Open science and research data management in Horizon Europe

Contents

Introduction

This document provides guidance on addressing the **open science** and **research data management** requirements when completing Horizon Europe funding applications.

It is for use by those applying funding under the **Horizon Europe** programme, including applicants for **Marie Skłodowska-Curie Action** (MSCA) and **European Research Council** (ERC) funding.

What you need to do

Horizon Europe and MSCA applications

If you are based at the University of Reading, or applying to be based here (e.g. for an MSCA Postdoctoral Fellowship), and you are **co-ordinating a proposal**, or a **sole applicant**, the **open science** and **research data management** sections of the application form (in part B 1.2) must be reviewed by the Research Data Manager prior to submission.

If you are not the co-ordinating partner on a collaborative proposal, the open science and research data management sections do not need to be reviewed by UoR. But if you will have significant data management responsibilities in the project, and you have been asked to draft or provide input into these sections, the Research Data Manager can review/advise as needed.

Draft application forms should be sent to the Research Data Manager directly or via your Research Development Manager no later than 5 working days before the application deadline.

All applications

Any costs included in the budget for **data storage**, **computing** and **archiving** must be approved by the Research Data Manager. The Research Data Manager can advise on any costs for **Open Access publishing**. Budgets (with the draft proposal) should be sent to the Research Data Manager directly or via your Research Development Manager no later than 5 working days before the application deadline.

Contact

Contact the Research Data Manager at researchdata@reading.ac.uk / 0118 378 6161.

General guidance on data management planning is available on the Research Data Management website.

Key documents

Standard application form (SAF), p. 33-34. Guidance on completing Part B, section 1.2 (Methodology) on open science and research data management.

Programme Guide (PG), p. 37-53. Detailed information about open science practices recommended by Horizon Europe, and how these are evaluated in the application.

Annotated model grant agreement (AGA), p. 152-161. Contains detailed information regarding Open Science requirements under Horizon Europe.

Data Management Plan template, recommended when drafting the Data Management Plan deliverable for awarded grants.

Overview of application form requirements

Horizon Europe standard applications

See: Standard application form (SAF), Part B 1.2

Applications within the main Horizon Europe programme must include sections, each up to one page in length, describing how appropriate open science practices will be implemented as part of the project's methodology, and how research data and outputs other than publications generated/collected during the project will be managed. Other parts of the application form where relevant information will need to be provided are: B 3.2 Capacity of participant and consortium as a whole; Tables 3.1b/c Work package description and Deliverables; and Section A List of publications and other outputs.

Marie Skłodowska-Curie Actions (MSCA)

See e.g.: Marie Skłodowska-Curie Actions Postdoctoral Fellowships application form

Applications for MSCA funding must include section up to half a page in length describing how appropriate open science practices will be implemented as part of the project's methodology, and how research data and outputs other than publications generated/collected during the project will be managed. See Part B 1.2 of the application form.

European Research Council (ERC)

See: ERC Work Programme 2021

ERC grant applications are evaluated on the criteria of scientific excellence alone, and there is no requirement to address open science and data management in the application.

Open science in Horizon Europe

See: PG, p. 37-53.

There are five **mandatory** open science practices within the Horizon Europe programme which must be addressed in applications:

- Immediate open access to scientific publications;
- Responsible management of research data in line with the FAIR principles, and open access to research data;
- Open information about the research outputs/tools/instruments needed to validate the conclusions of scientific publications or to validate/re-use research data;
- Open digital or physical access to results needed to validate conclusions;
- Immediate Open Access for all research outputs in case of public emergency.

There are also **recommended** open science practices endorsed by Horizon Europe. Adoption of these, where relevant to your project, **will result in a higher evaluation score**. These are:

- Measures to increase reproducibility: for example, providing access to software code, models, algorithms and workflows as well as data;
- Open peer review: a peer review model adopted by some publishers which may enhance the robustness and transparency of the peer review process;
- Citizen, civil society and end-user engagement: may be relevant for research that involves public participation in the design, delivery or assessment of research;
- Early and open sharing: making outputs accessible as early as possible is encouraged, especially where timeliness is an important factor, e.g. in health research;
- Preregistration: typically relevant for hypothesis-driven empirical research;
- Registered reports: a model for publishing empirical research offered by some publishers, where a study proposal is reviewed and accepted for publication before the research is undertaken;
- Preprints: early pre-peer review dissemination of results where timeliness is an important factor.

Refer to the PG for detailed information the above open science practices.

You can find out more about open science practices and their application by visiting the University of Reading's Open Research Handbook.

Completing the application form: Horizon Europe/MSCA

The following guidance describes what you need to do as the co-ordinating partner to complete the open science and data management sections in Part B 1.2 of the Horizon Europe SAF. It can also for completing the corresponding sections of an MSCA application, but note that for MSCA the combined length of both sections is not expected to exceed half a page, and can therefore be less detailed.

Section B 1.2 Open science and research data management

See: SAF, Part B 1.2

Open science

Describe how appropriate open science practices will be implemented as part of your methodology, adapted to the nature of your work, in a way that will increase the chances of the project delivering on its objectives. If you believe none of the open science practices are appropriate for the project, you must provide a justification. This section can be up to one page in length.

You should explain as a minimum how you will meet the **mandatory** open science requirements. Describing how you will adopt **recommended** practices, as appropriate to your project, will result in a higher evaluation score.

Some guidance on the mandatory open science practices follows. For more detailed guidance on mandatory and recommended open science practices, see PG, p. 37-53.

Open Access

The following text can be used (and adapted as required) to describe the project's Open Access policy:

> Peer-reviewed publications produced by the project will be made Open Access in accordance with the terms of the grant agreement. For each publication, an electronic copy of the accepted manuscript will be deposited in relevant institutional publications repositories, and made available under a CC BY or other acceptable licence. Metadata of deposited publications will be published under a CC0 licence, and will include all information required under the grant agreement.

Responsible management of research data in line with the FAIR principles

This will be addressed in more detail in the research data management section. Here it is sufficient to state that data will be managed in line with the FAIR principles as set out in the data management section.

Open information about the research outputs/tools/instruments

This requirement should be met by the use of repositories to preserve and share data, software code and other materials required to validate results. A repository will publish discoverable online information about an item, and in most cases will assign a persistent unique identifier such as DOI to the item, so that it can be cited and linked to from related research publications.

Open digital or physical access to results needed to validate conclusions

As above, this requirement should be met by the use of repositories, where it is expected that materials will be shared under an open licence, unless there is a reason to restrict access either temporarily (e.g. pending patent protection) or permanently (e.g. because of the sensitive nature of the data), in which case an item might be made available under more restrictive terms. Where this is the case, the reason for restriction should be specified.

Immediate Open Access for all research outputs in case of public emergency

This requirement is unlikely to apply in most cases, but might be met by use of preprints to accelerate publication of results before peer-reviewed publication, and by pre-publication release of important data or other outputs.

Research data management

Describe how research data and outputs other than publications generated/collected during the project will be managed in line with the FAIR principles (see below). This section can be up to one page in length.

You are expected to address each of the points below in this section.

Types of data*: (e.g. experimental, observational, images, text, numerical) and their estimated size; if applicable, combination with, and provenance of, existing data.*

Clearly identify and characterise each type of data and quantify in relevant terms, e.g. by sample size, number of records, etc. and digital data volume where it is substantial.

***Findability of data**: types of persistent and unique identifiers (e.g. digital object identifiers) and trusted repositories that will be used.*

Refer to University guidance on choosing a data repository. Most data repositories assign unique identifiers to datasets in the form of either digital object identifiers (DOIs) or accession IDs.

All University researchers are eligible to deposit data in the University of Reading Research Data Archive (up to 20 GB free of charge). There may be external services that are more suitable for your data if they serve subject communities or manage specific types of data. For example, the public databases of the European Bioinformatics Institute are the most suitable repositories for genetic data. The EC also provides its own repository for the outputs of EC-funded research, Zenodo. There is no charge to use this service.

Note: 'Personal websites and databases, publisher websites, as well as cloud storage services (Dropbox, Google drive, etc) are not considered repositories. Academia.edu, ResearchGate and similar platforms do not allow open access under the terms required and are NOT considered repositories' (AGA, p. 156). These services do not satisfy the Horizon Europe requirements.

***Accessibility of data**: IPR considerations and timeline for open access (if open access not provided, explain why); provisions for access to restricted data for verification purposes.*

You may wish to provide for protection of data that will be subject to commercial exploitation (e.g. patenting or licensing), but it should be possible to share these once IP protection has been secured.

Provide a timeline for making the data openly available. If you plan to restrict access to the data, explain what provisions will be put in place to enable access for verification purposes. Maintaining data in closed storage and undertaking to provide access on request is not encouraged as a permanent solution. If there is a need to restrict access to data, e.g. because although anonymised they present a higher risk of identification through possible linkage to other databases, then repositories that can manage controlled access to datasets should be used where possible. See guidance on where to archive restricted data.

***Interoperability of data**: standards, formats and vocabularies for data and metadata.*

Outline any standards that will be used for curation and documentation of your data. If you will be using specific data formats or metadata standards to describe particular data types, e.g. to comply with specialist data repository requirements or to conform to community standards, identify these and provide relevant information. (For example, the European Nucleotide Archive specifies data formats for submission.) Search for relevant standards at FAIRsharing. For more information about metadata standards, see the RDA Metadata Directory.

***Reusability of data**: Licenses for data sharing and re-use (e.g. Creative Commons, Open Data Commons); availability of tools/software/models for data generation and validation/interpretation /re-use.*

Research data should be made available under open licence where there is no reason to restrict access. This can be through the Creative Commons Attribution Licence (CC-BY 4.0), the Creative Commons Zero Public Domain Declaration (CC0) or any other suitable open licence.

Information about any tools/software/models instruments used to generate the data, and required to validate, interpret or re-use the data, should be provided in documentation accompanying the data deposited in the repository. Information must include a detailed description of the tool/instrument, how to access it, and any dependencies on commercial products. Where code has been written to generate or analyse the data, this should be archived, either alongside data or separately, under a suitable open licence wherever possible. A code repository platform (e.g. GitHub) is a valid development and collaboration environment, but versions of code used to generate published results should be archived to a suitable data or other digital repository. The University provides guidance on best practice in the management and sharing of code.

***Curation and storage/preservation costs**; person/team responsible for data management and quality assurance.*

If you will be collecting a large amount of data (as a guide, requiring more than 100 GB of storage capacity at UoR), you may need to include data storage costs in your budget. The amount requested should be consistent with the data volume anticipated under Types of data (above). For guidance on University storage services and costs see the section on storage and computing below.

Many data centres and data repositories do not charge for the deposit of data, but this is not always the case. It may depend on the volume of the data to be deposited. For example, TB-scale volumes of modelling output generated in climate/weather research could be deposited in the NERC Centre for Environmental Data Analysis (CEDA) Archive, but applicants would need to obtain a quote from CEDA. The Archaeology Data Service is a rare example of a data centre that levies a deposit charge as standard.

Costs for use of external services are eligible for EC grant funding.

Indicate who will be responsible for data management and quality assurance.

Section B 3.2 Capacity of participant and consortium as a whole

See: SAF, Part B, section 3.2, p. 40.

Show that your consortium includes expertise and/or track record in open science practices relevant to your project. This is not required if you have provided justification above that open science practices are not relevant for your project.

Tables 3.1b/c Work package description and Deliverables

Remember to include the DMP in your project deliverables. It must be submitted in the first six months of the award. An updated DMP deliverable must also be produced mid-

project (for projects longer than twelve months) and at the end of the project (where relevant).

Section A List of publications and other outputs

See: SAF, Part A, p. 13.

Publications are expected to be Open Access. Datasets and other outputs are expected to be FAIR and 'as open as possible, as closed as necessary'.

Costs for research data and Open Access

Storage and computing

You should consider any requirements you will have for resources related to the storage and processing of research data, and ensure all eligible costs are included in your budget. You will need to consider:

- how much data you will need to store during the project, where data will be stored, and any associated costs;
- whether any dedicated computing resource is required for computing-intensive proposals, and if so at what specification and cost.

Data collected/held at the University should be stored using University-managed infrastructure, which will provide data security, replication in separate data centres, automated backup and file recovery. For the different options available, and information about costs, please read the guidance here.

Costs would apply if you had a need to store substantial volumes of data at the University (as a guide > 100 GB) or to use the University's high-performance computing facilities.

If you have computing-intensive requirements, custom specifications of CPU, memory, storage and GPU can be purchased from the University on a pro rata basis. Information is available in the Academic Computing Team website. DTS can advise if you need assistance with your specification.

Open Access publishing

Any costs for publishing Open Access articles should be included in your budget where UoR is expected to be the corresponding institution for an article. Only costs for article processing charges (APCs) in fully Open Access journals or platforms can be charged to the project. Check that the chosen journal is reputable and fully Open Access by using the Directory of Open Access Journals (DOAJ) before submission.

As a ballpark figure you can estimate a cost of £3,500 including VAT per article. Note that the University has agreements with some publishers which allow corresponding authors based at the University to publish in their journals at no additional cost.
The SciFree tool can help with identifying journals included in these deals. Collaborating institutions may also have publisher agreements in place that would allow them to cover costs, providing the corresponding author was affiliated to them. If you know the journals you plan to publish in, you may be able to estimate any costs more precisely.

Any costs for Open Access books, monographs and edited volumes, and Open Access chapters if they are published in fully Open Access books or platforms, should be included in your budget.

Book publishing costs vary considerably between publishers. As a ballpark figure you can estimate a cost of £15,000 including VAT for a book processing charge (BPC). You should estimate £2,500 (including VAT) for a chapter processing charge (CPC). If you have already identified a potential publisher, you may be able to ask them for a more precise estimate.

If you have questions about costing up your publishing requirements, please contact oarequests@reading.ac.uk.

Ethics and security

See: SAF, Part A 4

You will be required to complete an Ethics Issues table as part of the application and, if you have identified any ethics issues related to aspects of your research (including, but not limited to, management of research data), you must also submit an ethics self-assessment. You can use this section of your application to explain how ethical aspects of data management will be handled, for example the measures you will adopt to ensure the secure storage and communication of sensitive data, or your proposed procedure for obtaining informed consent for long-term preservation of data and data sharing from research participants.

For guidance on completing this part of the application, see the guide How to complete your ethics self-assessment. You can also refer to the guidance on research ethics and data protection on the University's research data management website.

After the start of your project

Include a deliverable for an initial DMP at month 6 in the project

If your grant is awarded, you will be required to submit a first version of a full data management plan (DMP) within the first six months. Horizon Europe provides a DMP template for this purpose. The template includes questions to help you develop your plan with the appropriate level of detail.

The DMP is to be regarded as a living document, and should evolve as the project progresses. The initial deliverable will be a first iteration, which can be expanded in the course of the project. For projects of longer than 12 months an updated DMP deliverable is expected at the halfway point in the project, and a final updated DMP deliverable at the project end.

Where the University is the co-ordinating partner or has a substantial role in collection and processing of data, PIs are encouraged to contact the Research Data Manager for assistance in preparing the DMP.

If you anticipate you will not be able to deposit or share your data, a DMP must still be completed, within which the decision to restrict access to your data must be justified.

If you require support in writing the DMP deliverable, contact the Research Data Manager.

Deposit your data in an appropriate repository

Horizon Europe requires that research data be deposited in a data repository as specified and within deadlines set out in the DMP.

This does not necessarily mean data have to be made open, as access can be restricted where this is necessary. Some data repositories (including the University's Research Data Archive) can manage sensitive data under a controlled access policy. For more information see the Restricted data section of this web page.

But as a rule data should be made openly available, unless there is a good reason why this is not possible, which must be documented in your DMP. In cases where access to data will be entirely restricted, a metadata record describing the dataset should still be created, so that the existence of the data can be ascertained. The University's Research Data Archive can be used for this purpose

Open data should be licensed by default under the Creative Commons Attribution Licence (CC-BY 4.0), the Creative Commons Zero Public Domain Declaration (CC0) or any other suitable open licence. It is standard practice for data to be shared no later than any related results are published. In exceptional cases deposit can be delayed beyond the end of the project.

Make your publications Open Access

See: AGA, p. 155-158.

You may publish your research outputs in the venue of your choice, whether this is closed (subscription-based), hybrid, or Open Access, providing the conditions below are met.

- Your publication **must** be deposited in a trusted repository, in a machine-readable format. You should deposit your publication in CentAUR, the University of Reading's institutional repository, at the point of acceptance. Open Access to your publication must be provided immediately from the date of publication.
- The version deposited in the repository must be either the author final manuscript (incorporating all revisions following peer review) or the final published version, known as the version of record.
- The deposited version must be licensed under the latest version of a Creative Commons Attribution licence (CC-BY) or an equivalent licence. Monographs may use more restrictive licences excluding commercial uses and the creation of derivative works.
- Information must be included in your publication about any tools and instruments needed to validate your conclusions. These may include data, software, algorithms, protocols, models, workflows, electronic notebooks and other materials. Information should include a detailed description of the research output/tool/instrument, how to access it, any dependencies on commercial products, version/type, parameters, etc.