ERA-CAPS

Data Sharing Policy

March 2014

ERA-CAPS Data Sharing Policy

1. Principles

ERA-CAPS' view on research data is informed by the overarching principles declared by the Organisation for Economic Cooperation and Development in its 2007 publication "Principles and Guidelines for Access to Research Data from Public Funding":

- Publicly-funded research data are a public good, produced in the public interest;
- Publicly-funded research data should be openly available to the maximum extent possible.

2. General Requirements

In order to facilitate the re-use, re-integration and repurposing of research data, for the development of new knowledge and research directions and to ensure that maximum impact is realised from ERA-CAPS funding, all research projects applying for an ERA-CAPS grant are required to:

A. deliver
1. a Data Management Plan along with the grant application;
2. a Data Management Plan progress or update report together with the projects' mid-term and final reports;

B. designate, from among the research project staff, a Data Officer who shall
1. oversee
   i) the preparation of the Data Management Plan;
   ii) its execution, compliance and eventual recast, should there be a need for repurposing the plan during the research project's lifetime;
2. act as the project's contact point to account for the implementation of the Data Management Plan and any other related issue before the ERA-CAPS Evaluation Panels and the Call Secretariat.

Provisions for data sharing set out in the Data Management Plans must be compatible and not conflict with the ERA-CAPS Intellectual Property Rights Conditions.

3. Specific and Conditional Requirements

A. The Data Management Plan must always have a reference to the expected research results, including the description of datasets, data types and quantities.

1. If no data is expected to be produced during the research project, the Data Management Plan is still required: it must simply refer the fact that no data is expected to be produced and explain why.

B. If data is expected to be produced during the research project, participants must use the English language for data handling (classification, description) and the associated metadata.

C. If data is expected to be produced in a digital format and to be made accessible online, it must be assigned a persistent identifier, such as a Digital Object Identifier[1].

D. If the Data Management Plan foresees self-archiving of the project's data in an online repository:
   1. the repository of choice must conform to the OpenAIRE Guidelines For Data Archives[2] (see Section 5);
   2. data must always be deposited in a publicly-maintained repository, either institutional or community-based/disciplinary repository, regardless of the fact that it may also be deposited in a scientific publisher's private repository following the publication of a research result.

E. If the applicants expect to publish the research results, the underlying data supporting the publication must be made freely available for sharing and re-use.
   1. this data must be made public as soon as possible, preferably at the same time as the result is published and consistent with the published guidelines relevant to the consortium;

F. If data and/or results are made available online, a hyperlink to the dataset or publication shall be placed in the project's webpage within ERA-CAPS website.

4. Assessment of the Data Management Plan

The Data Management Plan will be reviewed by scientific experts who will assess its adequacy and effectiveness in relation to the best practices and the state-of-the-art of data management and sharing modalities within the research project's specific scientific discipline.

It may then be subject to a revision proposal if deemed inadequate, that applicants must comply with.

If, within the specific research area of the project, there are yet no commonly agreed or widely-accepted standards or best practices, applicants must indicate in the Data Management Plan what are their options for describing, handling, processing, archiving and sharing their data and the reasoning behind such choices, with a view on providing the maximum possible availability and re-usability of the data and on the interoperability of their adopted data standards and formats with those of the wider scientific framework.

---

[1] For more information, please refer to: http://www.doi.org/

[2] For more information, please refer to:
https://guidelines.openaire.eu/wiki/OpenAIRE_Guidelines:_For_Data_Archives

5. Guidelines and Recommendations for Compliance

ERA-CAPS acknowledges that as molecular plant sciences include many different research subareas with specific associated methods, techniques and traditions, there is also a wide variety of material considered as data and corresponding data types and handling methods.

The ERA-CAPS Data Sharing Policy is purposely flexible and not overly prescriptive so as to embrace this complex and multi-faceted reality, further complicated by the fact that participants are operating within an international setting, subject to different legal constraints from country to country.

The required Data Management Plan consists of a simple statement detailing, in accordance with the state-of-the-art of data management in their scientific disciplines and communities, how will applicants manage and make available, to the maximum possible extent, the data they expect to produce during the research project.

It is up to the applicants to identify the data expected to arise from their research and decide what methods, processes, techniques and formats should be applied for the best possible data handling, processing, sharing and storage.

Drafting the Data Management Plan

There are a number of topics that a Data Management Plan should address.

Namely, any appropriate Data Management Plan should include information regarding:

- Data areas, types and quantities
  - the volume, type and content of data that will be generated e.g. experimental measurements, numerical data, models, records and images;
- Standards, metadata and capture methods
  - the standards and methodologies that will be adopted for data collection, management and further dissemination, and why these have been selected. If no established standards are expected to be used, there should be an explanation on how datasets will be described;
- Physical and/or digital resources and infrastructures (including third party resources) to store, preserve and make data available
  - for instance, offline and online archives and repositories;
- Ethics and Intellectual Property issues
  - ownership of data and ethical issues related to the data;
  - declaration of waivers – any partial or total restrictions on data sharing due to the need to protect privacy, confidentiality, security, intellectual property rights and other rights and, if applicable, any envisaged mechanisms to lift such restrictions;
- Secondary use and provisions for data sharing, public access and reuse
  - further intended and/or foreseeable research uses for the completed dataset(s);

- - - o planned mechanisms for making data available, e.g. through self-archiving in public database repositories or upon request to the authors, including access mechanisms where appropriate;
- Timeframes
    - o timescales for public release of data
    - o the duration of time that data is expected to be preserved;
- Responsibility of each party involved in the research project towards the management and custody of research data
- Budget information for the management of data.
- Provisions for managing the data after the end of the project.

Best Practices

Even if the actual content of the Data Management Plans may vary, there are some general best practices that should be observed when elaborating the Data Management Plans, depending on the specific choices made, to aim for the widest possible dissemination and the best possible preservation of data quality.

- Data should always be unequivocally identified by using some coherent and consistent classification system.
- Where standardised formats and metadata exist for the data expected to be produced, such norms should always be applied instead of any proprietary formats.
    - o For those cases where it is unavoidable to use a specific format or a specific software to handle the data, there should be provisions to ensure that:
        - ▪ instructions are clearly given so that the data is retrievable and perfectly understandable;
        - ▪ or that the specific software is also made available, as well as the instructions on how to use it, so that any third party out of the research team can access, interpret and re-use the data.
- Free, public, direct, immediate and online access to data, without any other particular requirements, should be promoted whenever possible.
    - o If data is made available online, it must be assigned a persistent identifier. The use of a Digital Object Identifier is highly recommended.
- Self-archiving of data in an online repository is highly recommended.
    - o While leaving the ultimate decision about the specific repository of choice to the applicants if a project decides to share their data this way, the chosen repository must nevertheless conform to the OpenAIRE Guidelines For Data Archives. This guarantees the repository's own standard of quality, the retrievability and accessibility of deposited data and the repository's interoperability with other digital infrastructures;
    - o Publicly-managed over privately-owned repositories are preferred. This is the best way to guarantee
        - ▪ long-term preservation of data

- the ability to keep on par with upgradable and interoperable quality standards
- and the perpetuity of free and open access to the research data, irrespective of market fluctuations that affect private companies.

- If sharing of the research data is impaired by any type of restriction, participants should ensure that such restrictions are clearly set out and justified in the Data Management Plan as well as any envisaged mechanisms to lift such restrictions.
    - If restrictions on data sharing are specifically related to privacy issues, participants should consider to make provisions to anonymise the data, if feasible, so that it can be safely released without compromising third parties' privacy.
- Data should be available as soon as it is reasonably possible and its preservation and accessibility should be ensured for the greatest relevant amount of time.
- In any case, it should be clear for other potential users of the research data what is allowed for them to do with it. In that sense, if the data is allowed to be freely re-used, any interested party should be made fully aware of this, by assigning the data an adequate licence, such as a CC0 Licence[3].

Relation with Publications of Research Results

Data supporting published research results

As stated above, whenever the results of an ERA-CAPS funded research project are subject to publication, participants must make the data supporting the research results freely available for sharing and re-use.

Such data must be made public as soon as possible, but preferably at the same time as the result is published. The timescale for making data available must always be consistent with the published national or institutional data management guidelines relevant to the consortium.

Publications of research results

Participants are recommended to publish their research results in an Open Access format, either by directly publishing in an Open Access Journal or by depositing the research article in an Open Access public repository.

As stated in the Intellectual Property Rights Conditions:

"*The modalities of open access can be achieved either through a publication in a scientific journal being in open access or with an agreement of the subscription-based journal. In that last case (i) a reasonable time-limit in which only the subscribers of the scientific journal will have access to it may be applied if the open access is accepted free of charge; and (ii) the concerned article will have to be accessible without restriction either directly on the journal's website or on a public archive.*"

---

[3] For more information, please refer to: http://creativecommons.org/about/cc0

Furthermore, it is required for participants to ensure, by engaging with the party managing the ERA-CAPS website, that an hyperlink to the publication or the online datasets is placed in the respective project's webpage within the ERA-CAPS website.

Intellectual Property Rights

ERA-CAPS participants shall ensure that their Data Management Plans are consistent and do not conflict with the Intellectual Property Rights Conditions set out by ERA-CAPS.

The Intellectual Property Rights Conditions declare "(…)*the Participants shall endeavour to strike a balance between open access of raw data for international projects and restrictions on access to these Results*".