Horizon 2020 Data Management guidance

# Introduction

## Who is this for?

This guidance is for those co-ordinating collaborative applications for EC funding through the Horizon 2020 programme, and may also be relevant to partners in collaborative applications, especially where they will have a substantial role in collection and processing of research data in the project.

Open research data is the default for all thematic areas of Horizon 2020, as implemented through the Open Research Data Pilot. Projects supported by the European Research Council under Horizon 2020 may opt in to this Pilot on an individual and voluntary basis. Applicants for EC funding need to be aware of and address open data requirements in applications and awarded projects.

Where applicants opt in to the Open Research Data Pilot, they are required to complete the data management sub-section in the Impact section of the application. Awarded projects are required to submit a first version of a full Data Management Plan within the first six months of the project.

Projects can opt out of the Pilot on certain grounds, but applicants must be aware of the general obligations and allowable exemptions when applying for EC funding.

## EC guidance on Data management in H2020

The H2020 **Open access** guidance at <https://bit.ly/2bxhVbm>[[1]](#footnote-1) provides information about the scope of the Open Research Data Pilot and available opt-out exemptions.

The proposal and any parts relating to data management should be completed with reference to the **Data management** guidance at <https://bit.ly/2qjUABS>.[[2]](#footnote-2) This web page includes guidance on information to be provided in the proposal and the development of Data Management Plan deliverables during the project. You can also find here a Data Management Plan template, to be used by awarded projects to create the Data Management Plan deliverable. Links to guidance and a DMP template for ERC applicants are also provided.

The infographic **Open Research Data in Horizon 2020** (<http://bit.ly/2etuyHm>) explains clearly and simply the Open Research Data Pilot policy and how to comply.

## Proposal preparation and review

Where University of Reading applicants are co-ordinating a proposal, the data management section must be reviewed by Robert Darby, Research Data Manager, prior to submission. Draft sections can be sent to Robert directly or via your Research Development Manager and should be provided no later than 5 working days before the application deadline.

Contact Robert if you require preliminary guidance on completing the data management section or support in developing the Data Management Plan deliverable for your project. General guidance on data management planning is available on the Research Data Management website at <http://www.reading.ac.uk/res-data-management-planning.aspx>.

## Contact

Robert Darby, Research Data Manager: r.m.darby@reading.ac.uk / 6161.

# What is the scope of the Open Research Data Pilot?

## Areas of the programme

The Open Research Data Pilot applies across all thematic areas of the Horizon 2020 work programme, but in practice will be mostly relevant to Research and Innovation Actions (RIA), Innovation Actions (IA), and Coordination and Support Actions (CSA). Some instruments not suitable for data sharing are excluded:

* co-fund and prizes instruments;
* ERC proof of concept grants;
* ERA-NETs that do not produce data;
* SME phase 1 instruments.

For two-stage calls, information on participation in the Pilot will only be requested at Stage 2.

## Types of data covered

* The underlying data (with associated metadata) generated or collected by the project, i.e. those needed to validate the published results;
* any other data collected or generated by the project, e.g. data not directly associated with a publication, or raw data, that are judged to have value for sharing.

# What is required

Three key requirements may apply:

1. Opt in to or out of the Open Research Data Pilot in your proposal;
2. If you opt in to the Pilot, complete the data management section of your proposal;
3. If you opt in to the Pilot and the grant is awarded, submit a Data Management Plan deliverable within the first six months of the project.

## Opt in to or out of the Open Research Data Pilot

When completing a full proposal submission form you must indicate in the ‘Extended Open Research Data Pilot in Horizon 2020’ section whether you wish to opt in to or out of the pilot. If you opt out, you must indicate your reason(s) for not participating in the Pilot, i.e.:

* the project will not generate any data;
* to allow the protection of results (e.g. patenting);
* incompatibility with the need for confidentiality linked to security;
* incompatibility with privacy/data protection;
* achievement of the project’s main aim would be jeopardised;
* other reasons (to be specified).

Participation in the Pilot does not require all research data to be made open: you can opt in and exclude some data for any of the reasons above.

Applicants can also opt out of the Pilot completely at any stage (before or after the grant signature). In this case, the reason for opting out must be given.

Participation in the Pilot does not constitute part of the evaluation process. Proposals will not be penalised for opting out.

## Complete the data management section of the proposal

If you have opted in to the Pilot, in the Impact section of the main proposal you must include a sub-section detailing how participants will manage the research data generated and/or collected during the project, in particular addressing the following issues:

* What types of data will the project generate/collect?
* What standards will be used?
* How will this data be exploited and/or shared/made accessible for verification and re-use? If data cannot be made available, explain why.
* How will this data be curated and preserved?
* How will the costs for data curation and preservation be covered?

This section should be kept brief and to the point. You should not need more than half a page, or a page at most if large amounts of data will be involved and/or data management will be complex.

Note that measures to provide open access to peer-reviewed scientific publications which might result from the project should be addressed in a separate sub-section of the Impact section.

If you have not opted in to the Pilot, you are not required to complete this section, but you may still do so if it would be useful to discuss management of data in the project.

Some guidance on answering the questions above follows.

### What types of data will the project generate/collect?

Clearly identify and characterise each type of data. Describe the data contents, e.g. experimental measurements, models and model outputs, records, images, etc. Quantify where possible, e.g. by number of experiments or data volume.

### What standards will be used?

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, identify these and provide relevant information. For example, the European Nucleotide Archive specifies data formats for submission: <http://www.ebi.ac.uk/ena/submit/data-formats>. Search for relevant standards at <https://fairsharing.org/standards/>. For more information about Metadata Standards, see <http://rd-alliance.github.io/metadata-directory/>.

### Vehicles for data sharing, curation and preservation

The basic EC expectation is that data will be made FAIR, i.e. Findable, Accessible, Interoperable, and Re-usable, and the preference is for you to use data repositories that support open access to data to achieve these aims. Data repositories provide long-term preservation and curation, licence and manage access to data, and publish online metadata records (usually conforming to community standards) with unique identifiers (typically Digital Object Identifiers or DOIs), so that data can be easily discovered, cited and linked to.

Data can be deposited in any suitable repository.

All University researchers are eligible to deposit data in the University of Reading Research Data Archive, information here: <http://www.reading.ac.uk/res-research-data-archive.aspx>. The Archive will preserve and enable access to data in the long-term (10 years minimum). Up to 20 GB of data per project can deposited at no charge. Deposits greater than 20 GB may be subject to a charge and must be agreed in advance. If you intend to deposit more than 20 GB of data in the Archive, contact r.m.darby@reading.ac.uk to discuss.

There may be external services that are more suitable for your data if they serve particular subject communities or manage specific types of data. The EC also provides its own publications and data repository, Zenodo, which offers a guaranteed home for the outputs of EC-funded research. There is no charge to use this service. Zenodo can be accessed at <https://zenodo.org/>.

Guidance on choosing a suitable data repository can be found here: <http://www.reading.ac.uk/res-data-repository.aspx>.

### How will the costs for data curation and preservation be covered?

If you anticipate needing to store in excess of 20 GB of data in the Archive you must contact r.m.darby@reading.ac.uk to discuss costs when you are preparing your application.

There may be costs for archiving data in alternative repositories that you will need to include in your application. Costs for use of external services are eligible for EC grant funding.

### If data cannot be made available, explain why

If you are participating in the Pilot, but some of your data cannot be made available, you should explain this here, clearly identifying the relevant data and reason for their exclusion. For example, you may wish to provide for protection of data that will be subject to commercial exploitation (e.g. patenting or licensing), or clarify that some personal or commercially-privileged data will not be made available, in order to meet relevant ethical and/or legal obligations.

Where full disclosure of data will not be made, you should plan to make data as open as possible, e.g. by obtaining informed consent for data sharing from research participants, and by using methods such as redaction and anonymization to remove sensitive information from datasets prior to publication.

For advice on how to maximise data openness while protecting the interests of research participants, consult the guidance provided by the UK Data Service at <https://www.ukdataservice.ac.uk/manage-data/legal-ethical>.

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

If you have opted in to the Pilot and your grant is awarded, you will be required to deliver a first version of a full Data Management Plan (DMP) for the project within the first six months. A Data Management Plan template in provided is Annex 1 to the **Guidelines on FAIR Data Management in Horizon 2020** (<http://bit.ly/1Y0OMaI>). This template includes a set of 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. It is expected that an updated version will be prepared in the context of periodic project evaluations/assessments, and at the least in time for the final project review.

Where Reading is the project co-ordinator 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.

# Additional considerations

## 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. In particular 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.

Where possible, data collected/held at the University should be stored using the University network or your University OneDrive account, which will provide data security, replication in separate data centres, automated backup and file recovery.

For group access with small to medium-volume storage requirements, you can set up a collaborative share for the project of up to 100 GB at no cost (thereafter £1.20 per GB per year). For information about this service log in to the DTS Service Catalogue at <http://www.reading.ac.uk/internal/its/services/sercat2017.aspx> and select File Storage.

Your University OneDrive account allows you to store and share 5 TB + of data at no cost. Data must be stored and shared in accordance with the OneDrive Data Security Policy. Information about the service is available here: <http://www.reading.ac.uk/internal/its/services/OneDrive.aspx>.

For storage requirements greater than 100 GB you can use the University’s Research Data Storage service. The minimum allocation for research data storage is 500GB for 1 month, and 3 different tiers of storage are available:

* Basic: £106 per TB per year (replication in one location, no snapshots or backup)
* Research Cloud (can be used with computing platforms): £282 per TB per year (replication in one location, 2-week snapshots, cloud backup optional)
* Gold: £434 per TB per year (replication in two locations, snapshots and 3-month cloud backup)

For information about the service log in to the DTS Service Catalogue and select Research Data Storage.

If you need accessible storage to support collaboration with external partners, other cloud-based services (e.g. Dropbox) may be suitable. You should bear in mind risks to data security when using cloud services, especially where personal, confidential or sensitive data are involved. Make sure any service you intend to use provides appropriate security guarantees and adopt additional security measures, such as encryption of files and folders, where necessary.

The Academic Computing Team provides guidance on alternative cloud services at <https://research.reading.ac.uk/act/knowledgebase/off-site-cloud-storage/>.

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. For information is available in the Academic Computing Team portal at <https://research.reading.ac.uk/act/knowledgebase/research-cloud-service/>.

## Ethics and Security

You are not required to address ethical considerations related to research data in the data management section of the proposal, except in so far as they may prevent you from making some data available.

But 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 **How to complete your ethics self-assessment**, available from the bottom of the page at <http://bit.ly/2eoHoXE>.[[3]](#footnote-3)

1. Full URL: <http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm>. [↑](#footnote-ref-1)
2. Full URL: <http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm>. [↑](#footnote-ref-2)
3. Full URL: <http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/ethics_en.htm>. [↑](#footnote-ref-3)