CRAFT-OA | eosc

D1.7 First update DMP

Submission Date 2024.06.28	Version 1.0 – Submitted version
Due Date 2024.06.30	
Deliverable Title	First update DMP
Deliverable No.	D1.7
Lead beneficiary	IBL PAN
Contributing WP	WP1
Type	DMP – Data Management Plan
	HORIZON-INFRA-2022-EOSC-01 Grant Agreement: 101094397

PU	Public	X
SEN	Sensitive	
R-UE/UE-R	EU classified	

DOI:10.5281/zenodo.12607329

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
Authors	Nikodem Wołczuk (https://orcid.org/0000-0002-4303-2016), Sona Lisa Arasteh (https://orcid.org/0000-0002-5725-1922), Clara Armengou (https://orcid.org/0000-0003-2124-5165), Margo Bargheer (https://orcid.org/0000-0001-8246-8210), Arnaud Gingold (https://orcid.org/0009-0000-0466-5542), Sy Holsinger (https://orcid.org/0009-0005-8978-225X), Tabea Klaus (https://orcid.org/0000-0002-2791-1053), Maxim Kupreyev (https://orcid.org/0000-0002-8472-6954), Pierre Mounier (https://orcid.org/0000-0003-0691-6063), Lisa Müller (https://orcid.org/0009-0008-3029-1888), Antti-Jussi Nygård (https://orcid.org/0000-0002-1534-109X), Elio Pellin (https://orcid.org/0000-0002-4076-6743), Leonidas Pispiringas (https://orcid.org/0000-0002-7111-094X), Hanna Varachkina (https://orcid.org/0000-0002-5832-3983), Theresa Waldmann (https://orcid.org/0009-0002-2994-6660)

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	2024.06.01	Author: Nikodem Wołczuk (IBL PAN)	Initial draft
0.2	2024.06.14	Authors: Sona Lisa Arasteh (MWS), Clara Armengou (DOAJ), Margo Bargheer (UGOE), Arnaud Gingold (OPERAS-AMU), Sy Holsinger (OPERAS AISBL), Tabea Klaus (UGOE), Maxim Kupreyev (OPERAS), Pierre Mounier (OPERAS-AMU), Lisa Müller (UGOE), Antti-Jussi Nygård (TSV), Elio Pellin (UBERN), Leonidas Pispiringas (OpenAIRE AMKE), Hanna Varachkina (UGOE), Theresa Waldmann (UGOE)	Internal review of new structure and additions to DMP by WP leaders and others involved
0.3	2024.06.18	Author: Nikodem Wołczuk (IBL PAN)	Version submitted to 1st review
0.3	2024.06.20	Reviewer: Cezary Rosiński (IBL PAN)	1st review
0.4	2024.06.21	Author: Nikodem Wołczuk (IBL PAN)	Comments from review integrated, Version submitted to 2nd review
0.4	2024.06.21	Reviewer: Birgit Schmidt (UGOE)	2nd review
0.5	2024.06.25	Author: Nikodem Wołczuk (IBL PAN)	Comments from review integrated, submitted for final formal check
0.6	2024.06.25	Contributors: Tabea Klaus (UGOE)	Final formal check

Funded by the European Union

		Lisa Müller (UGOE)	
1.0	2024.06.28		Submitted version

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.

List of Acronyms

D	Deliverable
DDH	Diamond Discovery Hub
DMP	Data Management Plan
DOC	Word Document Format
DOI	Digital Object Identifier
EC	European Commission
EOSC	European Open Science Cloud
FAIR	Findability, Accessibility, Interoperability and Reusability
GDPR	General Data Protection Regulation
IPSPs	Institutional Publishing Service Providers
IPTPs	Institutional Publishing Tools and Technology Providers
JSON	Java Script Object Notation
maDMP	machine actionable Data Management Plan
OA	Open Access
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OJS	Open Journal Systems
PDF	Portable Document Format
PID	Persistent Identifier
PKP	Public Knowledge Project
RDA	Resource Description and Action
T	Task
WP	Work Package
XML	Extensible Markup Language

TABLE OF CONTENT

1 EXECUTIVE SUMMARY

Deliverable (D) 1.7 “First Update 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. D1.7 is a document based on the original CRAFT-OA project Data Management Plan (DMP) (D1.2) but updated with information acquired during the first 18 months of the project.

The Deliverable 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). This section also contains information on the changes made during the first DMP update.
- The second section provides a general overview of the project data, which is the basis for 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, revision and amendment. This document is one of the versions produced during the project and at every stage of the project may be characterised by a certain level of incompleteness or generality in some areas that will be improved as the project proceeds. All versions of DMP will be made publicly available on the Argos[1] platform and in CRAFT-OA’s community in the Zenodo[2] repository.

[1] https://argos.openaire.eu/splash/about/how-it-works.html

[2] https://zenodo.org/communities/craft-oa/

Funded by the European Union

2 METHOD

In large-scale projects, such as CRAFT-OA, it is important to adopt an adequate strategy when preparing the DMP. With large and diverse amounts 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 EC's[3] guidelines.

2.1 Data Management Plan requirements

It is important to specify what scope the DMP should cover. Any research data that are 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), personal data, software developed, etc. The described output data are 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 FAIR data principles[4] 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 standardise the information produced in this way, it was decided to look for a tool that would enable it. The answer to this type of need is a machine actionable Data Management Plan (maDMP) which is an extension of the standard DMP approach[5].

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 available in the Horizon Europe (HORIZON) Reference Documents: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/how-to-participate/reference-documents

[4] Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3:160018. https://doi.org/10.1038/sdata.2016.18

[5] More information: Miksa, T., Walk, P., & Neish, P. (2020). *RDA DMP Common Standard for Machine-actionable Data Management Plans*. Zenodo. https://doi.org/10.15497/rda00039 and 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

and reuse of information contained in the document. Relying on ma DMPs allows integration of standardised information, e.g. employing persistent identifiers (PIDs), and ensures the unification of at least part of the information.

2.3 Tool

The Argos platform, an online service developed by OpenAIRE[6] and EUDAT[7] that can be used to create a maDMP and manage its subsequent versions, was selected for the creation of the CRAFT-OA DMP. It allows the simultaneous collaboration of many users while working on one DMP. Working with Argos consists of 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), Word Document Format (DOC), Extensible Markup Language (XML) and Resource Description and Action (RDA) Jaca Script 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 WPs. Leaders entered dataset information into Argos using available questionnaires using the Horizon Europe template (based on EC guidelines) as the basis. WP leads used their knowledge of the tasks outlined in the WPs they led to present the data details as reliably as possible. After preparing the datasets, the descriptions were additionally verified. The result is information that represents the management of data on the current project stage within individual WPs.

As far as possible, data of similar type, structure, origin and characterised 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

2.5 First update

After 18 months of the project, the original content of the DMP was reviewed to ensure that its structure and the information contained therein were clear and possibly well reflected in the approach to the data produced by the project.

When analysing the existing structure of the data descriptions in the DMP, the project team concluded that this structure was too broad and thus blurring the picture for potential audiences. The previous approach was based on descriptions of data created in individual WPs, which made the information difficult to manage and, due to the specificity of the materials, risked overlapping the described resources.

It was decided to create new descriptions, more focused on the similarity of certain groups of data and the method of storing and sharing them. The information contained in the 14 existing descriptions was analysed and combined into groups with similar characteristics. This resulted in seven descriptions of the data handled in the project:

- Project documentation
- Personal and contact data
- Analyses, requirements, recommendations, technical documentations, workflows
- Workshop and training materials
- Survey data
- Software
- Metadata

Following the changes to the Argos software, the descriptions have been made available to the WP leaders for review and to make the necessary changes and updates.

According to the project team, the new, simplified structure of the DMP will allow for better management of project data and will increase transparency of CRAFT-OA towards its users.

Funded by the European Union

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 (T) assigned in the project, some types of material are common concerning 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** – collaboratively produced materials for coordinating activities, such as notes from project meetings, are mainly produced and stored in the project management system Atlassian Confluence[9] or in Google Documents format and saved and stored on institutional servers e.g. as Word files 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.
- **Analyses, requirements, recommendations etc.** – all qualitative project documentation including analyses, recommendations or guidelines that have been produced during the project, whether intended to be made public (e.g. in the form of deliverables) or so produced only for the project activities. The storage method is Google Drive project space in Google Docs, Google Sheets or similar formats. Documents that were planned to be made available to users in advance, or those that have gained this status by their substantive content value, will be made public in an open repository under 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 following 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 number 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[10] principles. All personal data processed in the course of the project will be handled

[9] https://www.atlassian.com/software/confluence

[10] The FAIR principles provide guidelines to improve the Findability, Accessibility, Interoperability and Reusability (FAIR) of digital assets.

in compliance with GDPR. Details of the datasets produced by each WP 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 1 of the CRAFT-OA DMP in Argos. At the various stages of the project, not all data can be described completely. Still, 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 stage of the project or the specificity of the data). Some descriptions (or questionnaire parts) will be elaborated in subsequent versions of the DMP as the project develops. The template used to create dataset descriptions in Argos is Horizon Europe template[11]. For all the questions provided by the template read Appendix 1: Argos Horizon Europe template – list of questions.

4.1 Project documentation

This dataset contains internal project documentation and documents related to communication between project teams. It includes any document produced in any WP to coordinate work on tasks (e.g. rolling minutes, internal technical documentation), outputs of the project (deliverables, reports etc. describing the internal functioning of the project, which cannot be regarded as information available to the public) as well as other data created or collected for internal management.

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?

[11] 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.

New

1.1.5 What is its format?

Google Document Link File, .doc/.docx, .pdf

1.1.6 What is its expected size?

<1GB

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

• To share information

• To keep on record

• To make informed decisions

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

3.2.2 Data

3.2.2.2 How is the dataset / output shared?

Closed

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

The documents described are for internal project management only. Some project outputs (deliverables) are intended for internal purposes and shared only with the consortium and the European Commission.

3.2.3 Metadata

3.2.3.2 Under which license will metadata be provided?

Creative Commons Zero (CC0)

4.1 Allocation of resources

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

0

Euro

Storage

7.1 Other

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

No

4.2 Personal and contact data

This dataset includes personal data collected for project management as well as internal and external communication purposes from all WPs. It may contain mail addresses, names, affiliations, post addresses, sociodemographic data, mailing lists, event registration, sign-ups for newsletters, chat logs, meeting recordings, device IDs, location, and IP addresses. Collected data will be processed and stored protected enough to comply with GDPR requirements.

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

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 at 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.

In addition to the personal data, most of which was newly collected in the course of the project, the contact details from the DIAMAS project[12] of Institutional Publishing Service Providers (IPSPs) and Institutional Publishing Tools and Technology Providers (IPTPs) were used in selected cases, e. g. for the distribution of the survey in WP4. As participants from the DIAMAS survey, they had explicitly consented to further contact and inclusion in the IPTP/IPSP registries, which will be developed in the two projects. The use and processing of the reused data was also performed in a GDPR-compliant way.

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 after their consent. According to GDPR, it is possible at any time for project members and externals to withdraw consent from

[12] https://diamasproject.eu/

data 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 with the coordination and communication task force. 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.

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 long as the consent of the person is not withdrawn before) as long as necessary following the guidelines of the EC and the requirements of the partner institutions.

Only colleagues granted access by the coordination team can access the personal data for management purposes. This data is stored on institutional servers with backup routines and can be accessed using 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 by the Zenodo repository. The personal data of the project will only be stored (as long as the consent of the person is not withdrawn before) as long as necessary following 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

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

No

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 by the coordination team.

4.1 Allocation of resources

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

0

Storage

Indirect cost

no additional costs will be involved

4.1.2 How will this cost be covered?

Collaboration with other Projects

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

The data protection officer (DPO) for CRAFT-OA is Prof. Dr. Andreas Wiebe, DPO at 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-internally. Sociodemographic data concerning the website will be collected via a Matomo plug-in, meaning they are GDPR-

compliant. In case events are recorded, participants receive a legal disclaimer informing them about their rights and can choose to attend the event without being recorded. X (Twitter) Analytics, a third-party service, is used to collect data for KPIs from X (Twitter). Personal data will not be shared via Twitter except with the 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 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.

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

4.3 Analyses, requirements, recommendations, technical documentations, workflows

This dataset contains all project outputs such as analyses, requirements, recommendations, technical documentations, workflows etc. intended to be made available to users for further use.

It contains, for example, the results of the following Work Packages:

WP2: D2.2 Reusable curriculum for upskilling trainings; D2.3 Technical Gap Analysis & High-level Community Transition Plans

WP3: D3.1 Report on standards for best publishing practices and basic technical standards, inspired by the FAIR principles; D3.2 Report on challenges and help measures faced by OA journals and platforms

WP4: T4.4 – deployment and upgrading toolkits and recommendations for OJS and Lodel, based on a requirement survey.

WP5: 1) T5.1 – texts and tables documenting the technical requirements and recommendations of citation indexes and aggregators. 2) T5.2 – 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.

WP6: Recommendation for publishers for handling research supporting data; Federated Trust and Identity (T&I) technical requirements, architecture and best practices

WP7: Toolkit for IPTP Network; Institutional publishing technical living handbook

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

For task T5.1: The dataset will consist of a collection of freely available public information about technical requirements for publications referencing.

1.1.4 What is the type of the described dataset?

Derived or compiled

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

For task T5.1: The dataset will be used for: 1) informing the target group (diamond publishers) about the aggregators' technical requirements; 2) guiding the WP members in building the service that will help reference the publications on various service providers' platforms.

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

- Researchers
- Research communities
- Decision makers
- Industry
- Other

For WP3: Publishing platforms, publishers, and stand-alone Diamond OA journals

For task T5.1: The dataset targets mainly two groups within the research community: the diamond publishers, and the partners involved in the building of support services for the diamond publishers.

2.1 Publications

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
- Projects identifiers

DOI

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

Yes

For task T5.1: 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?

Zenodo CRAFT-OA Community

https://zenodo.org/communities/craft-oa/

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?

For task T5.1: Yes, to the file (textual, spreadsheets, presentation, etc.)

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: Analyses, requirements, recommendations, technical documentations, workflows

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?

Yes

Couldn't find it? Insert it manually

For task T5.1: 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?

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?

Yes

For task T5.1: 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
- Archiving
- Re-use
- Security

4.1.2 How will this cost be covered?

- Infrastructure Grant
- Other

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)

T5.2 Visibility Pathfinder

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

T5.2 Visibility Pathfinder

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

T5.2 Visibility Pathfinder

5.1 Data Security

5.1.1 What security measures are followed?

• Firewall

• Passwords

Public versions of the deliverables 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

4.4 Workshop and training materials

This dataset includes workshop and training materials produced in Work Packages 2, 3, 4 and 7.

Examples of elements of this dataset:

WP2: Training and Engagement Materials Package

WP3: Training materials on implementing technical standards and self-assessment toolkit

WP4: Part of the data created in task T4.4 will include online workshops in the form of “Meet the Experts” aimed at journal managers and technical staff.

WP7: Materials related to the outreach, dissemination and engagement activities, e.g. the recordings of the cross-project webinars organised in collaboration with the DIAMAS and

PALOMERA project[13] or the "Shaping Diamond OA: Criteria for Journals"[14] webinar held together with the DIAMAS 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

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?

20MB

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

[13] https://operas-eu.org/projects/palomera/

[14] Shaping Diamond OA: Criteria for Journals, 16 May 2024, https://www.youtube.com/watch?v=TpGbvrHT8VI

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

4.5 Survey data

This dataset contains data obtained through surveys carried out during the project in Work Packages 3, 4 and 7. These may include surveys targeting institutional IT departments and surveys targeting editorial staff of scholarly journals.

An additional set of data is a large extent of semi-structured interviews provided by the University of Bern 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?

For University of Bern data:

- 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

For University of Bern data:

The data are essentially interviews conducted with 11 platform operators and journal editors.

The qualitative study, conducted 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 start-ups or reorganizations of platforms and journals.

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

• Researchers

• Research communities

• Industry

• Other

For University of Bern data: 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 start-ups or reorganizations of platforms and journals.

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

• Researchers identifiers

• Organizations identifiers

DOI

ORCID

ROR

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

Yes

All data will be published on Zenodo. Zenodo is an 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.

3.1.1.3 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/

For University of Bern data: The DataCite Metadata Schema is used as the standard metadata schema for all published data.

3.1.1.6 Are the metadata searchable?

Yes

3.1.1.7 How are searchable metadata provided?

• Registry/Catalogue

• Linked Open Data

3.1.1.8 Are keywords provided in the metadata?

Yes

For University of Bern data: All results published on Zenodo or BORIS will be 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 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 an 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.

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

For University of Bern data: The questionnaire templates, the anonymized transcriptions, and the report will be made publicly available. The individual 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 on a local drive at the University of Bern and deleted after the end of the project.

For other surveys: The surveys will seek consent for data sharing.

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

For University of Bern data: The questionnaire templates, the anonymized transcriptions, and the report will be made publicly available. The individual 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

For University of Bern data: A README file with the information relevant to the project will be published together with the data. For this purpose, the following template will be 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

For University of Bern data: 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?

For University of Bern data: 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

For University of Bern data: 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 the project team check transcription and anonymization.

7.1 Other

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

No

4.6 Software

This dataset includes all of the software components developed in Work Packages 4, 5 and 6.

Examples of components being developed in individual WPs:

WP4: In tasks T4.1, T4.2 and T4.3 it is planned to contribute to the Public Knowledge Project (PKP) applications (Open Journal Systems (OJS), Open Monograph Press (OMP), Open Preprint Systems, PKP shared library, PKP plugins) and develop independent plugins for PKP

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

WP5: OJS plugin for the visibility pathfinder ready as MVP; Diamond Discovery Hub (DDH).

WP6: OJS connector for OpenAIRE Research Graph; OJS plugin for the European Open Science Cloud (EOSC) Interoperability Framework on Research Product Publishing; OJS plugins for integrating feedback from EOSC catalogue; OpenAIRE Publisher Dashboard Service.

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?

20MB

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 on 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

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

4.7 Metadata

This dataset contains all metadata produced and processed during the project.

WP5: T5.3 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 OA journals. The DDH dataset will therefore comprise metadata of journals and articles, some components of the OpenAIRE Graph, and potentially other sets of data.

OpenAIRE Graph Dump: The OpenAIRE Graph is one of the largest open scholarly record collections worldwide, all linked together, contextualised and openly available. 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 450 M 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, and DataCite. It also contains metadata from repositories registered in OpenDOAR, re3data.org, FAIRSharing.org, and the EOSC Service Catalogue. Among these, are prominent repositories such as UKPubMed, ArXiv, HAL, Zenodo, Figshare, Dryad, and RePEc.

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

Metadata in the Diamond Discovery Hub and from OpenAIRE Graph will be re-used and redesigned for new purposes.

1.1.4 What is the type of the described dataset?

Derived or compiled

1.1.7 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

1.1.8 What is its origin / provenance?

For task T5.3: Existing publishing platforms and services generating standard metadata for their publications.

For OpenAIRE Graph: the OpenAIRE Graph aggregates around 450 M 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, and DataCite. It also contains repositories registered in OpenDOAR, re3data.org, FAIRSharing.org, and the EOSC Service Catalogue. Among these, are prominent repositories such as UKPubMed, ArXiv, HAL, Zenodo, Figshare, Dryad, and RePEc.

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

- Researchers
- Research communities
- Decision makers
- Education
- The public
- Industry
- 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

• Reference

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

7.1 Other

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

No

5 List of References and Websites

All references and websites mentioned in the document were last checked for availability on 28.06.2024.

5.1 List of References

Miksa, T., Walk, P., & Neish, P. (2020). *RDA DMP Common Standard for Machine-actionable Data Management Plans*. Zenodo. 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

Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3:160018. https://doi.org/10.1038/sdata.2016.18

5.2 List of Websites

http://about.zenodo.org/policies/

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

https://diamasproject.eu/

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

https://github.com/OpenEdition/lodel

https://github.com/pkp

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

https://operas-eu.org/projects/palomera/

https://schema.datacite.org/

https://www.atlassian.com/software/confluence

https://www.eudat.eu

https://www.openaire.eu

https://www.rd-alliance.org

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

https://www.youtube.com/watch?v=TpGbvrHT8VI

https://zenodo.org/

https://zenodo.org/communities/craft-oa/

APPENDIX 1. ARGOS HORIZON EUROPE TEMPLATE – LIST OF QUESTIONS

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