yes

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

Couldn't find it? Insert it manually

If such vocabularies or standardised data models for community building exist, we will hopefully find them and use them.

3.4 Increasing data and other outputs reuse

3.4.2 What reusability and/or reproducibility methods are followed?

Readme files

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes

That's one of the goals of the project to build up communities of practice that will last beyond the funded period.

3.4.5 Is provenance well documented?

Yes

Provenance of community information is part of community building.

3.4.6 What documented procedures for quality assurance do you have in place?

Data conform to format specification

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

several person months

- Storage
- Re-use
- Security

Indirect cost

4.1.2 How will this cost be covered?

- Use of institution infrastructure
- Collaboration with other Projects
- Other

in-kind contributions by communities of IPSPs and IPTPs and consortium partners in their capacity of being part of these communities (is in their self-interest to keep the collected data up-to-date and reusable)

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Couldn't find it? Insert it manually

Margo Bargheer 0000-0001-8246-8210

5.1 Data Security

5.1.1 What security measures are followed?

Other

security measures will be decided when the balance of FAIR and GDPR compliance is clearer.

5.1.3 How will you preserve the described dataset/output in the long term?

Community building and community composition data is planned to merge into the Diamond Open Access capacity centre.

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset/output?

unknown

The deliverables D1.1 and D8.1 of CRAFT-OA specify the use of personal data.

6.1.2 Does the described dataset/output contain sensitive information?

Unknown

6.1.3 Does the described dataset/output contain personal data?

Yes

7.1 Other

7.1.1 Do you make use of other procedures for data management?

Yes



4.3 WP3: Documentation and outputs of WP3

WP3 focuses on enabling the professional competence of publishing platforms, publishers and stand-alone OA Diamond journals, by 1) using generally best publishing standards and practices, 2) interfacing using technical standards to better integrate journals into search engines, indexes, library catalogues, and discovery services, 3) identifying challenges and obstacles for OA journals and platforms to be able to comply with such standards and 4) identifying, developing, and implementing relevant help measures (i.e. training materials, workshops, etc.) for Diamond OA journals and platforms to comply with agreed standards.

This Dataset includes rolling minutes, the main outputs (D3.1, D3.2, M23) of the work package and the working documents related to the mentioned tasks and deliverables.

Dataset Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

1.1.4 What is the type of the described dataset?

Derived or compiled

1.1.5 What is its format?

.doc/.docx

.pdf for final version of the deliverable

1.1.6 What is its expected size?

<1 GB

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

- Researchers
- Research communities

- Decision makers
- The public
- Industry

Publishing platforms, publishers, and stand-alone Diamond OA journals

Supporting processes of professionalisation

Increasing interoperability, discoverability and technical sophistication

Supporting cultural change while doing so

Knowledge transfer to researchers as authors

Funders and other policymakers

Establishing funding compliance criteria

Global aggregators

EOSC, EU catalogues, aggregators:

Facilitated harvesting

Optimised metadata

Improving functionality according to FAIR

Improving showcase function of EU research area (ERA) results by high quality

2.1 Publications

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset/output?

Data identifiers

DOI

3.1.1.2 Will you provide metadata for the described dataset/output?

No

3.2.1 Repository

3.2.1.1 In which repository will the dataset/output be deposited?

Zenodo

Zenodo

<https://zenodo.org>

Repository hosted by Zenodo

3.2.1.2 Is the selected repository a trusted source?

Yes

- Follows repository standards
- Assigns PIDs
- Supports mid- and long-term preservation

3.2.1.5 Does the repository(ies) assign datasets/outputs with persistent identifiers?

Yes

3.2.1.6 Does the repository(ies) resolve the identifiers to a digital object?

Yes, DOI.

3.2.1.7 Does the repository support versioning?

Yes

3.2.2 Data

3.2.2.1 What is the described dataset/output title?

CRAFT-OA - WP3 Dataset

3.2.2.2 How is the dataset/output shared?



Open

3.2.2.5 Are there any methods or tools required to access the dataset/output?

No

3.2.2.8 Is the described dataset/output supported by a data access committee?

No

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset/output can not be openly shared?

Yes

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

3.2.3.3 Do metadata provide information about how to access the described dataset/output?

Yes

3.2.3.4 Will metadata remain available after the dataset/output is no longer available?

Yes

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

No

3.3.7 Does the described dataset/output provide qualified references with other outputs?

No

3.4 Increasing data and other outputs reuse



3.4.3 Will you provide the described dataset/output in the public domain?

Yes

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes

3.4.5 Is provenance well documented?

No

3.4.6 What documented procedures for quality assurance do you have in place?

Set up of scientific and technical committee

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset/output?

no

6.1.2 Does the described dataset/output contain sensitive information?

No

6.1.3 Does the described dataset/output contain personal data?

No

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

4.4 WP4: Technical documentation for development tasks T4.1, T4.2 and T4.3

Documentation for requirement engineering and development work in tasks T4.1, T4.2 and T4.3.

Dataset Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Other

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

1.1.4 What is the type of the described dataset?

Derived or compiled

1.1.6 What is its expected size?

20 MB

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To make informed decisions
- To develop a product

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

- Industry

- Other

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.2 Datasets

2.2.1 Does the described output use or support any published dataset?

No

2.3 Software

2.3.1 Does the described output use or support any software?

Yes

<https://github.com/pkp>, <https://github.com/OpenEdition/lodel>

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset/output?

Data identifiers

DOI

3.1.1.2 Will you provide metadata for the described dataset/output?

No

3.2.2 Data

3.2.2.2 How is the dataset/output shared?

Open

3.2.2.5 Are there any methods or tools required to access the dataset/output?

No

3.2.2.8 Is the described dataset/output supported by a data access committee?

No

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset/output can not be openly shared?

Yes

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

No

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset/output?

Creative Commons Attribution 4.0

3.4.3 Will you provide the described dataset/output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

0 Euro

- Storage

- Archiving

Direct cost

4.1.2 How will this cost be covered?

Collaboration with other Projects

5.1 Data Security

5.1.3 How will you preserve the described dataset/output in the long term?

Where possible, released in the PKP Documentation Hub and maintained by the PKP Community.

Other parts released in a repository decided by the CRAFT-OA project.

6.1 Ethical aspects

6.1.2 Does the described dataset/output contain sensitive information?

No

6.1.3 Does the described dataset/output contain personal data?

No

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

4.5 WP4: Software components developed in tasks T4.1, T4.2 and T4.3

Contributions to the Public Knowledge Project applications (Open Journal Systems, Open Monograph Press, Open Preprint Systems, PKP shared library, PKP plugins) and developed independent plugins for PKP applications and Lodel.

The contributions to the PKP applications will be merged to the PKP repositories in Github and will not be maintained separately. Plugins developed in the project will be released in Github.

Dataset Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Software

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

1.1.4 What is the type of the described dataset?

Other

1.1.6 What is its expected size?

20 MB

1.1.7 Why are you collecting/generating or re-using it?

- To develop a product
- To improve a product

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

- Industry

- Other

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.2 Datasets

2.2.1 Does the described output use or support any published dataset?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset/output?

Other

URL

3.1.1.2 Will you provide metadata for the described dataset/output?

No

3.2.1 Repository

3.2.1.1 In which repository will the dataset/output be deposited?

GitHub

3.2.1.2 Is the selected repository a trusted source?

Yes

3.2.1.5 Does the repository(ies) assign datasets/outputs with persistent identifiers?

No

3.2.1.7 Does the repository support versioning?

Yes

3.2.2 Data

3.2.2.2 How is the dataset/output shared?

Open

3.2.2.5 Are there any methods or tools required to access the dataset/output?

No

3.2.2.9 Please specify how the dataset/output will be accessed during and after the project ends

Publicly available in Github during and after the project.

3.2.3 Metadata

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

3.2.3.3 Do metadata provide information about how to access the described dataset/output?

No

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

3.3.3 Have you applied a standard schema for your (meta)data?

No

3.3.4 Will you provide a mapping to more commonly used ontologies?

No

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset/output?

GNU General Public License 3.0

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

0 Euro

- Storage
- Archiving

Direct cost

4.1.2 How will this cost be covered?

Collaboration with other Projects

5.1 Data Security

5.1.3 How will you preserve the described dataset/output in the long term?

Contributions to the PKP applications will be maintained/preserved by the PKP community.

How to maintain the developed plugins after the project runtime will be the responsibility in CRAFT-OA WP7

6.1 Ethical aspects

6.1.2 Does the described dataset/output contain sensitive information?

No

6.1.3 Does the described dataset/output contain personal data?

No

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

4.6 WP4: Survey data in T4.4

Survey data gathered in task T4.4. These may include surveys targeting institutional IT departments and surveys targeting editorial staff of scholarly journals.

Dataset Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

1.1.4 What is the type of the described dataset?

Observational

1.1.6 What is its expected size?

5 MB

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To make informed decisions

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

- Industry
- Other

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

Yes

<https://github.com/pkp/>

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset/output?

Data identifiers

DOI

3.1.1.2 Will you provide metadata for the described dataset/output?

Yes

3.1.1.3 What type(s) of metadata?

Descriptive

3.1.1.4 Do the metadata use standardised vocabularies?

No

3.1.1.6 Are the metadata searchable?

Yes

3.1.1.7 How are searchable metadata provided?

Registry/Catalogue

3.2.2 Data

3.2.2.2 How is the dataset/output shared?

Open

The survey will seek consent for data sharing.

3.2.3 Metadata

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

0 Euro

- Storage
- Archiving

Direct cost

4.1.2 How will this cost be covered?

Other

6.1 Ethical aspects

6.1.2 Does the described dataset/output contain sensitive information?

No

6.1.3 Does the described dataset/output contain personal data?

Yes

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No



4.7 WP4: Deployment and upgrading toolkits T4.4

Deployment and upgrading toolkits and recommendations for OJS and Lodel, based on a requirement survey.

Dataset Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Workflows

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

1.1.4 What is the type of the described dataset?

Derived or compiled

1.1.6 What is its expected size?

10 MB

1.1.7 Why are you collecting/generating or re-using it?

To share information

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

- Research communities
- Industry
- Other

2.1 Publications

2.1.1 Does the described output support any scientific publication?

No

2.1.2 Is there a data availability statement provided along with the publication?

No

2.3 Software

2.3.1 Does the described output use or support any software?

Yes

<https://github.com/pkp>, <https://github.com/OpenEdition/lodel>

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset/output?

None

3.1.1.2 Will you provide metadata for the described dataset/output?

Yes

3.1.1.4 Do the metadata use standardised vocabularies?

No

3.1.1.6 Are the metadata searchable?

Yes

3.2.2 Data

3.2.2.2 How is the dataset/output shared?

Open

3.2.2.5 Are there any methods or tools required to access the dataset/output?

No

3.2.3 Metadata

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

3.3.7 Does the described dataset/output provide qualified references with other outputs?

No

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

0 Euro

- Storage
- Archiving

4.1.2 How will this cost be covered?

Other

6.1 Ethical aspects

6.1.2 Does the described dataset/output contain sensitive information?

No

6.1.3 Does the described dataset/output contain personal data?

No

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

4.8 WP4: Workshop materials T4.4

Task T4.4 will include online workshops in the form of “Meet the Experts” aimed for journal managers and technical staff. The materials of these workshops will be published as a dataset.

Dataset Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Other

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

1.1.4 What is the type of the described dataset?

Derived or compiled

1.1.6 What is its expected size?

20 MB

1.1.7 Why are you collecting/generating or re-using it?

To share information

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

- Research communities
- Industry
- Other

2.3 Software

2.3.1 Does the described output use or support any software?

Yes

<https://github.com/pkp>

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset/output?

None

3.1.1.2 Will you provide metadata for the described dataset/output?

Yes

3.1.1.3 What type(s) of metadata?

Descriptive

3.1.1.4 Do the metadata use standardised vocabularies?

No

3.1.1.6 Are the metadata searchable?

Yes

3.1.1.7 How are searchable metadata provided?

Registry/Catalogue

3.2.2 Data

3.2.2.2 How is the dataset/output shared?

Open

3.2.2.5 Are there any methods or tools required to access the dataset/output?

No

3.2.3 Metadata

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

0 Euro

- Storage
- Archiving

Direct cost

4.1.2 How will this cost be covered?

Other

6.1 Ethical aspects

6.1.2 Does the described dataset/output contain sensitive information?

No

6.1.3 Does the described dataset/output contain personal data?

No

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No



4.9 WP5: T5.1 Documentation on aggregators technical requirements

The dataset contains texts and tables documenting the technical requirements and recommendations of citation indexes and aggregators.

Dataset Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Other

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

The dataset will consist of collection of freely available public information about technical requirements for publications referencing.

1.1.4 What is the type of the described dataset?

Other

1.1.5 What is its format?

The data collected will consist essentially of textual information.

1.1.6 What is its expected size?

Around 1 GB

1.1.7 Why are you collecting/generating or re-using it?

- To share information
- To make informed decisions

- To develop a product

The dataset will be used for:

- inform the target group (diamond publishers) about the aggregators technical requirements
- guide the WP members in building the service that will help referencing the publications on various service providers platforms

1.1.8 What is its origin/provenance?

Final list and hyperlinks will be provided in future.

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

- Research communities
- Other

The dataset targets mainly two groups within the research community: the diamond publishers, the partners involved in the building of support services for the diamond publishers.

2.2 Datasets

2.2.1 Does the described output use or support any published dataset?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset/output?

Data identifiers

DOI

The final deliverable will be made publicly available on a trusted repository providing DOIs.

3.1.1.2 Will you provide metadata for the described dataset/output?

Yes

The final deliverable will use Dublin core metadata.

3.1.1.3 What type(s) of metadata?

- Descriptive

- Reference

3.1.1.4 Do the metadata use standardised vocabularies?

No

3.1.1.6 Are the metadata searchable?

Yes

3.1.1.7 How are searchable metadata provided?

Registry/Catalogue

3.1.1.8 Are keywords provided in the metadata?

Yes

3.1.1.9 Are metadata harvestable?

Yes

3.2.1 Repository

3.2.1.1 In which repository will the dataset/output be deposited?

<https://zenodo.org/>

3.2.1.2 Is the selected repository a trusted source?

Yes

Follows repository standards

3.2.1.5 Does the repository(ies) assign datasets/outputs with persistent identifiers?

Yes

3.2.1.6 Does the repository(ies) resolve the identifiers to a digital object?

yes: to the file (textual, spreadsheets, presentation, etc.)

3.2.2 Data

3.2.2.1 What is the described dataset/output title?

Requirements for diamond journals' global visibility

3.2.2.2 How is the dataset/output shared?

Open

3.2.2.5 Are there any methods or tools required to access the dataset/output?

No

3.2.2.8 Is the described dataset/output supported by a data access committee?

No

3.2.2.9 Please specify how the dataset/output will be accessed during and after the project ends

Openly accessible via zenodo and partners' websites.

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset/output can not be openly shared?

Yes

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

3.3.3 Have you applied a standard schema for your (meta)data?

Yes

Couldn't find it? Insert it manually

Dublin Core

3.3.7 Does the described dataset/output provide qualified references with other outputs?

No

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset/output?

Creative Commons Attribution 4.0

3.4.3 Will you provide the described dataset/output in the public domain?

No

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes

3.4.5 Is provenance well documented?

Yes

Information sources for the establishment of the dataset will be explicitly mentioned and referenced.

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

0 Euro

Storage

Costs of FAIRification are included in the general grant for the project management.

4.1.2 How will this cost be covered?

Infrastructure Grant

5.1 Data Security

5.1.1 What security measures are followed?

Passwords

Public version of the deliverable will be openly accessible, working documents will be stored separately and accessible only to the partners of the project upon identification.

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset/output?

no

6.1.2 Does the described dataset/output contain sensitive information?

No

6.1.3 Does the described dataset/output contain personal data?

No

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

4.10 WP5: T5.2 Visibility Pathfinder

Task 5.2 focuses on the journal visibility, accessibility, and functionality by identifying and selecting a checklist of properties that journals should satisfy in order to ensure maximum visibility through indexes and aggregators in the scholarly communication domain.

This dataset includes results of an analysis of status quo, documentation for technical requirements, and design of the workflows of the Visibility Pathfinder, a tool that will allow journals to self-assess their current level of visibility and provide feedback on how to improve it via the CRAFT-OA DDH, including the technical description of its integration within existing publishing systems (e.g. OJS) and a set of visuals to indicate the level of compliance to the journal users.

Dataset Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Workflows

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

1.1.4 What is the type of the described dataset?

Derived or compiled

1.1.6 What is its expected size?

unknown

1.1.7 Why are you collecting/generating or re-using it?

To develop a product

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

- Researchers
- Research communities
- Decision makers

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset/output?

Projects identifiers

3.1.1.2 Will you provide metadata for the described dataset/output?

No

3.2.2 Data

3.2.2.1 What is the described dataset/output title?

T5.2 Visibility pathfinder data

3.2.2.2 How is the dataset/output shared?

Shared

3.2.3 Metadata

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

3.4 Increasing data and other outputs reuse

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

unknown

- Storage
- Archiving
- Re-use
- Security

4.1.2 How will this cost be covered?

Infrastructure Grant

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

a. Radek Gomola (orcid:0000-0002-6903-9937)

b. Martina Dvořáková (orcid:0009-0004-6521-2773)

Couldn't find it? Insert it manually

c. Michal Růžička (orcid:0000-0001-5547-8720)

Couldn't find it? Insert it manually

5.1 Data Security

5.1.1 What security measures are followed?

- Firewall
- Passwords

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset/output?

no

6.1.2 Does the described dataset/output contain sensitive information?

No

6.1.3 Does the described dataset/output contain personal data?

No

7.1 Other



7.1.1 Do you make use of other procedures for data management?

No

4.11 WP5: T5.3 Diamond Discovery Hub

(Provisory description) The Diamond Discovery Hub (DDH) will increase the visibility of diamond journals on a series of major platforms and aggregators. To that end, it will collect, harmonize, enrich, and disseminate metadata of diamond journals. The DDH dataset will therefore comprise metadata of journals and articles, some components of the OpenAIRE Graph, and potentially other sets of data.

Dataset Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Other

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

Data in the DDH will be re-used and redesigned for new purposes.

1.1.4 What is the type of the described dataset?

Derived or compiled

1.1.6 What is its expected size?

unknown to date

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information

1.1.8 What is its origin/provenance?



Existing publishing platforms and services generating standard metadata for their publications.

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

- Researchers
- Research communities
- The public
- Industry
- Other

2.2 Datasets

2.2.1 Does the described output use or support any published dataset?

No

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset/output?

- Data identifiers
- Researchers identifiers
- Organizations identifiers

PIDs listed above will be part of the service but can not be defined at this stage.

3.1.1.2 Will you provide metadata for the described dataset/output?

Yes

3.1.1.3 What type(s) of metadata?

- Descriptive
- Structural
- Reference

3.2.3 Metadata

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)



4.12 WP6: T6.1: EOSC onboarding

The OJS connector for OpenAIRE Graph only process metadata on the articles intended for publication in the OpenAIRE Graph, i.e. authors of the article, their institutional contacts etc. This information is for clean reasons expected to be published by the subjects. We will make every reasonable effort to ensure that the data is accurate and not lost. Should the author contact the editors with a request to correct the data, the corrections will be distributed via the plugin in the standard way.

For the OJS plugin for the EOSC Research Product Publishing Framework, we will follow the same principle, but we are unable to be more specific for them as of now as the investigation of EOSC Research Product Publishing Framework requirements for the implementation in the project will be done later during the project.

Dataset Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Other

1.1.2 Is it physical or digital?

Digital

3.2.3 Metadata

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

4.13 WP7: UBern Dataset

UBern is an associated partner in the Horizon Europe project CRAFT OA. Our contribution to the project consists to a large extent of semi-structured interviews with platform service providers and journal editors.

Dataset Description

1.1 Brief description of the described research output

1.1.1 What kind of research output are you describing?

Research Data

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

New

1.1.4 What is the type of the described dataset?

Other

1.1.5 What is its format?

- Questionnaire templates (PDF)
- Audio files interviews (MP4)
- Interview transcriptions (PDF)
- Reports (PDF)
- Consent forms (Word, PDF)

1.1.6 What is its expected size?

The expected data volume will not exceed 1 GB.

1.1.7 Why are you collecting/generating or re-using it?



- To obtain information
- To share information
- To make informed decisions

The data are essentially interviews that we conduct with 11 platform operators and journal editors.

The qualitative study, which we are conducting with semi-structured interviews, will provide information on what platform service providers and journal editors think needs to be improved to make Diamond-OA publishing more sustainable and easier. This information will be incorporated into the CRAFT-OA project.

The questionnaires and transcriptions of the interviews in anonymized form will be made publicly available for reuse. They can be used for work in the field of information science and scholarly communications or in start-ups or reorganizations of platforms and journals.

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

- Researchers
- Research communities
- Other

The questionnaires and transcriptions of the interviews in anonymized form will be made publicly available for reuse. They can be used for work in the field of information science and scholarly communications or in start-ups or reorganizations of platforms and journals.

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset/output?

- Data identifiers
- Researchers identifiers
- Organizations identifiers

DOI

ORCID

ROR

3.1.1.2 Will you provide metadata for the described dataset/output?



Yes

Yes, all data will be published on Zenodo. Zenodo is a EC-co-funded, repository, for publications and data. A DOI is automatically assigned to all items and Zenodo allows for versioning. Both publications and research data can be published and Zenodo provides means to link them. Data is stored in the CERN cloud infrastructure. Zenodo is compliant with the open data requirements of Horizon Europe and OpenAIRE. Making data accessible.

3.1.1.3 What type(s) of metadata?

Descriptive

3.1.1.4 Do the metadata use standardised vocabularies?

Yes

Couldn't find it? Insert it manually

3.1.1.5 Please provide URL/Description of used vocabularies

<https://schema.datacite.org/>

The DataCite Metadata Schema is used as the standard metadata schema for all published data in this work package.

3.1.1.6 Are the metadata searchable?

Yes

3.1.1.7 How are searchable metadata provided?

Linked Open Data

3.1.1.8 Are keywords provided in the metadata?

Yes

All results published on Zenodo will provided with search keywords together with their metadata. Keywords for open data will be selected wherever possible from controlled vocabularies that are suitable for the specific type of the data. Where no suitable vocabularies are available free form Keywords will be used.

3.1.1.9 Are metadata harvestable?

Yes

Yes, metadata (CC0) is freely accessible via open interfaces.

3.2.1 Repository



3.2.1.1 In which repository will the dataset/output be deposited?

Zenodo

<https://zenodo.org/>

Zenodo is a EC-co-funded, repository, for publications and data. A DOI is automatically assigned to all items and Zenodo allows for versioning. Both publications and research data can be published and Zenodo provides means to link them. Data is stored in the CERN cloud infrastructure. Zenodo is compliant with the open data requirements of Horizon Europe and OpenAIRE.

3.2.1.2 Is the selected repository a trusted source?

Yes

3.2.1.5 Does the repository(ies) assign datasets/outputs with persistent identifiers?

Yes

3.2.1.6 Does the repository(ies) resolve the identifiers to a digital object?

A DOI is automatically assigned to all items and Zenodo allows for versioning. Both publications and research data can be published and Zenodo provides means to link them. Data is stored in the CERN cloud infrastructure. Zenodo is compliant with the open data requirements of Horizon Europe and OpenAIRE. Making data accessible.

3.2.2 Data

3.2.2.2 How is the dataset/output shared?

Open

The questionnaire templates, the anonymized transcriptions, and the report will be made publicly available.

The consent forms and the audio files will not be made publicly available for privacy reasons. The audio files and the not anonymized transcriptions will be stored in the campus storage of the University of Bern and deleted after the end of the project.

3.2.2.9 Please specify how the dataset/output will be accessed during and after the project ends

The questionnaire templates, the anonymized transcriptions, and the report will be made publicly available.

The consent forms and the audio files will not be made publicly available for privacy reasons. The audio files and the not anonymized transcriptions will be stored in the campus storage of the University of Bern and deleted after the end of the project.

3.2.2.10 Please specify how long after the project has ended the dataset/output will be made accessible for

According to Zenodo's general policies (<http://about.zenodo.org/policies/>), "items will be retained for the lifetime of the repository. This is currently the lifetime of the host laboratory CERN, which currently has an experimental programme defined for the next 20 years at least."

3.2.3 Metadata

3.2.3.1 Will you provide metadata even if the described dataset/output can not be openly shared?

No

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

3.2.3.4 Will metadata remain available after the dataset/output is no longer available?

Yes

3.3 Making data and other outputs interoperable

3.3.1 Does your (meta)data use a controlled vocabulary?

No

3.4 Increasing data and other outputs reuse

3.4.1 What internationally recognised licence will you use for your dataset/output?

Creative Commons Attribution 4.0

3.4.2 What reusability and/or reproducibility methods are followed?

Readme files

3.4.3 Will you provide the described dataset/output in the public domain?

Yes

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

Yes

A ReadMe file with the information relevant to the project is published together with the data. For this purpose, the following template is adapted:

https://www.ub.unibe.ch/unibe/portal/unibiblio/content/e6304/e583799/e573822/e1085861/e1085890/e1085891/pane1085902/e1198324/Readme_Template_EN_v2_20220511_eng.txt

The data is published under a Creative Commons Attribution 4.0 (CC-BY 4.0) licence.

4.1 Allocation of resources

4.1.1 What will be the cost of making the described output FAIR?

50 Swiss Franc

Storage

4.1.2 How will this cost be covered?

Other

4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

Elio Pellin (orcid:0000-0002-4076-6743)

5.1 Data Security

5.1.1 What security measures are followed?

Passwords

5.1.2 What conditions do the security measures meet?

The data is stored on servers managed by the IT of the University of Bern. Only the project staff involved have access to the data. Access is regulated by password protection. The data on the university servers are automatically backed up daily.

Questionnaire templates for the interviews will be stored on Microsoft OneDrive, provided by the University of Bern. The servers are located and managed in Switzerland. Due to the non-sensitive nature of the templates, this should not pose a security risk.

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset/output?

no

6.1.2 Does the described dataset/output contain sensitive information?

Yes

6.1.4 What are the methods used for processing and accessing sensitive/personal information?

Anonymising data where necessary

A consent form is obtained from all interview partners. Data will only be published in anonymised form and identifying data (audio files) will be deleted.

We let a person from project team check transcription and anonymization.

7.1 Other

7.1.1 Do you make use of other procedures for data management?

No

4.14 OpenAIRE Graph Dump

The OpenAIRE Graph is one of the largest open scholarly record collections worldwide, key in fostering Open Science and establishing its practices in the daily research activities. Conceived as a public and transparent good, populated out of data sources trusted by scientists, the Graph aims at bringing discovery, monitoring, and assessment of science back in the hands of the scientific community.

Imagine a vast collection of research products all linked together, contextualised and openly available. For the past years OpenAIRE has been working to gather this valuable record. It is a massive collection of metadata and links between scholarly research products such as articles, datasets, software, and other research products, entities like organisations, funders, funding streams, projects, communities, and data sources.

As of today, the OpenAIRE Graph aggregates around 450Mi metadata records with links collected from 131K data sources trusted by scientists, including:

- Open Access journals registered in DOAJ
- Crossref
- Unpaywall
- ORCID
- Microsoft Academic Graph
- Datacite

And repositories registered in OpenDOAR, re3data.org, FAIRSharing.org, and the EOSC Service Catalogue. Among these, prominent repositories such as:

- UKPubMed
- ArXiv
- HAL
- Zenodo
- Figshare
- Dryad
- Repec

After cleaning, deduplication, enrichment and full-text mining processes, the graph is analysed to produce statistics for the OpenAIRE MONITOR, the Open Science Observatory, made discoverable via the OpenAIRE EXPLORE and programmatically accessible via OpenAIRE Public APIs. Last but not least, the Graph data are openly available and can be used by third-parties to create added value services.

In the context, of CRAFT-OA, this dataset will be used in the following Work Packages:

- WP5: Journal Visibility and Discoverability Improvement
- WP6: Journals' Data Reuse and Uptake in EOSC: T6.1: EOSC onboarding, T6.2: OpenAIRE Publisher Dashboard Service

Learn more about the OpenAIRE Graph at <https://graph.openaire.eu>.

Discover the content of the graph on OpenAIRE EXPLORE and OpenAIRE's API for developers.

Dataset Description

1.1 Brief description of the described research output

1.1.2 Is it physical or digital?

Digital

1.1.3 Are you generating or re-using it?

Re-used

1.1.7 Why are you collecting/generating or re-using it?

- To obtain information
- To share information
- To keep on record
- To make informed decisions
- To develop a product
- To improve a product
- To combine with other data

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

- Researchers
- Research communities
- Decision makers
- Education
- Other

3.1.1 Making data findable, including provisions for metadata

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset/output?

- Data identifiers
- Researchers identifiers
- Organizations identifiers
- Projects identifiers
- National identifiers
- Other

3.1.1.2 Will you provide metadata for the described dataset/output?

Yes

3.1.1.3 What type(s) of metadata?

- Descriptive
- Administrative
- Structural

3.1.1.4 Do the metadata use standardised vocabularies?

Yes

3.1.1.6 Are the metadata searchable?

Yes

3.2.3 Metadata

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

4.1 Allocation of resources

4.1.2 How will this cost be covered?

Other

5.1 Data Security

5.1.2 What conditions do the security measures meet?

- Data access
- Data sharing

6.1 Ethical aspects

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset/output?

no

6.1.2 Does the described dataset/output contain sensitive information?

Unknown

5 REFERENCES

All references and websites mentioned in the document were last checked for availability on 31.07.2023.

5.1 List of References

European Commission. (2021). *Horizon Europe Data Management Plan Template*. Version 1.0. 05.05.2021. https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/temp-form/report/data-management-plan_he_en.docx

Miksa, T.; Walk, P. & Neish, P. (2020). *RDA DMP Common Standard for Machine-actionable Data Management Plans*. <https://doi.org/10.15497/rda00039>

Simms, S.; Jones, S.; Mietchen, D. & Miksa, T. (2017). Machine-actionable data management plans (maDMPs). *Research Ideas and Outcomes* 3: e13086. <https://doi.org/10.3897/rio.3.e13086>

5.2 List of Websites

<https://argos.openaire.eu/splash/about/how-it-works.html>

<https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/reference-documents>

<https://www.eudat.eu>

<https://www.go-fair.org/fair-principles/>

<https://github.com/RDA-DMP-Common/RDA-DMP-Common-Standard/tree/master>

<https://www.openaire.eu>

<https://www.rd-alliance.org>

<https://zenodo.org>



APPENDIX 1. ARGOS HORIZON EUROPE TEMPLATE – LIST OF QUESTIONS

1. SUMMARY

1.1. BRIEF DESCRIPTION OF THE DESCRIBED RESEARCH OUTPUT

1.1.1 What kind of research output are you describing?

1.1.2 Is it physical or digital?

1.1.3 Are you generating or re-using it?

1.1.4 What is the type of the described dataset?

1.1.5 What is its format?

1.1.6 What is its expected size?

1.1.7 Why are you collecting/generating or re-using it?

1.1.8 What is its origin/provenance?

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

2. LINKS BETWEEN OUTPUTS

2.1. PUBLICATIONS

2.1.1 Does the described output support any scientific publication?

2.1.2 Is there a data availability statement provided along with the publication?

2.2. DATASETS

2.2.1 Does the described output use or support any published dataset?

2.3. SOFTWARE

2.3.1 Does the described output use or support any software?

3. FAIR PRACTICES

3.1. MAKING DATA AND OTHER OUTPUTS FINDABLE, INCLUDING PROVISIONS FOR METADATA

3.1.1. MAKING DATA FINDABLE, INCLUDING PROVISIONS FOR METADATA

3.1.1.1 What type(s) of persistent identifier(s) are used for the described dataset/output?

3.1.1.2 Will you provide metadata for the described dataset/output?

3.2. MAKING DATA AND OTHER OUTPUTS OPENLY ACCESSIBLE

3.2.1. REPOSITORY

3.2.1.1 In which repository will the dataset/output be deposited?

3.2.1.2 Is the selected repository a trusted source?

3.2.1.4 Add appropriate arrangements made with the repository(ies) where the described dataset will be deposited

3.2.1.5 Does the repository(ies) assign datasets/outputs with persistent identifiers?

3.2.1.6 Does the repository(ies) resolve the identifiers to a digital object?

3.2.1.7 Does the repository support versioning?

3.2.2. DATA

3.2.2.1 What is the described dataset/output title?

3.2.2.2 How is the dataset/output shared?

3.2.2.3 What is the reason of limiting access to the dataset/output?

3.2.2.4 If an embargo applies, please specify when the dataset/output will be made available.

3.2.2.5 Are there any methods or tools required to access the dataset/output?

3.2.2.6 Please provide information about the method(s) needed to access the dataset/output.

3.2.2.7 Please provide information about the tools needed to access the dataset/output.

3.2.2.8 Is the described dataset/output supported by a data access committee?

3.2.2.9 Please specify how the dataset/output will be accessed during and after the project ends

3.2.2.10 Please specify how long after the project has ended the dataset/output will be made accessible for

3.2.3. METADATA

3.2.3.1 Will you provide metadata even if the described dataset/output can not be openly shared?

3.2.3.2 Under which license will metadata be provided?

3.2.3.3 Do metadata provide information about how to access the described dataset/output?

3.2.3.4 Will metadata remain available after the dataset/output is no longer available?

3.3. MAKING DATA AND OTHER OUTPUTS INTEROPERABLE

3.3.1 Does your (meta)data use a controlled vocabulary?

3.3.2 If you created the vocabulary, where can it be found?

3.3.3 Have you applied a standard schema for your (meta)data?

3.3.5 What is the methodology followed?

3.3.6 What community-endorsed interoperability best practices are followed?

3.3.7 Does the described dataset/output provide qualified references with other outputs?

3.4. INCREASING DATA AND OTHER OUTPUTS REUSE

3.4.1 What internationally recognised licence will you use for your dataset/output?

3.4.2 What reusability and/or reproducibility methods are followed?

3.4.3 Will you provide the described dataset/output in the public domain?

3.4.4 Do you intend to ensure (re)use by third parties after your project finishes?

3.4.5 Is provenance well documented?

3.4.6 What documented procedures for quality assurance do you have in place?

4. ALLOCATION OF RESOURCES

4.1. ALLOCATION OF RESOURCES

4.1.1 What will be the cost of making the described output FAIR?

4.1.2 How will this cost be covered?



4.1.3 Identify the people who will be responsible and their role(s) in the management of the described output

5. SECURITY

5.1. DATA SECURITY

5.1.1 What security measures are followed?

5.1.2 What conditions do the security measures meet?

5.1.3 How will you preserve the described dataset/output in the long term?

6. ETHICAL ASPECTS

6.1. ETHICAL ASPECTS

6.1.1 Are there any ethical or legal issues that can have an impact on sharing the described dataset/output?

6.1.2 Does the described dataset/output contain sensitive information?

6.1.3 Does the described dataset/output contain personal data?

6.1.4 What are the methods used for processing and accessing sensitive/personal information?

7. OTHER ISSUES

7.1. OTHER

7.1.1 Do you make use of other procedures for data management?

7.1.2 Documentation of other procedures

