Web Tutorial 1: Research data management in EPM projects

Michaela Kuepferling

INRiM

RDM in JRP/CSP proposals

Breakout session 1

Open Science in EPM

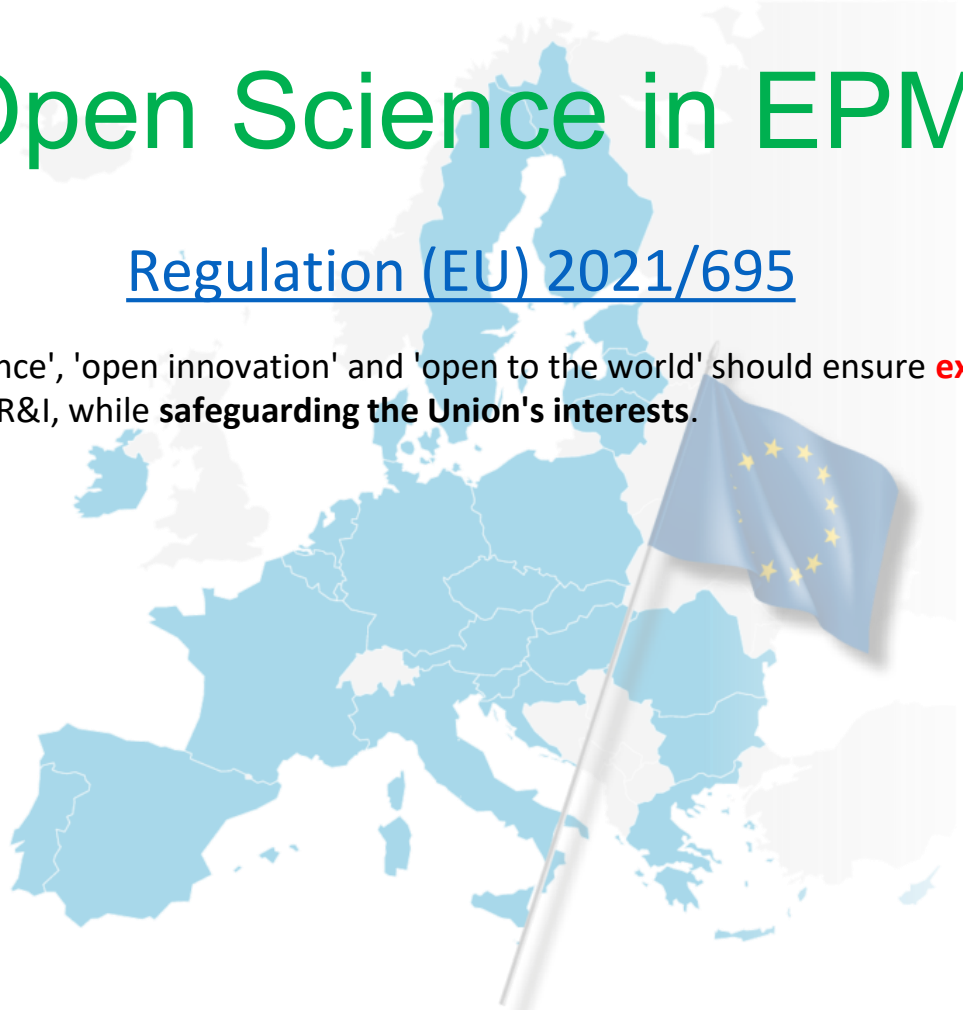
The Partnership refers directly to the Horizon Europe programme

Regulation (EU) 2021/695

Open Science in EPM

Regulation (EU) 2021/695

The concepts of 'open science', 'open innovation' and 'open to the world' should ensure **excellence** and the **impact** of the Union's **investment** in R&I, while **safeguarding the Union's interests**.



Open Science in EPM

Regulation (EU) 2021/695

The concepts of 'open science', 'open innovation' and 'open to the world' should ensure **excellence** and the **impact** of the Union's **investment** in R&I, while **safeguarding the Union's interests**.

- open access to scientific publications and research data to increase the **quality, impact** and benefits of science
- **accelerate** the advancement of knowledge by making it more **reliable**, efficient and accurate, more easily **understood** by society
- following the principle 'as open as possible, as closed as necessary', while ensuring the possibility of exceptions taking into account the legitimate interests of the beneficiaries
- responsible management of research data, which should comply with the 'FAIR principles', in particular through data management plans
- use of the European Open Science Cloud (EOSC) and the European Data Infrastructure

Open Science in EPM

Regulation (EU) 2021/695

The concepts of 'open science', 'open innovation' and 'open to the world' should ensure **excellence** and the **impact** of the Union's **investment** in R&I, while **safeguarding the Union's interests**.

- **open access** to scientific publications and research data to increase the **quality, impact** and benefits of science
- **accelerate** the advancement of knowledge by making it more **reliable**, efficient and accurate, more easily understood by society
- following the principle **'as open as possible, as closed as necessary'**, while ensuring the possibility of exceptions taking into account the legitimate interests of the beneficiaries
- responsible management of research data, which should comply with the **'FAIR principles'**, in particular through data management plans
- use of the European Open Science Cloud (EOSC) and the European Data Infrastructure

Open Science in EPM

Regulation (EU) 2021/695

The concepts of 'open science', 'open innovation' and 'open to the world' should ensure **excellence** and the **impact** of the Union's **investment** in R&I, while **safeguarding the Union's interests**.

- **open access** to scientific publications and research data to increase the **quality, impact** and benefits of science
- **accelerate** the advancement of knowledge by making it more **reliable**, efficient and accurate, more easily understood by society
- following the principle **'as open as possible, as closed as necessary'**, while ensuring the possibility of exceptions taking into account the legitimate interests of the beneficiaries
- responsible management of research data, which should comply with the **'FAIR principles'**, in particular through data management plans
- use of the **European Open Science Cloud (EOSC) and the European Data Infrastructure**

Open Science in EPM

Regulation (EU) 2021/695

The concepts of 'open science', 'open innovation' and 'open to the world' should ensure **excellence** and the **impact** of the Union's **investment** in R&I, while **safeguarding the Union's interests**.

- **open access** to scientific publications and research data to increase the **quality, impact** and benefits of science
- **accelerate** the advancement of knowledge by making it more **reliable**, efficient and accurate, more easily understood by society
- following the principle **'as open as possible, as closed as necessary'**, while ensuring the possibility of exceptions taking into account the legitimate interests of the beneficiaries
- responsible management of research data, which should comply with the **'FAIR principles'**, in particular through data management plans
- use of the **European Open Science Cloud (EOSC) and the European Data Infrastructure**

open science practices as an integral part of the proposed methodology for your project

RDM in Horizon Europe

HEU-annotated grant agreement 01.05.2024

COMMUNICATION, DISSEMINATION, OPEN SCIENCE AND VISIBILITY (— ARTICLE 17)

- Open science: open access to scientific publications
- Open science: research data management
- Open science: additional practices

RDM in Horizon Europe

HEU-annotated grant agreement 01.05.2024

Research data management — The process within the research lifecycle that includes the organisation, storage, preservation, security, quality assurance, allocation of persistent identifiers (PIDs) and rules and procedures for sharing of data including licensing.

RDM in Horizon Europe

HEU-annotated grant agreement 01.05.2024

- **Open science: research data management**

Beneficiaries must manage responsibly the digital research data generated in the action (‘data’) in line with the **FAIR (Findable, Accessible, Interoperable, Reusable) principles**. They should also ensure open access to research data via a trusted repository under the principle **‘as open as possible, as closed as necessary’**.

RDM in Horizon Europe

HEU-annotated grant agreement 01.05.2024

- **Open science: research data management**

Beneficiaries must manage responsibly the digital research data generated in the action ('data') in line with the **FAIR (Findable, Accessible, Interoperable, Reusable) principles**. They should also ensure open access to research data via a trusted repository under the principle **'as open as possible, as closed as necessary'**.

- Other research outputs other: software, algorithms, code, protocols, models, workflows, electronic notebooks etc.
- Reused data

RDM in Horizon Europe

[Programme guide Horizon]

RDM in Horizon Europe

European Commission

Programme guide Horizon

- Persistent **identifiers** (PIDs) are key in ensuring the findability of research outputs, including data.

RDM in Horizon Europe

European Commission

Programme guide Horizon

- Persistent **identifiers** (PIDs) are key in ensuring the findability of research outputs, including data.
- Standardised **metadata** frameworks: to enhance the potential reuse of research outputs.

RDM in Horizon Europe

Programme guide Horizon

- Persistent **identifiers** (PIDs) are key in ensuring the findability of research outputs, including data.
- Standardised **metadata** frameworks: to enhance the potential reuse of research outputs.
- Trusted **repositories**: for the deposition of and access to publications and research data.

RDM in Horizon Europe

European Commission

Programme guide Horizon

- Persistent **identifiers** (PIDs) are key in ensuring the findability of research outputs, including data.
- Standardised **metadata** frameworks: to enhance the potential reuse of research outputs.
- Trusted **repositories**: for the deposition of and access to publications and research data.
- Data management plans (**DMP**s): a cornerstone for responsible management of research outputs.

RDM in Horizon Europe

European Commission

Programme guide Horizon

- Persistent **identifiers** (PIDs) are key in ensuring the findability of research outputs, including data.
- Standardised **metadata** frameworks: to enhance the potential reuse of research outputs.
- Trusted **repositories**: for the deposition of and access to publications and research data.
- Data management plans (**DMP**s): a cornerstone for responsible management of research outputs.
- European Open Science Cloud (**EOSC**): open, trusted virtual environment enabling circa 2 million European researchers to store, share, process, analyse, and reuse research digital objects.

RDM in JRP/CSP proposals

RDM in JRP/CSP proposals

EUROPEAN PARTNERSHIP

Co-funded by the European Union

METROLOGY PARTNERSHIP

EURAMET

The management of research data and other research outputs is mandatory.

RDM in JRP/CSP proposals

EUROPEAN PARTNERSHIP

Co-funded by the European Union

METROLOGY PARTNERSHIP

EURAMET

The management of research data and other research outputs is mandatory.

Guide 4/7: Writing JRPs/CSPs

RDM in JRP/CSP proposals

Guide 4/7: Writing JRPs/CSPs

4.7.6 Section B2.f: Open science

(1 page maximum)

- Early and open sharing of research outputs. (e.g. Pre-print servers such as Zenodo, arXiv, Preprints, TechRxiv, ChemRxiv, BioRxiv and medRxiv; sharing with CIPM CCs, EURAMET TCs, standardisation TCs/WGs and OIML or WELMEC TCs)
- Measures to ensure the reproducibility of research outputs. (research outputs/tools/instruments needed to validate and guarantee access to results; accreditations or compliance with ISO/IEC 17025 or ISO/IEC 17034; round robin or comparisons)
- Providing open access to research outputs (such as data, software, models, algorithms, and workflows; also to a wider community, e.g. EOSC Portal or registered repositories
- Participation in open peer-review (use of Open Research Europe (ORE))
- Involving all relevant knowledge actors including citizens, civil society and end users in the co-creation of R&I agendas and contents (such as citizen science)

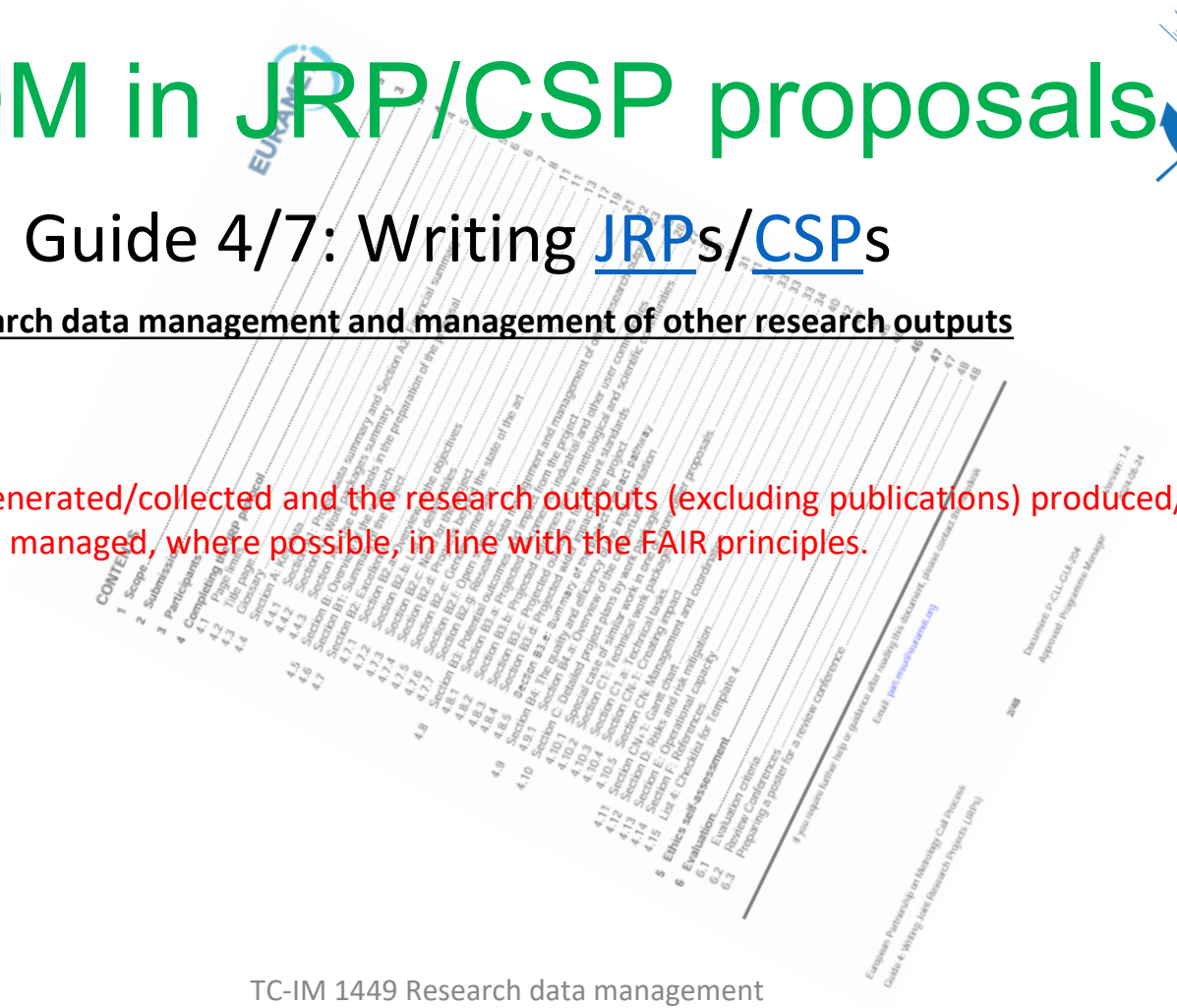
RDM in JRP/CSP proposals

Guide 4/7: Writing JRPs/CSPs

4.7.7 Section B2.g: Research data management and management of other research outputs

(1 page maximum)

Describe how the data generated/collected and the research outputs (excluding publications) produced/reused during the project will be managed, where possible, in line with the FAIR principles.



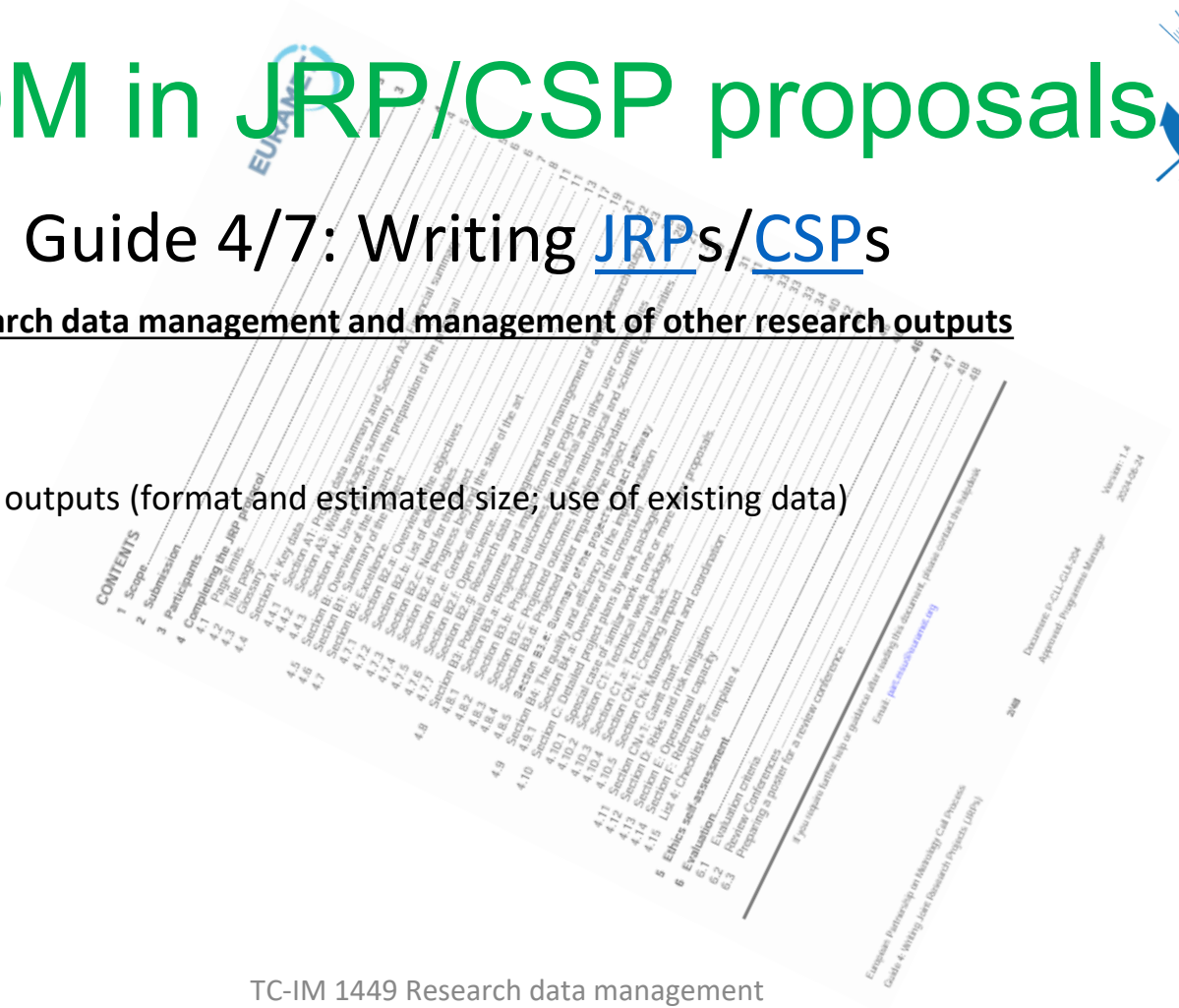
RDM in JRP/CSP proposals

Guide 4/7: Writing JRPs/CSPs

4.7.7 Section B2.g: Research data management and management of other research outputs

(1 page maximum)

- Types of data/research outputs (format and estimated size; use of existing data)



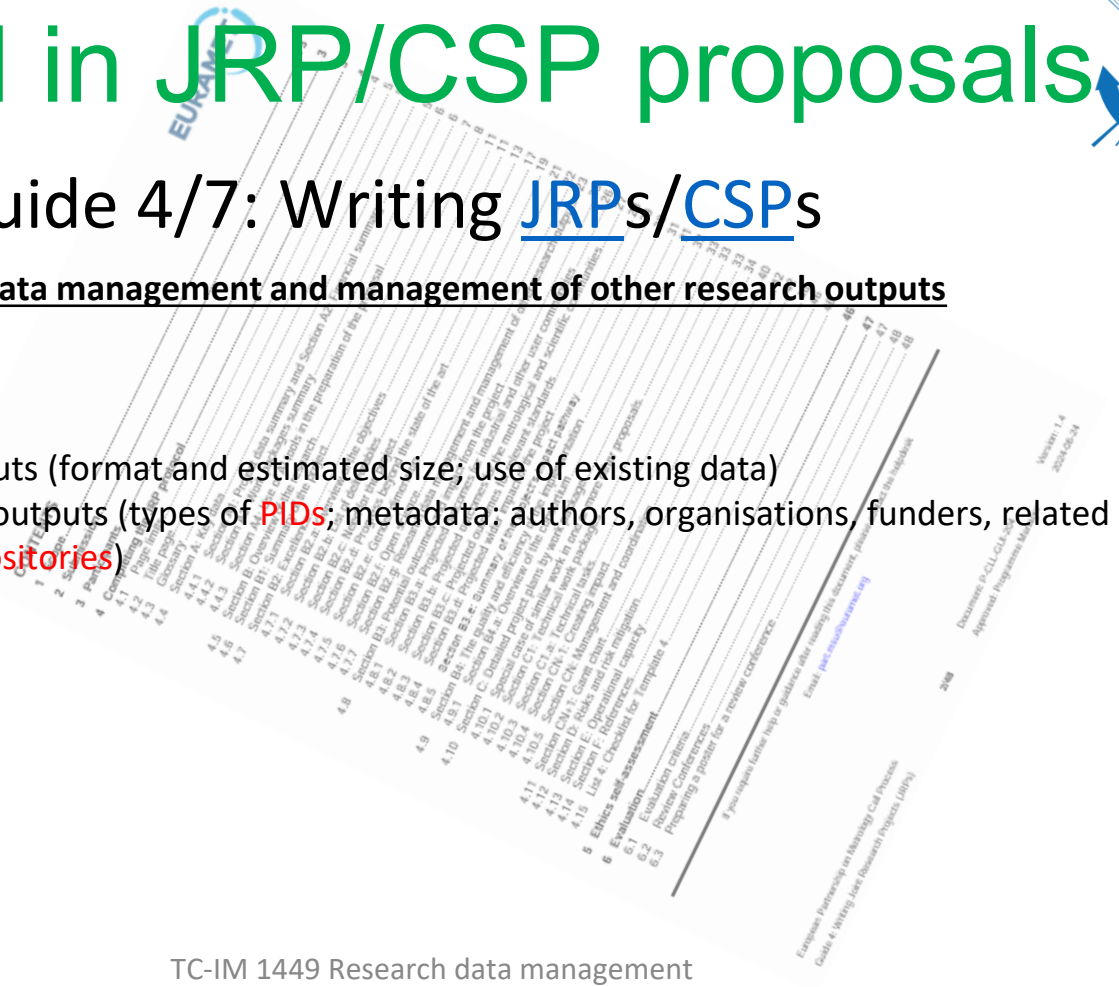
RDM in JRP/CSP proposals

Guide 4/7: Writing JRPs/CSPs

4.7.7 Section B2.g: Research data management and management of other research outputs

(1 page maximum)

- Types of data/research outputs (format and estimated size; use of existing data)
- Findability of data/research outputs (types of PIDs; metadata: authors, organisations, funders, related publications, research outputs; trusted repositories)



RDM in JRP/CSP proposals

Guide 4/7: Writing JRPs/CSPs

4.7.7 Section B2.g: Research data management and management of other research outputs

(1 page maximum)

- Types of data/research outputs (format and estimated size; use of existing data)
- Findability of data/research outputs (types of PIDs; metadata: authors, organisations, funders, related publications, research outputs; trusted repositories)
- Accessibility of data/research outputs (openly accessible of if and why restrictions apply; who will be responsible for decisions; e.g. Data Access Committee (DAC))

RDM in JRP/CSP proposals

Guide 4/7: Writing JRPs/CSPs

4.7.7 Section B2.g: Research data management and management of other research outputs

(1 page maximum)

- Types of data/research outputs (format and estimated size; use of existing data)
- Findability of data/research outputs (types of PIDs; metadata: authors, organisations, funders, related publications, research outputs; trusted repositories)
- Accessibility of data/research outputs (openly accessible of if and why restrictions apply; who will be responsible for decisions; e.g. Data Access Committee (DAC))
- Interoperability of data/research outputs (use of standards, formats and vocabularies for data/metadata)

RDM in JRP/CSP proposals

Guide 4/7: Writing JRPs/CSPs

4.7.7 Section B2.g: Research data management and management of other research outputs

(1 page maximum)

- Types of data/research outputs (format and estimated size; use of existing data)
- Findability of data/research outputs (types of PIDs; metadata: authors, organisations, funders, related publications, research outputs; trusted repositories)
- Accessibility of data/research outputs (openly accessible of if and why restrictions apply; who will be responsible for decisions; e.g. Data Access Committee (DAC))
- Interoperability of data/research outputs (use of standards, formats and vocabularies for data/metadata)
- Reusability of data/research outputs (use of licences for data sharing/re-use, e.g. Creative Commons, Open Data Commons; provide tools, software and models for data generation and validation, interpretation, and reuse)

RDM in JRP/CSP proposals

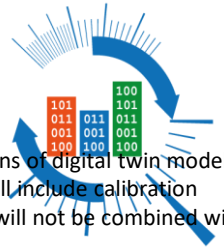
Guide 4/7: Writing JRPs/CSPs

4.7.7 Section B2.g: Research data management and management of other research outputs

(1 page maximum)

- Types of data/research outputs (format and estimated size; use of existing data)
- Findability of data/research outputs (types of PIDs; metadata: authors, organisations, funders, related publications, research outputs; trusted repositories)
- Accessibility of data/research outputs (openly accessible of if and why restrictions apply; who will be responsible for decisions; e.g. Data Access Committee (DAC))
- Interoperability of data/research outputs (use of standards, formats and vocabularies for data/metadata)
- Reusability of data/research outputs (use of licences for data sharing/re-use, e.g. Creative Commons, Open Data Commons; provide tools, software and models for data generation and validation, interpretation, and reuse)
- Participant, person or team responsible for data management and quality assurance Explain who will be responsible, including whether a DAC will be established.

Example 1: JRP



Types of data/research outputs

The data generated will be from spectroradiometric measurements, calibrations of spectroradiometers, comparisons of array spectroradiometers and validations of digital twin models. The project will collect the following data: images in JPEG format, numerical data in CSV format and text description data in Markdown format. Research outputs will include calibration methods, comparison protocols and digital twins. Existing data will be from: participants, the scientific literature, simulations and measurements. Project data will not be combined with existing data.

Findability of data/research outputs

Data/research outputs will be findable with a unique and persistent identifier (PID) e.g. DOI, GIT Commit/tag, Handle. The metadata will provide bibliographic information and information on funding and licensing terms. Where applicable, it will include PIDs for authors (ORCID), organisations (ROR, ISNI), funders (ROR, GRID, FundRef), and related publications and research outputs (DOI, URN, ISBN, Handle). The data/research outputs will be deposited in trusted repositories. Software and research protocols will be stored, documented and version-controlled in online repositories.

Accessibility of data/research outputs:

The data needed to validate the results presented in scientific publications/research outputs will be made openly available by default unless there is a specific reason not to publish them. Other data/research outputs will be made available on a case-by-case basis if relevant for third parties. The coordinator, relevant participant(s) and the project management board will be responsible for IPR / access considerations. Open access will be decided on a case-by-case basis in agreement with the data/research output owners (as confidentiality may be required for proprietary information). IPR will be managed in line with the DMP, Consortium Agreement, Grant
Agreement and DCE plan. Open access will be granted as soon as is reasonably possible (i.e. there may be IPR-related delays). Data and research outputs will remain accessible for the lifetime of the repository. Users will be required to acknowledge the project and the funding source, as per the CC BY license.

Interoperability of data/research outputs

The datasets will use the trusted repository's basic metadata schema for administrative data, which is compliant with the recommended standards used by DataCite, BASE search and OpenAIRE. To guarantee interoperability, individual datasets/research outputs will be described using affirmed discipline-specific vocabularies, standards, formats and methodologies. No mappings will be necessary: as the datasets will be described using standard terminologies. Datasets deposited in trusted repositories will include qualified references to other datasets from the same project and/or previous research.

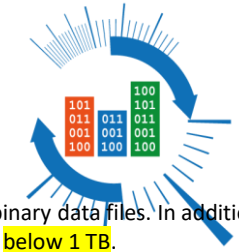
Reusability of data/research outputs

The data/research outputs will either be licensed using a CC BY license or a license with equivalent rights. Users will be required to acknowledge the consortium and the source of the data in any resulting publications. Templates will be documented with embedded instructions to maintain reusability. Data will be in a common format that can be read using widely available software. All data, published in open-access journals, will be usable by third parties after deposition in a trusted repository. Unpublished data will be made available for reuse on a case-by-case basis.

Participant, person or team responsible for data management and quality assurance

Instead of establishing a Data Access Committee (DAC), the coordinator/participants will manage the data/research outputs and quality assurance, will coordinate updates to the DMP and will decide which data/research outputs will be kept and for how long. The participant(s) that produce the data/research outputs will be responsible for organising its backup and storage, archiving and deposition in repositories.

Example 2: CSP



Types of data/research outputs

The data/research outputs generated will be mainly from reports, training materials. Examples of such data may include text data, images, Excel datasheets and binary data files. In addition, outputs will include publications, guides, documented calibration methods and reports. The estimated overall size of the data/research outputs is expected to be below 1 TB.

Findability of data/research outputs

The data/research outputs (reports, training material) will be findable as each will be identifiable with a DOI, Commit/tag, Handle, persistent and unique identifier. The metadata will provide information on the following: datasets (description, date of deposit, author(s), venue, and embargo); the EPM funding; grant project name, acronym and number; licensing terms; persistent identifiers and authors. Where applicable, the metadata will include persistent identifiers for related publications and research outputs. The data/research outputs will be deposited and published in trusted repositories located using the Registry of Research Data Repositories and, where applicable, on the project web platform (which will contain both a link to that repository and the material itself). Repositories will be selected appropriate to the character of the result/output.

Accessibility of data/research outputs

The consortium does not envisage IPR applying to the data, the relevant participant(s) will be responsible for any IPR considerations related to the research outputs. This will be managed using the DMP, the CA, the GA and the project's exploitation plan. The data/research outputs will remain accessible for the lifetime of the repository.

Interoperability of data/research outputs

The datasets will use the trusted repository's basic metadata schema for administrative data, which is compliant with the recommended standards used by Data Cite and Open AIRE. For individual datasets, the following discipline-specific vocabularies, standards, formats, and methodologies will be used. Mapping will not be required as the terminology used will be chosen to be compatible with the existing literature. The project's datasets that will be deposited in the chosen trusted repository will include qualified references to other datasets from the same project and/or previous research.

Reusability of data/research outputs

The data/research outputs will either be licensed under the latest available version of the CC BY license or a license with equivalent rights as set out in the GA. Users will be required to acknowledge the consortium and the source of the data in any resulting publications. Alternatively, the Public Domain Dedication License (CC 0) will be used.

Availability of tools, software and models for data generation and validation, interpretation, and re-use

The templates will be documented with embedded instructions to maintain reusability by minimising the number of files required. The data are in a common format and can be read using widely available software. Any data published in open-access journals will be usable by third parties after the datasets have been deposited in a trusted repository. The data that does not relate to peer-reviewed publications will be made available for re-use on a case-by-case basis.

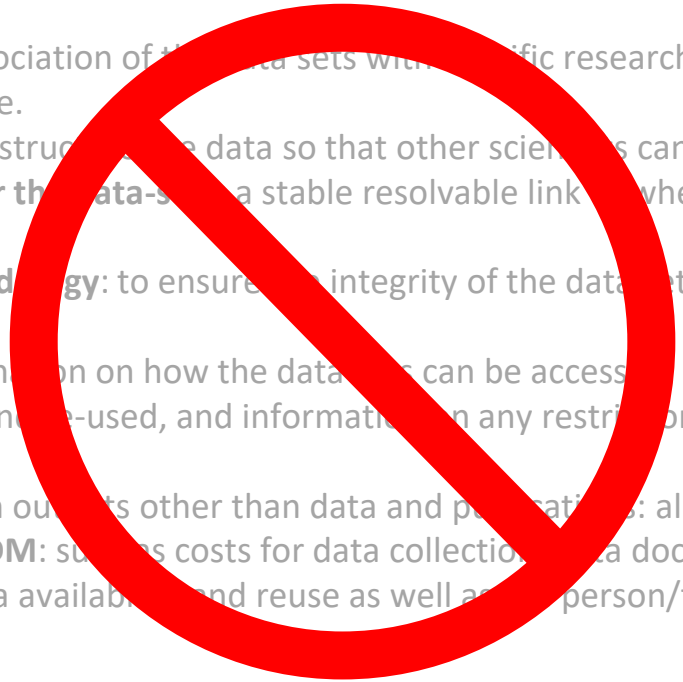
Participant, person, or team responsible for data management and quality assurance

Instead of establishing a Data Access Committee (DAC), the coordinator, with support from the Impact work package leader and other participants, will have overall responsibility for the management of data/research outputs and quality assurance. The coordinator will be responsible for updates to the data management plan and for deciding on a case-by-case basis which data/research outputs will be kept and for how long. The participant(s) that produced the data will be responsible for organising its backup and storage, archiving, and with deposition of the data/research outputs in the chosen repositories.

The DMP

• **Data set description**: allowing association of the data sets with specific research as well as information on data types and an estimate of the data set's size.
• **Standards and metadata**: used to structure the data so that other scientists can make an assessment.
• **Name and persistent identifier for the data-sets**: a stable resolvable link to where the data sets can be directly accessed.
• **Curation and preservation methodology**: to ensure the integrity of the data sets and the period during which they will be maintained.
• **Data sharing methodology**: information on how the data sets can be accessed, including the terms-of-use or the license under which they can be accessed and re-used, and information on any restrictions that may apply or relevant security and privacy considerations.
• **Output management**, for research outputs other than data and publications: all points apply to them.
• **Costs and personnel related to RDM**: such as costs for data collection, data documentation, data storage, data access and security, data preservation, data availability and reuse as well as the person/team responsible for data management and quality assurance processes.

The DMP



- **Data set description**: allowing association of the data sets with specific research as well as information on data types and an estimate of the data set's size.
- **Standards and metadata**: used to structure the data so that other scientists can make an assessment.
- **Name and persistent identifier for the data-set**: a stable resolvable link where the data sets can be directly accessed.
- **Curation and preservation methodology**: to ensure the integrity of the data sets and the period during which they will be maintained.
- **Data sharing methodology**: information on how the data can be accessed, including the terms-of-use or the license under which they can be accessed and re-used, and information on any restrictions that may apply or relevant security and privacy considerations.
- **Output management**, for research outputs other than data and publications: all points apply to them.
- **Costs and personnel related to RDM**: such as costs for data collection, data documentation, data storage, data access and security, data preservation, data availability and reuse as well as the person/team responsible for data management and quality assurance processes.

Questions?