European Research Council

Scientific Council

Established by the European Commission

Open Research Data and Data Management Plans

Information for ERC grantees

by the ERC Scientific Council

Version 4.0

11 August 2021

European Commission

This document is regularly updated in order to take into account new developments in this rapidly evolving field. Comments, corrections and suggestions should be sent to the Secretariat of the *ERC Scientific Council Working Group on Open Science* via the address erc-open-access@ec.europa.eu.

The table below summarizes the main changes that this document has undergone.

HISTORY OF CHANGES			
Version	**Publication date**	**Main changes**	**Page (in the relevant version)**
1.0	23.02.2018	• Initial version	
2.0	24.04.2018	• Part 'Open research data and data deposition in the Physical Sciences and Engineering domain' added • Minor editorial changes; faulty link corrected • Contact address added	15-17 6, 10 2
3.0	23.04.2019	• Name of WG updated • Added text to the section on 'Data deposition' • Reference to FAIRsharing moved to the general part from the Life Sciences part and extended • Added example of the Austrian Science Fund in the section on 'Policies of other funding organisations'; updated links related to the German Research Foundation and the Arts and Humanities Research Council; added reference to the Science Europe guide • Small changes to the text on 'Image data' • Added reference to the Ocean Biogeographic Information (OBIS) • Reformulation of the text related to Biostudies • New text in the section on 'Metadata' in the Life Sciences part • Added reference to openICPSR • Added references to ioChem-BD and ChemSpider • Change of header 'Geophysics' into 'Earth system science' • Information on EPOS updated • Minor editorial changes and updates	2 5 7 8 9 10 11 11 13 16/17 17 17 whole document
3.1	03.07.2019	• Added reference to OpenNeuro	14
4.0	11.08.2021	• Integration of the concept of 'data product' • Integration of references to the new requirements under Horizon Europe and related guidance • Added reference to the ARGOS tool • Moved reference to the Guidelines on FAIR Data Management in Horizon 2020 • Added reference to GitHub and Zenodo-GitHub / Dryad-Zenodo integrations • Updated reference to Science Europe Practical Guide • Added a section on 'Where to obtain further help and support' • Added reference to the ELIXIR Research Data	3-5 4-9 6 6, 10 8 10 10-11 12

HISTORY OF CHANGES			
Version	**Publication date**	**Main changes**	**Page (in the relevant version)**
		Management Kit (RDMkit) • Updated subsection on 'Image data' • Added references to MGI and Xenbase • Updated subsection on 'Astronomy' • Added reference to ICOS, ESGF, Pangaea and NCEI Paleoclimatology data • Added references to NOMAD and Materials Cloud Archive • Updated subsection on 'Particle physics' • Change of header 'Software engineering' to 'Computer science' • Removed reference to QUALINET • Editorial changes and clarifications; updates of links	 12-13 13 19 21 21 21 22 22 whole document

Open Research Data and Data Management Plans

Information for ERC grantees

The ERC has supported the cause of open science from its start in 2007, and continues to do so today. Open access to publications from ERC funded projects is already mandatory; the next step in the development of open science is making research data also publicly available when possible. This will benefit science by increasing the use of data and by promoting transparency and accountability.

The ERC embraces the FAIR data principles: research data should be **f**indable, **a**ccessible, **i**nteroperable and **r**e-usable. This means that data should be:

- identified in a persistent manner using community conventions, and described using sufficiently rich metadata;
- stored in such a way that they can be accessed by humans and machines;
- structured in such a way that they can be combined with other datasets;
- licensed or having terms-of-use that spell out how they can be used by others.

The article by Wilkinson et al. on "The FAIR Guiding Principles for scientific data management and stewardship"[1] provides a detailed discussion of the FAIR principles.

Not all data can or should be preserved in the long term. In some cases, the sheer size of raw data may mean that only derived data products[2] can be archived. In such cases, the corresponding metadata should remain FAIR and reference the decision not to retain the data. The criteria for prioritisation, appraisal and selection of the data to be retained should be detailed in the Data Management Plan.

Likewise, not all data can be made fully open. Where data raise privacy or security concerns, controls and limits on data access will be required. In some cases, it will be appropriate for researchers to delay or limit access to data in order to secure intellectual property protection.[3] There may also be other reasons to keep data closed. Any restrictions on access should be explicit and justified in the Data Management Plan, and such data should still be managed in line with the FAIR principles.

For researchers, the move to FAIR data means that they have to think about what data their research will produce, how these data will be described, and how they can be made available in such a way as to benefit science and society in general. This means that they have to draw up a Data Management Plan and find suitable data repositories.

---

[1] 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)

[2] Here and in the sequel, the term 'data products' is used to mean derived data that satisfy certain standards that depend on the specific discipline or research field. For an example from astronomy, see the "ESO Science Data Products Standard" (https://www.eso.org/sci/observing/phase3/p3sdpstd.pdf).

[3] In this context the following report by the European Commission may be of interest: Crouzier, T., Barbarossa, E., Grande, S., Triaille, J.P., *IPR, Technology Transfer & Open Science,* Publications Office of the European Union, Luxembourg, 2017 (https://doi.org/10.2760/789864)

ERC requirements

All ERC projects funded under the Work programmes 2017 to 2020 participate by default in the Horizon 2020 Open Research Data (ORD) pilot, with the possibility for grantees to opt out at any time[4]. For projects funded under the Work programmes 2015 and 2016 grantees can opt into the pilot if they so wish.

ERC grantees of projects that take part in the Horizon 2020 ORD pilot are required to submit a Data Management Plan (DMP) within six months after the start of their grant. Grantees are required to deposit their research data in a repository and provide open access at least to those data, including associated metadata, needed to validate the results in their publications. Access to other data, including associated metadata, has to be provided as specified in the DMP.[5]

Under Horizon Europe (Work programmes 2021 and onwards), grantees of all ERC projects that generate research data have to submit a DMP (at the latest six months after the start of the project), deposit such data in a 'trusted' repository and provide access to them, under the principle "as open as possible, as closed as necessary". There are also a number of requirements concerning the bibliographic and administrative metadata of deposited data, which also have to be made openly accessible to enhance findability and facilitate reuse.

Under Horizon Europe it is not possible to opt out completely from these obligations, but exceptions to the requirement to provide open access to data and metadata are possible. Grantees funded under Horizon Europe are advised to pay careful attention to the requirements detailed in the Horizon Europe Model Grant Agreement (MGA)[6] and the explanations provided in the Horizon Europe Annotated Grant Agreement (AGA)[7].

Data Management Plans

As practices with regard to data management, storage, and sharing differ widely across disciplines, the ERC uses a general set of requirements that DMPs should meet.

A DMP should provide information on:

1. *Dataset description:*
   Grantees should provide a sufficiently detailed description, including the scientific focus and technical approach, to allow association of their datasets and derived data products with specific research themes.
2. *Standards and metadata:*
   Grantees should describe the protocols used to structure their data and indicate the metadata standards applied. This will allow other scientists to make an assessment, to

[4] In case of opt-out after the signature of the grant agreement, a formal amendment must be requested.
[5] See Article 29.3 of the Horizon 2020 ERC Model Grant Agreement (https://ec.europa.eu/research/participants/data/ref/h2020/mga/erc/h2020-mga-erc-multi_en.pdf) and the (ERC specific) annotations in the Horizon 2020 Annotated Grant Agreement (https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/amga/h2020-amga_en.pdf).
[6] See Annex 5 (Article 17) of the Horizon Europe Model Grant Agreement (https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/agr-contr/general-mga_horizon-euratom_en.pdf).
[7] See annotations to Annex 5 (Article 17) in the Horizon Europe Annotated Grant Agreement (https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/aga_en.pdf).

attempt to reproduce the conclusions derived from the dataset (and possibly even the dataset itself), and potentially reuse the data for further research. If available, grantees should provide a reference to the community data standards with which their data conform and that make them interoperable with other datasets of similar type.

3. *Name and persistent identifier for the datasets:*
   Grantees should plan to use repositories that will provide a unique and persistent identification (an identifier) of their datasets and derived data products, and a stable resolvable link to where they (or, as a minimum, their metadata) can be directly accessed.

4. *Curation and preservation methodology:*
   Grantees should provide information on the standards that will be used to ensure the integrity of their datasets, and the period during which they will be maintained. Grantees should also explain whether and how their datasets will be preserved and kept accessible in the longer term. If applicable, they should detail the criteria for prioritisation, appraisal and selection of the datasets to be retained. If raw data cannot be stored (e.g. because they are too large or modified in (quasi-)real-time), grantees should describe what data products will be derived, and how these will be preserved and kept accessible. If available, grantees should provide a reference to the public data repository in which their datasets or data products will reside.

5. *Data sharing methodology*
   Grantees should provide information on how their datasets and/or data products can be accessed, including the terms-of-use or the licence under which they can be accessed and re-used, and information on any restrictions that may apply. It is also important to specify and justify the timing of data sharing. This could be, for example, as soon as possible after the data collection, or at the end of the project. For data that underlie publications it could be, for example, at the time of publication or pre-publication.

Grantees should demonstrate that their approach to data management planning is in line with the FAIR principles by providing adequate information on these five topics.

The ERC does not prescribe a specific format for the DMPs that its grantees need to submit, because practices and standards differ widely across disciplines. However, grantees funded under Horizon 2020 are encouraged to use the ERC template[8] that is available on the European Commission's Funding & Tenders Portal:

- ERC Data Management Plan Template:
  http://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-erc-tpl-oa-data-mgt-plan_en.odt

A very convenient on-line tool to formulate a DMP according to the requirements of the ERC (as laid down in the template) and of several other research funding organisations is provided by the Digital Curation Centre:

- DMPonline tool: https://dmponline.dcc.ac.uk

[8] A DMP template for ERC projects funded under Horizon Europe is in preparation.

The ARGOS tool (a joint effort of OpenAIRE and EUDAT) allows generating machine actionable DMPs:

- ARGOS tool: https://argos.openaire.eu/

Grantees funded under Horizon 2020 should also keep in mind the following guidance document[9]:

- Guidelines on Implementation of Open Access to Scientific Publications and Research Data in projects supported by the European Research Council under Horizon 2020: https://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf

The following document by the European Commission is also instructive:

- Guidelines on FAIR Data Management in Horizon 2020: https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

Writing a DMP should not be regarded as a purely administrative exercise. Rather, it should provide a positive stimulus to thinking about how the data generated within a project will be stored, managed and safeguarded, and possibly shared to be reused. It should be part of the research process from the outset. As a project progresses, the data generated may well change in type and volume. It is therefore useful to envisage a DMP as a dynamic framework, which should be maintained and modified as the research advances. Planning for submission early in the research cycle will facilitate the publication process. Good data management will save time, safeguard information and increase the visibility and impact of the research outcomes.

The ERC recognises that data annotation and deposition are time-consuming activities. ERC grant money can be specifically earmarked for this purpose, for example to contribute to the salary of a research assistant or to the costs of a commercial provider.

Data deposition

The ERC is convinced of the importance of data and their value to the scientific community. Data deposition can be complementary to publication, but data can also be deposited without an associated publication. The ERC considers data as an important scientific output; therefore data deposition should always be accompanied by a reference to the ERC grant number in the metadata.

Publications present the pertinent data underlying conclusions made in a research paper, and publishers increasingly require that all relevant data are made available to the community. The ERC expects data underlying publications by ERC grantees to adhere to the FAIR principles. Researchers often generate additional data, not directly linked to publications, which shape the way their projects develop, and these also can constitute a valuable resource. Funders and indeed the public in general are anxious that all valid data be managed in line with the FAIR principles in order to promote scientific progress; the European Commission has adopted a policy of open data for all research that it finances.

[9] Specific guidance for grantees funded under Horizon Europe is in preparation.

Data dumping is of course to be avoided, especially where datasets are huge. It is important that data be of sufficient technical and scientific quality as well as being sufficiently annotated and structured to be useful to the community. Ultimately, it is for the individual investigator to decide which data merit conservation and/or sharing. Where the scientific content is concerned, it is necessary to bear in mind that what seems of little interest in the context of a particular project may be relevant to other lines of investigation and therefore of potential interest to the research community. So-called negative results may also be of potential value.

When looking for a repository for research data, grantees should first check whether there is a thematic/community database where the data could be archived. Irrespective of the repository chosen, grantees should always check whether it is sustainable in the longer term and:

- stores the data in a safe way;
- makes sure that the data will remain findable (via the use of a persistent identifier), as well as accessible and re-usable;
- describes the data in a standard way, using accepted metadata standards;
- allows the depositor to specify a licence governing access and re-usability of the data.

Grantees funded under Horizon Europe should check that the repository they have chosen satisfies the requirements for a 'trusted' repository and allows them to encode the detailed metadata required by the Horizon Europe Model Grant Agreement[10]. One way for a repository to demonstrate that it is 'trusted' is to be certified (see below), but there are other criteria that are also sufficient. Grantees may also want to check whether the repository is an OpenAIRE content provider[11], in which case at least some of the metadata encoded in the repository may be ingested directly into the Horizon Europe grant reporting system.

There are a number of organisations that carry out a certification of data repositories. The following links may be useful:

- Core Trust Seal (this list includes repositories certified by the Data Seal of Approval and/or the World Data System): https://www.coretrustseal.org/why-certification/certified-repositories/
- Nestor seal (DIN-Norm 31644): https://www.langzeitarchivierung.de/Webs/nestor/EN/Services/nestor_Siegel/nestor_siegel_node.html

Since 2012, there is also an ISO standard for trusted digital repositories (ISO 16363). However, the uptake of the related certification has been minimal so far.[12]

[10] See Annex 5 (Article 17) of the Horizon Europe Model Grant Agreement (https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/agr-contr/general-mga_horizon-euratom_en.pdf) and the annotations to Annex 5 (Article 17) in the Horizon Europe Annotated Grant Agreement (https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/common/guidance/aga_en.pdf).
[11] See https://explore.openaire.eu/search/find?active=datasources.
[12] See http://www.iso16363.org/iso-certification/certified-clients/.

General repositories for research data

The following repositories are of interest to researchers in all domains:

- Zenodo (not-for-profit, hosted by CERN): https://zenodo.org
- Dryad (not-for-profit membership organisation): https://datadryad.org/

A popular repository for software, code, algorithms etc. is

- GitHub (private company, subsidiary of Microsoft): https://github.com/

Zenodo can be used in combination with GitHub to provide a DOI to software.[13] Recently a Dryad–Zenodo integration has been established that allows researchers to publish software and code in Zenodo through the Dryad submission process.[14]

Also popular are:

- Figshare (free service provided by a private company): https://figshare.com
- Open Science Framework (not-for-profit, developed and maintained by the Center for Open Science[15]): https://osf.io
- Harvard Dataverse (not-for-profit, hosted by the Institute for Quantitative Social Studies IQSS at Harvard University): https://dataverse.harvard.edu

While some of these repositories, such as Zenodo, are supported by public money, some others, such as Dryad, may charge a fee. Some degree of data curation may be provided, but this is often not the case. Figshare is a commercial company that provides data management services to individuals and will advise about data curation and data deposition through a cloud provider. The company also works with institutions to enable them to curate their academic research outputs and host their data on their own machines.

For an extensive overview of data repositories across all disciplines, see:

- Registry of Research Data Repositories (re3data.org): https://www.re3data.org

At the European level, EUDAT bundles a large number of general and discipline-specific repositories:

- EUDAT Collaborative Data Infrastructure (CDI): https://eudat.eu/eudat-cdi

A growing number of universities and research institutes host a repository for use by their research staff. Most of these institutional repositories were originally set up for storing (open access) publications, but dedicated research data repositories also occur. In order for an institutional repository to be acceptable as a research data archive, it is essential that the university/institute has a data policy guaranteeing the support for data storage and sharing into the future. Grantees supported by Horizon Europe should also ensure that the repository fulfils all the other requirements for a 'trusted' repository.

Individual researchers may also set up their own focussed database. There are many such initiatives, which may be open to the community and can play a useful role. However, in

[13] See https://guides.github.com/activities/citable-code/.
[14] See https://blog.datadryad.org/2021/02/08/doing-it-right-a-better-approach-for-software-amp-data/ for more details.
[15] https://cos.io

contrast to public data repositories, these are generally not deposition databases, and as long as they depend on a single individual and/or funding source, long-term sustainability is challenging. In addition to the major problem of perennity, curation of the data may not always be adequate, with problems of quality, correct annotation, renewal (whether the database is up to date) etc.. This can complicate access and also compromises re-use. For all these reasons such initiatives are generally not suitable for the long-term archiving of research data generated by ERC projects in line with the FAIR principles.

Many journal websites contain lists of repositories. In addition, there are an increasing number of commercial publishers that offer authors opportunities to store the research data underlying their publications. Grantees should be aware that these solutions are unlikely to be in line with the FAIR principles and the requirements of the ERC grant agreement.

If in doubt about how to deposit data, in what format etc., it is recommended to consult the repository directly.

Metadata and data preparation

In order to make stored data findable, accessible, interoperable and reusable (FAIR), it is not enough to store raw data; they need to be properly documented and described using informative metadata.

Defining appropriate metadata depends on the discipline and/or the methodology that was used to produce the data. Discipline-specific repositories often have detailed requirements for describing data that are stored in that repository.

A generally accepted minimum standard for describing information on the web, including research data, is Dublin Core. Further information on this metadata standard is available at:

- Dublin core: https://dublincore.org/

For more information on disciplinary metadata standards see also

- Digital Curation Centre: https://www.dcc.ac.uk/guidance/standards/metadata

and the Metadata Directory that has been set up under the auspices of the Research Data Alliance:

- RDA Metadata Directory: https://rd-alliance.github.io/metadata-directory/

A curated resource on data and metadata standards, inter-related to databases and data policies can be found at

- FAIRsharing: https://fairsharing.org/.

From its first incarnation as BioSharing.org – which focused on the life sciences – FAIRsharing has evolved into a resource that serves users across all disciplines.[16]

As indicated earlier, grantees funded under Horizon Europe also need to ensure that their chosen repository allows the submission of the bibliographic and administrative metadata as detailed in the Horizon Europe MGA.

[16] Sansone, S.-A. et al. (2019). FAIRsharing as a community approach to standards, repositories and policies. *Nature Biotechnology,* volume 37, pages 358-367 (https://doi.org/10.1038/s41587-019-0080-8)

Policies of other funding organisations

As the movement towards FAIR research data management progresses, various national funding agencies have formulated policies and specified requirements for DMPs that might be informative when drawing up a DMP, for example:

- Austrian Science Fund (FWF): "Research Data Management" https://www.fwf.ac.at/en/research-funding/open-access-policy/research-data-management/
- Netherlands Organisation for Scientific Research (NWO): "Research data management" https://www.nwo.nl/en/research-data-management
- German Research Foundation (DFG): "DFG Guidelines on the Handling of Research Data" https://www.dfg.de/en/research_funding/proposal_review_decision/applicants/research_data/index.html
- Swiss National Science Foundation (SNSF): "Open Research Data" http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx
- The Research Council of Norway (RCN): "Open Access to Research Data": https://www.forskningsradet.no/en/Adviser-research-policy/open-science/open-access-to-research-data/
- UK Arts and Humanities Research Council (AHRC): "Data Management Plan": https://ahrc.ukri.org/documents/guides/research-funding-guide1/ (Funding Guide: "4. Application Guidance of the Funding Guide"; see also "5. Assessment Criteria and Peer Review")

In January 2021 Science Europe published its

- "Practical Guide to the International Alignment of Research Data Management – Extended Edition" https://www.scienceeurope.org/our-resources/practical-guide-to-the-international-alignment-of-research-data-management/.

The original version of this guide was released in early 2019. Developed by experts from Science Europe member organisations and in consultation with the broader research stakeholder community, the guide presents core requirements for DMPs and criteria for the selection of trustworthy repositories, as well as some guidance to organisations on how to put these into practice. Following the successful uptake of the guide by several organisations, the extended edition also features a new rubric to facilitate the evaluation of a DMP.

Where to obtain further help and support

Useful information on how to better align research data and software with the FAIR principles can be found in the collection

- "Top 10 FAIR Data & Software Things": https://librarycarpentry.org/Top-10-FAIR/.

These are brief guides (stand-alone, self-paced training materials) for different disciplines/topics that can be used by the research community to understand how they can make their research data and software more FAIR.

The Open Science infrastructure

- OpenAIRE: https://www.openaire.eu/

provides support[17] on many different aspects of research data management and also runs a helpdesk[18]. OpenAIRE also coordinates a network of National Open Access Desks (NOADs)[19] across Europe who can provide assistance on general topics related to FAIR research data.

Grantees who require help with specific questions can contact the relevant support structures within their institution or their reference Research Infrastructure. In some countries, there are national data centres or competence centres that can provide tailor-made assistance. For questions related to the Open Science requirements in the ERC grant agreement, grantees are invited to contact the ERC Executive Agency at erc-open-access@ec.europa.eu.

~~~~~~~~~~~~~~~~~

In what follows, more specific information is given for ERC grantees in the Life Sciences and in the Physical Sciences and Engineering, and for those working in the Social Sciences and Humanities. This includes references to specialised repositories for specific disciplines as well as more general domain-specific information.

Note that this information is provided 'as is', i.e. it does not reflect any particular preference on part of the ERC as to which repositories, protocols, metadata or sharing methodologies an ERC grantee chooses to use.

[17] See the article "New support resources on Open Science practices and OpenAIRE services" https://www.openaire.eu/new-support-resources-on-open-science-practices-and-openaire-services
[18] https://www.openaire.eu/support/helpdesk
[19] https://www.openaire.eu/contact-noads
~~~~~~~~~~~~~~~~~

Open research data and data deposition in the Life Sciences domain

The life sciences have a long tradition of open access data repositories. Submission of datasets to an established public repository is considered good scientific practice and is often also a condition for publication. The public repositories ensure that data are correctly curated, accessible and maintained in the long term. Data publication through such a repository will help grantees make their data FAIR. In addition, some publishers are implementing formal data citation in the reference list of papers, which will provide a mechanism to attribute credit to datasets. In this context see the paper "A Data Citation Roadmap for Scientific Publishers" by Cousijn et al.[20].

A useful resource for grantees in the life sciences is the ELIXIR Research Data Management Kit (RDMkit), which can be accessed at https://rdmkit.elixir-europe.org/. It provides examples of good research data management practices and offers guidelines, information, and pointers to help researchers with problems throughout the data lifecycle.

Established public repositories

ELIXIR, the ESFRI research infrastructure for life sciences data, has compiled a list of recommended repositories:

- ELIXIR Deposition Databases for Biomolecular Data: https://elixir-europe.org/platforms/data/elixir-deposition-databases

Many of these are based at the EMBL-EBI (European Bioinformatics Institute; for advice on data deposition see https://www.ebi.ac.uk/submission/) with established partner databases in other parts of the world. The

- NCBI resource site: https://www.ncbi.nlm.nih.gov/guide/sitemap/

also provides a list of data repositories, although many do not take public submissions.

Image data

In the rapidly developing area of microscopy and bioimage data, solutions for public archiving and re-use of image datasets are currently being built.

The new European research infrastructure for biological and biomedical imaging

- Euro-BioImaging: https://www.eurobioimaging.eu/

covers a wide range of imaging approaches and is very active in coordinating community solutions for managing and analysing image data. Euro-BioImaging has been granted the status of an ERIC[21]; it currently has 15 members and one observer.

In close collaboration with Euro-BioImaging and ELIXIR, EMBL has launched the new

- Bioimage Archive: https://www.ebi.ac.uk/bioimage-archive/,

[20] Cousijn, H. et al. (2018). A Data Citation Roadmap for Scientific Publishers. *Scientific Data* 5, 180259 (https://doi.org/10.1038/sdata.2018.259)

[21] European Research Infrastructure Consortium (https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures/eric_en)

which accepts image datasets linked to publications, as well as reference image datasets. The Bioimage Archive underpins and works closely with existing more specialist image data resources that support deposition and annotation of different types of image data. These include the

- Electron Microscopy Public Image Archive (EMPIAR): https://www.ebi.ac.uk/pdbe/emdb/empiar/

for electron microscopy as well as correlative light and electron microscopy data, and the

- Image Data Repository (IDR): https://idr.openmicroscopy.org/about/

for cell and tissue scale light microscopy data.

Health sciences and clinical data

Many community databases exist in this area. Different 'clinical speciality' related databases are available, such as:

- National Database for Autism Research (NDAR): https://ndar.nih.gov/

Clinical research outputs tend to be handled nationally because of varying national regulations about confidentiality, where data from individuals are concerned. Personal data poses additional 'consent' challenges, and the development of public databases requires 'controlled access' for data protection. This is a rapidly evolving area where community standards and repositories will be established in the coming years. As standards emerge, the ERC will adopt best practice as recommended by each research community. However, for information, all clinical trials should normally be registered at the outset, in one of the publicly accessible registries identified by the World Health Organisation:

- International Clinical Trials Registry Platform (ICTRP): https://www.who.int/ictrp/en/

Other types of repositories

In a number of research areas, the research community has generated specific archives. These may be repositories that aggregate data from multiple underlying repositories, so that they can be easily found and used by the community. This is the case for organism-based research with examples such as:

- FlyBase – A Database of Drosophila Genes & Genomes: https://flybase.org/
- WormBase: https://wormbase.org/
- The Zebrafish Information Network (ZFIN): https://zfin.org/
- The Mouse Genome Informatics (MGI) data base: http://www.informatics.jax.org/
- Xenbase - The Xenopus model organism knowledgebase: http://www.xenbase.org/entry/

National and international research consortia may also create databases. This is exemplified by a number of databases in the domain of biodiversity, such as:

- Global Biodiversity Information Facility (GBIF): https://www.gbif.org/
- Ocean Biogeographic Information (OBIS): https://obis.org/

Incorporating data into these resources can be very valuable for promoting research within the community, but additional deposition of the data into an established public data-type-focussed repository is highly recommended to ensure long-term curation, preservation and findability.

Data management in domains where established databases are not available

Many institutions have data storage facilities for unstructured data for which there is no existing dedicated community repository. This category includes data generated by functional studies where, for example, a cell component is removed and then complemented by another molecule, or where behavioural studies are carried out to test brain function in an animal model. Unstructured data are accepted by repositories such as Dryad, Zenodo or Figshare, as mentioned in the general part of this document.

In the case that the data behind a study are archived in multiple resources or locations, the ERC encourages grantees to deposit the study metadata, including links to the data location(s), in a recognised resource such as BioStudies (https://www.ebi.ac.uk/biostudies/). This also allows life sciences data for which there is no thematic repository to be deposited. Often a BioStudies record corresponds to the data behind a paper and so can be used to provide a simple link from the paper to the data behind the study via the accession number.

Metadata

In the life sciences, the key community deposition databases have strict metadata standards that are required for deposition of data to make them FAIR. Therefore, much of the thinking of what metadata should be supplied is provided and managed in this way.

Activities surrounding standardisation of metadata (such as cross-data resource identifier mapping, mapping of textual metadata labels to ontology and standard vocabulary terms, standardisation of computational workflows and application programming interfaces (APIs), and schematic mark-up of the data) can be facilitated by reusing existing mature interoperability resources. The section on 'Interoperability' on the ELIXIR website (https://elixir-europe.org/platforms/interoperability) recommends interoperability tools for the purpose of making the data FAIR via the following resources:

- The ELIXIR Recommended Interoperability Resources (RIRs): https://www.elixir-europe.org/platforms/interoperability/rirs
- Bioschemas: https://bioschemas.org/
- Common Workflow Language: https://www.commonwl.org/

Additionally, the 'Tools' section on the ELIXIR website (https://elixir-europe.org/platforms/tools) provides links and guidance on good practice for open source software development in the life sciences.[22]

[22] Jiménez, R.C., Kuzak, M., Alhamdoosh, M. et al. (2017). Four simple recommendations to encourage best practices in research software [version 1; referees: 3 approved]. *F1000Research* 6:876 (https://doi.org/10.12688/f1000research.11407.1)

Open research data and data deposition in the Social Sciences and Humanities domain

The situation with regard to open data in the SH domain, both in terms of infrastructure (repositories) as well as protocols and standards, is rapidly evolving. There are many initiatives, at the national and supra-national levels, that aim to provide researchers with the necessary tools and information.

Characteristic features of the disciplines that together make up the ERC's SH domain is their variety, in terms of topics, epistemologies, and methodologies. This is reflected also in the data that SH projects produce: quantitative datasets; experimental data; observational data; interviews; archival data; human artefacts; medical and genetic data; and so forth. In addition, the various kinds of data cross-cut the disciplinary divisions, as several disciplines produce different kinds of data, depending on the methodologies used.

Also, particular restrictions may apply to making data open depending on the discipline. Data may include copyrighted material, such as literary texts or images, or archival materials to which access is restricted. In other cases, data may include privacy-sensitive material, such as video recordings of parent-child interactions or interviews.

For this reason, it is not possible to provide a single set of guidelines for the entire SH domain. Therefore, this document aims to provide some general and some discipline-specific references that ERC grantees can use to draw up DMPs that are adequate for their discipline and their specific project, and that meet the FAIR principles.

In what follows more information is given on:

- general repositories
- discipline-specific repositories
- metadata and data preparation

General repositories

There are many options available for SH scholars, both general as well as discipline-specific, not-for-profit as well as commercial. The list below mentions a number of well-known repositories for use by social sciences and humanities disciplines, but it is certainly not exhaustive.

An important selection of repositories for SH scholars is provided by CESSDA:

- Consortium of European Social Science Data Archives (CESSDA): https://cessda.net/

CESSDA is an ERIC[23] with currently 22 members, all of them national agencies that operate on a not-for-profit basis.

Many of the CESSDA repositories also cover (some of) the humanities in addition to the social sciences.

[23] European Research Infrastructure Consortium (https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures/eric_en)

The geographical coverage of CESSDA is growing. Among the EU countries missing at the time of writing are some Southern European Countries (Italy, Spain) and most EU-13 countries.

Also of interest to researchers in the SH domain is ICPSR:

- Inter-university Consortium for Political and Social Research (ICPSR): https://www.icpsr.umich.edu/

ICPSR is a not-for-profit membership organisation that maintains a data archive in the social and behavioural sciences:

- openICPSR: https://www.openicpsr.org/openicpsr/

Currently, ICPSR has a membership of more than 750 universities, government agencies, and other institutions.

Discipline-specific repositories

There are a number of repositories that are discipline-specific, and that are usually maintained by discipline-specific organisations or consortia.

Linguistics

- Linguistics Linked Open Data (LLOD): https://linguistic-lod.org/

LLOD is maintained by the Open Knowledge Foundation's Working Group on Open Data in Linguistics (https://linguistics.okfn.org).

- European Research Infrastructure for Language Resources and Technology (CLARIN): https://www.clarin.eu
  Depositing Services offered by CLARIN Centres: https://www.clarin.eu/content/depositing-services

CLARIN is an ERIC, like CESSDA. Its geographical coverage is wide, with currently 21 national consortia as full members and three consortia as observers. Among the EU countries, Spain, Ireland, Luxembourg and several EU-13 countries are currently not (yet) represented among the CLARIN membership.

Historical sciences

Repositories for the historical sciences are mostly at the institutional or national level. A number of CESSDA archives also accept historical datasets.

Archaeology

There are only few repositories dedicated to archaeology. Most of these have a national focus, such as:

- Archaeological Data Service (ADS) in the UK: https://archaeologydataservice.ac.uk/
- e-Depot for Dutch Archaeology (EDNA): https://dans.knaw.nl/en/about/services/easy/e-depot-for-dutch-archaeology

EDNA was established by the Data Archiving and Networked Services (DANS) and the Cultural Heritage Agency (RCE) to archive digital research data of Dutch archaeologists in a sustainable manner and make them available. The data are stored in EASY (https://easy.dans.knaw.nl/), the online archiving system of DANS.

Arts and humanities

- Digital Research Infrastructure for the Arts and Humanities (DARIAH): https://www.dariah.eu/

DARIAH is another ERIC. It is a pan-European infrastructure for arts and humanities scholars working with computational methods. It has 19 members, one observer and several cooperating partners in seven non-member countries. Among the EU countries, missing at the time of writing are Spain and a number of EU-13 countries.

Note that several CESSDA archives also accept humanities datasets.

Psychology

The Leibniz Institute for Psychology Information (https://leibniz-psychology.org/) has developed a data-sharing platform specialized for psychology research:

- PsychData: https://www.psychdata.de/

For an extensive overview of data repositories in psychology, see the article "Finding a Home for Your Science" by DeSoto.[24]

Of interest for researchers working in the psychology subdomain of cognitive neuroscience is the platform

- OpenNeuro: https://openneuro.org/,

which allows the sharing of MRI, MEG, EEG, iEEG, and ECoG data.

Demography

- Data Sharing for Demographic Research (DSDR): https://www.icpsr.umich.edu/icpsrweb/DSDR/

DSDR is housed within the Inter-university Consortium for Political and Social Research (ICPSR) mentioned earlier.

CESSDA archives will normally also accept demographic datasets.

Metadata and data preparation

A general overview of SH metadata standards can be found on the SH-specific pages of the DCC:

[24] DeSoto, K.A. (2016). Finding a Home for Your Science. *Observer*, Volume 29, Issue 5 (https://www.psychologicalscience.org/observer/finding-a-home-for-your-science)

- Digital Curation Centre (DCC): https://www.dcc.ac.uk/resources/subject-areas/social-science-humanities

The DCC website lists metadata standards for, among others, archaeology, social and policy studies, economics, heritage studies.

For metadata and data preparation in the social sciences, see the following guide on the website of the Inter-university Consortium for Political and Social Research (ICPSR):

- Guide to Social Science Data Preparation and Archiving: https://www.icpsr.umich.edu/icpsrweb/content/deposit/guide/

For metadata and data preparation in linguistics, see:

- Section on 'Standards and Formats', CLARIN website: https://www.clarin.eu/content/standards-and-formats

Open research data and data deposition in the Physical Sciences and Engineering domain

The PE domain has a large number of data repositories. In the following section a number of areas are addressed in some detail. This list should by no means be considered as an exhaustive one, rather as a collection of representative examples in a rapidly evolving landscape.

Discipline-specific repositories

Astronomy

The Strasbourg astronomical Data Center is dedicated to the collection and worldwide distribution of astronomical data and related information:

- Strasbourg astronomical Data Center (CDS): https://cdsweb.u-strasbg.fr/

It hosts a variety of repositories of multi-wavelength data and provides useful interfaces, e.g. the SIMBAD astronomical database (https://simbad.u-strasbg.fr/simbad/), the world reference database for the identification of astronomical objects; VizieR (https://vizier.u-strasbg.fr/viz-bin/VizieR), the catalogue service for the CDS reference collection of astronomical catalogues and tables published in academic journals; and the Aladin interactive software sky atlas for access, visualization and analysis of astronomical images, surveys, catalogues, databases and related data (https://aladin.u-strasbg.fr/aladin.gml). Astronomers can upload their own data to CDS, provided that they are related to a publication in a refereed journal. See the following link for details: http://cdsarc.u-strasbg.fr/submit/Make_your_data_visible.pdf

Chemistry

The use of public repositories and databases in chemistry is still developing, with the majority of the progress happening in the area of structural chemistry. The

- Worldwide Protein Data Bank: https://www.wwpdb.org/

manages the archives of the Protein Data Bank, which provides a repository of information about the 3D structures of proteins, nucleic acids, and complex assemblies.

Another key resource in use in this area is the

- Cambridge Crystallographic Data Centre: https://www.ccdc.cam.ac.uk/

for small molecule crystallography data.

- UniProt: https://www.uniprot.org/

covers direct sequencing data for proteins, and both

- ProteomeXchange: http://www.proteomexchange.org/

and the

- PRIDE Archive – proteomics data repository: https://www.ebi.ac.uk/pride/archive/

deal with mass spectrometry proteomics data.

A network of repositories for open access Computational Chemistry research results is

- ioChem-BD: https://www.iochem-bd.org.

A free chemical structure database providing fast text and structure search access to over 67 million structures from hundreds of data sources is

- ChemSpider: https://www.chemspider.com/.

Maintained by the Royal Society of Chemistry, it also encourages researchers to upload their own data.

Earth system science

Digital seismic waveform data in standardized format are available via the International Federation of Digital Seismograph Networks (FDSN, formed in 1985), which provides a huge amount of accessible data via the various on-line data centres, all accessible via the FDSN website:

- Federation of Digital Seismograph Networks (FDSN) : https://www.fdsn.org/webservices/datacenters/

The Data Management Center of IRIS – Incorporated Research Institutions for Seismology (https://www.iris.edu/) in the US is one of the hubs for seismology that serves the international FDSN community, also archiving historical data from pre-digital sources:

- IRIS DMC: https://ds.iris.edu/ds/nodes/dmc/data/types/

Likewise,

- UNAVCO: https://www.unavco.org/

archives and distributes geodetic data (GPS/GNSS, InSAR) for research purposes.

Geochemists also have on-line databases, for example a relational database of peer-reviewed summary data on the geochemistry of all reservoirs in the earth (https://earthref.org/GERM/). Data from geomagnetic observatories around the world can be obtained through the 'Intermagnet' program (https://www.intermagnet.org/). The

- European Plate Observing System (EPOS): https://www.epos-ip.org/

is a collaborative framework where many diverse communities of geoscientists and engineers aim at providing open access to geophysical, geochemical and geological data pertaining to the solid earth as well as visualization and modelling tools. At present, EPOS includes ~300 research institutions from 25 European countries. In October 2018, the European Commission granted EPOS the legal status of an ERIC[25], which is currently joined by 13 countries: Belgium, Denmark, France, Greece, Iceland, Italy, the Netherlands, Norway, Poland, Portugal, Slovenia and the United Kingdom, with Switzerland participating as an observer.

[25] European Research Infrastructure Consortium (https://ec.europa.eu/info/research-and-innovation/strategy/strategy-2020-2024/our-digital-future/european-research-infrastructures/eric_en)

The Integrated Carbon Observation System (ICOS), likewise, is an ERIC whose goal is to provide essential long-term observations required to understand the present state and predict future behaviour of the global carbon cycle greenhouse gas emissions. It provides free access to all ICOS data through the

- ICOS Carbon Portal (https://www.icos-cp.eu/)

as well as to links with inventory data.

In the area of earth system modelling, global model simulations under the Coupled Model Intercomparison Project (CMIP) and the Paleoclimate Modelling Intercomparison Project (PMIP) are archived and can be deposited at the

- Earth System Grid Federation (ESGF): https://esgf-node.llnl.gov/projects/esgf-llnl/.

The following two archives for palaeoclimate and paleoenvironmental data are part of the global data centres for environmental science:

- Pangaea: https://www.pangaea.de/,

which also receives and archives general environmental data, and the

- NCEI (formerly NCDC) Paleoclimatology data: https://www.ncdc.noaa.gov/data-access/paleoclimatology-data.

Materials science

The Crystallography Open Database contains the crystalline structures of a large number of systems. Researchers can contribute with their own results:

- Crystallography Open Database (COD): https://www.crystallography.net/cod/

RefractiveIndex.INFO (https://refractiveindex.info) contains the dielectric functions of various materials. The

- NOMAD repository and archive: https://nomad-lab.eu/index.php?page=repo-arch

contains ab initio electronic-structure data from density-functional theory and methods beyond. It makes scientific data citable and keeps scientific data for at least 10 years for free. NOMAD also facilitates research groups to share and exchange their results. The

- Materials Cloud Archive: https://archive.materialscloud.org/

provides FAIR & long-term storage of research data from computational materials science, with particular focus on sharing the full provenance of calculations.

Particle physics

Scattering data providing mostly documentation of published results (data points from plots and tables) are deposited at the

- Durham High Energy Physics Database (HEPData): https://hepdata.net/.

The website http://nucastrodata.org/index.html hyperlinks all online nuclear astrophysics datasets, hosts the Computational Infrastructure for Nuclear Astrophysics (CINA), and provides a mechanism for researchers to share files online.

Computer science

In computer science (but also physics, astronomy etc.) one research output is the development of code.

- Github: https://github.com

is an extremely popular platform to publish such output, and while behind Github is a commercial company, public projects can be stored for free.

Telecommunications

A library of test instances for **S**urvivable fixed telecommunication **N**etwork **D**esign is provided by

- SNDlib: http://sndlib.zib.de/home.action.

It contains realistic network design test instances available to the research community and serves as a standardized benchmark for testing, evaluating, and comparing network design models and algorithms. Every user can contribute by submitting new test instances, new solutions or dual bounds for existing test instances.

- Video Quality Experts Group (VQEG): https://www.its.bldrdoc.gov/vqeg/video-datasets-and-organizations.aspx

collects websites containing video content, including video test sequences. The

- Consumer Digital Video Library (CDVL): https://www.cdvl.org/

provides a repository of video content that is suitable for determining the effectiveness of consumer video processing applications and quality measurement algorithms. Users can share and download high-quality uncompressed video clips, which can be filtered using a clip descriptor and recommended usage guidance.

Metadata

In the situation where there is no public or community database for a data type, the ERC encourages grantees to deposit the metadata, including links to the data location, in a recognised resource.

A good example where standards for metadata have been established is given by the Virtual Observatory (VO) with the vision that astronomical datasets and other resources should work as a seamless whole. Many projects and astronomical data centres worldwide are working towards this goal via the International Virtual Observatory Alliance (IVOA - https://ivoa.net/). The IVOA debates and agrees the technical standards that are needed to make the VO possible. It also acts as a focus for VO aspirations, a framework for discussing and sharing VO ideas and technology, and body for promoting and publicising the VO.