

2018-4

## Open Research Data and Data Management Plans: Information for ERC Grantees

European Research Council

Follow this and additional works at: <https://arrow.tudublin.ie/dataguide>

---

### Recommended Citation

European Research Council, "Open Research Data and Data Management Plans: Information for ERC Grantees" (2018). *Guides*. 3.  
<https://arrow.tudublin.ie/dataguide/3>

This Article is brought to you for free and open access by the Data@TU Dublin at ARROW@TU Dublin. It has been accepted for inclusion in Guides by an authorized administrator of ARROW@TU Dublin. For more information, please contact [arrow.admin@tudublin.ie](mailto:arrow.admin@tudublin.ie), [aisling.coyne@tudublin.ie](mailto:aisling.coyne@tudublin.ie), [vera.kilshaw@tudublin.ie](mailto:vera.kilshaw@tudublin.ie).



**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 2.0  
24 April 2018



European  
Commission

Horizon 2020  
European Union funding  
for Research & Innovation

This document will be 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 Access](#) via the address [erc-open-access@ec.europa.eu](mailto:erc-open-access@ec.europa.eu).

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

<b>HISTORY OF CHANGES</b>			
<b>Version</b>	<b>Publication date</b>	<b>Change</b>	<b>Page</b>
1.0	23.02.2018	<ul style="list-style-type: none"> <li>Initial version</li> </ul>	
2.0	24.04.2018	<ul style="list-style-type: none"> <li>Part 'Open research data and data deposition in the Physical sciences and Engineering domain' added.</li> <li>Minor editorial changes; faulty link corrected.</li> <li>Contact address added.</li> </ul>	15-17  6, 10 2

# 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 mandatory. The next step in the development of open science is making research data publicly available when possible. This will benefit science by increasing the use of data and by promoting transparency and accountability.

The ERC embraces the so-called ‘FAIR data principles’: research data should be **findable, accessible, interoperable and re-usable**. This means that data should be:

- identified in a persistent manner using community conventions, and described using sufficiently rich meta-data;
- 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 data sets;
- licensed or have terms-of-use that spell out how they can be used by others.

The article “The FAIR Guiding Principles for scientific data management and stewardship” (Wilkinson et al., 2016, *Scientific Data* 3:160018, <https://doi.org/10.1038/sdata.2016.18>) provides a detailed discussion of the FAIR principles.

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. Any such restrictions on access should be explicit and justified, and such data should still be managed in line with the FAIR principles.

For researchers, the move to open 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 depositories.

## ERC requirements

### *Data Management Plans*

All ERC projects funded under the Work programmes 2017 and later participate by default in the Horizon 2020 Open Research Data (ORD) pilot, with the possibility for grantees to opt out at any time. 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 ORD pilot are required to submit a data management plan (DMP) within six months after the start of their grant.

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. *Data set description:*  
Grantees to provide a sufficiently detailed description, including the scientific focus and technical approach, to allow association of their data sets with specific research themes.
2. *Standards and metadata:*  
Grantees to describe the protocols and standards used to structure their data (i.e. fully reference the metadata) so that other scientists can make an assessment and reproduce the dataset. If available, to provide a reference to the community data standards that their data conform to and that make them interoperable with other data sets of similar type.
3. *Name and persistent identifier for the data sets:*  
Grantees should plan to use depositories that will provide a unique and persistent identification (an identifier) of for their data sets and a stable resolvable link to where their datasets can be directly accessed. Submission to a public depository normally provides this; many institutional depositories provide similar services.
4. *Curation and preservation methodology:*  
Grantees to provide information on the standards that will be used to ensure the integrity of their data sets and the period during which they will be maintained, as well as how they will be preserved and kept accessible in the longer term. If available, to provide a reference to the public data depository in which their data will reside.
5. *Data sharing methodology*  
Grantees to provide information on how their data sets can be accessed, including the type of license 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 underlying publications it could be, for example, at the time of publication or pre-publication.

A DMP that provides adequate information on these five topics will meet the FAIR principles.

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 are encouraged to use the ERC template that is available on the Horizon 2020 Participant 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](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>

Grantees should also keep in mind the following guidance document:

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

Writing a DMP should not be regarded as simply an administrative hurdle. Rather, it should provide a positive stimulus to thinking about how the data generated within a project will be stored, managed and safeguarded, and should be part of the research process from the onset. 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 ear-marked 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.

Publications present the pertinent data underlying conclusions made in a research paper and publishers increasingly require that all relevant data should be made available to the community. The ERC expects data underlying publications by ERC grantees to be available. 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 made available in order to promote scientific progress; the European Commission has adopted a policy of open data for all research that they finance. Data dumping is of course to be avoided and 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 what 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 depository for research data it is important to check whether the depository:

- 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 meta-data standards;
- and specifies a license governing access and re-usability of the data.

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

- Core Trust Seal (this list includes depositories certified by the Data Seal of Approval and accredited by the World Data System): <https://www.coretrustseal.org/why-certification/certified-repositories/>
- Nestor seal (DIN-Norm 31644): <http://www.dnb.de/Subsites/nestor/EN/Siegel/siegel.html>
- ISO 16363 certified depositories: <http://www.iso16363.org/iso-certification/certified-clients/>

### *General depositories for research data*

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

- Zenodo (not-for-profit, hosted by CERN): <https://zenodo.org>:
- Dryad (not-for-profit membership organisation): <http://www.datadryad.org>
- Figshare (free service provided by private company): <https://figshare.com>
- Open Science Framework (not-for-profit, developed and maintained by the Center for Open Science<sup>1</sup>): <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 depositories, 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 depositories 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 depositories:

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

A growing number of universities and research institutes host a depository for use by their research staff. Most of these institutional depositories are originally set up for storing (open access) publications, but dedicated research data depositories also occur. In order for an institutional depository to be acceptable as a trusted archive, it is essential that the university/institute has a data policy guaranteeing the support for data storage and sharing.

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 most often, in contrast to public data depositories, these are 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

---

<sup>1</sup> <https://cos.io>

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.

Many journal websites contain lists of depositories. In addition, there are an increasing number of commercial publishers that offer authors opportunities to store the research data underlying their publications.

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

What proper metadata are depends on the discipline and/or the methodology that was used to produce the data. Discipline-specific depositories often have detailed requirements for describing data that are stored in that depository.

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: <http://dublincore.org>

For more information on disciplinary metadata standards see also

- Digital Curation Centre: <http://www.dcc.ac.uk/resources/metadata-standards>

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

- RDA Metadata Directory: <http://rd-alliance.github.io/metadata-directory/>

### *Policies of other funding organisations*

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

- Netherlands Organisation for Scientific Research (NWO): "Data management protocol"  
<http://www.nwo.nl/en/policies/open+science/data+management>
- German Research Foundation (DFG): "DFG Guidelines on the Handling of Research Data"  
[http://www.dfg.de/en/research\\_funding/proposal\\_review\\_decision/applicants/submitting\\_proposal/research\\_data/index.html](http://www.dfg.de/en/research_funding/proposal_review_decision/applicants/submitting_proposal/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](http://www.snf.ch/en/theSNSF/research-policies/open_research_data/Pages/default.aspx)
- The Research Council of Norway (RKN): "Open Access to Research Data":



[https://www.forskningradet.no/en/Article/Open\\_access\\_to\\_research\\_data/1240958527698](https://www.forskningradet.no/en/Article/Open_access_to_research_data/1240958527698)

- UK Arts and Humanities Research Council (AHRC): “Technical plan”:  
<http://www.ahrc.ac.uk/funding/research/researchfundingguide/attachments/technicalplan/>  
see also: “Peer review updates and guidance”:  
<http://www.ahrc.ac.uk/peerreview/peer-review-updates-and-guidance/>

The following document by DG Research of the European Commission is also instructive:

- “Guidelines on FAIR Data Management in Horizon 2020”:  
[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf)

~~~~~

In what follows more specific information is given for ERC grantees in the Life Sciences, the Physical Sciences and Engineering, and for those working in the Social Sciences and Humanities. This will include references to specialised depositories for specific disciplines where such are available, and more general information in other cases.

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 depositories, protocols, meta-data or sharing methodologies an ERC grantee chooses to use.

## Open research data and data deposition in the Life Sciences domain

The Life Sciences have a long tradition of open access data depositories. Submission of datasets to an established public depository is considered good scientific practice and is often also a condition for publication. The public depositories ensure that data are correctly curated, accessible and maintained in the long term. Data publication through such a depository will make your data FAIR. In addition, several 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” (Cousijn et al, 2017, bioRxiv, <https://doi.org/10.1101/100784>).

### *Established public depositories*

The ELIXIR European consortium, the ESFRI research infrastructure for life science data, has compiled a list of recommended depositories:

- ELIXIR Deposition Database list:  
<https://www.elixir-europe.org/platforms/data/elixir-deposition-databases>

Many of these are based at the European Bioinformatics Institute (for advice on data deposition see <http://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 depositories, although many do not take public submissions.

Further information on a broad selection of databases, policies and standards relevant to the life sciences are available on <https://www.FAIRsharing.org>

### *Imaging data*

In the rapidly developing area of microscopy and imaging data, solutions for public archiving of data sets are currently being built. There is already an electron microscopy public image archive:

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

The new European research infrastructure Euro-BioImaging covers a wide range of imaging approaches:

- European Research Infrastructure for Imaging Technologies in Biological and Biomedical Sciences (Euro-BioImaging): <http://www.eurobioimaging.eu/>

Euro-BioImaging is actively promoting the development of a public BioImage Archive in strategic collaboration with the EBI and ELIXIR. Euro-Bioimaging's interim service, the Image Data Repository (IDR), already accepts several types of light microscopy data (<https://idr.openmicroscopy.org/about/>).

### *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 data bases requires ‘controlled access’ for data protection. This is a rapidly evolving area where community standards and depositories 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 publically accessible registries identified by the World Health Organisation:

- International Clinical Trials Registry Platform (ICTRP): <http://www.who.int/ictcp/en/>

### *Other types of depositories*

In a number of research areas, the research community has generated specific archives. These may be depositories which aggregate data from multiple underlying depositories, 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: <http://flybase.org/>
- WormBase: <http://www.wormbase.org/>
- The Zebrafish Information Network (ZFIN): <https://zfin.org/>

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/>

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 depository is highly recommended to ensure long-term curation and preservation.

### *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 depository. 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 depositories such as:

- Dryad Digital depository: <https://datadryad.org/>
- Zenodo: <https://zenodo.org/>
- Figshare: <https://figshare.com/>

Data files for which there is no depository, including imaging data, can be deposited in BioStudies (<https://www.ebi.ac.uk/biostudies/>, see below).

## Metadata

It is essential that data deposited in data storage facilities should be correctly annotated and standardised and be accessible to the research community. 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, such as:

- BioStudies: <https://www.ebi.ac.uk/biostudies/>

Many studies generate multiple types of data, deposited in more than one depository. These may be cross-referenced in a community database. Deposition of metadata in the Biostudies database is also a convenient way to link data via data identifiers to wherever they are physically deposited. Often a BioStudies record corresponds to the data behind a paper and so usefully provides a link to the funding behind the study.

For a discussion of metadata resources, see the section on ‘Interoperability’ on the ELIXIR website at <https://elixir-europe.org/platforms/interoperability>.

## 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 (depositories) 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 feature 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 data sets; experimental data; observational data; interviews; archival data; human artefacts; medical and genetic data; and so forth. And 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 copy-righted 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 depositories
- discipline-specific depositories
- metadata and data preparation

### *General depositories*

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 depositories for use by social sciences and humanities disciplines, but it is certainly not exhaustive.

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

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

CESSDA is a so-called 'ERIC', a European Research Infrastructure Consortium ([http://ec.europa.eu/research/infrastructures/index\\_en.cfm?pg=eric](http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=eric)), i.e., an international entity established by the European Commission that has national governments or consortia as its members. Currently, CESSDA has seventeen members, all of them national agencies that operate on a not-for-profit basis.

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

The geographical coverage of CESSDA is growing. Missing at the time of writing are some Southern European Countries (Italy, Spain) and some EU-13 countries (e.g., Poland, Romania, Bulgaria, Estonia, Latvia). Another prominent absentee is Ireland.

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

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

ICPSR is a not-for-profit membership organisation that maintains a data archive in the social and behavioural sciences. Currently ICPSR has a membership of more than 750 academic institutions and research organizations.

## **Discipline-specific depositories**

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

### *Linguistics*

- Linguistics Linked Open Data (LLOD): <http://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 19 national consortia as full members and 2 consortia as observers. Still missing at the time of writing are Spain, Ireland, and Romania.

### *Historical sciences*

Depositories for the historical sciences are mostly at the institutional or national level. A number of CESSDA archives also accept historical data sets.

### *Archaeology*

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

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

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):  
<http://www.dariah.eu>

DARIAH is another ERIC. It is a pan-European infrastructure for arts and humanities scholars working with computational methods. It has seventeen members and eight co-operating partners. Missing at the time of writing are Spain, and a number of EU-13 countries.

Note that several CESSDA archives also accept humanities data sets.

### *Psychology*

The Leibniz Institute for Psychology Information (ZPID)<sup>2</sup> has developed a data-sharing platform specialized for psychology research:

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

For an extensive overview of data depositories in psychology, see the article “Finding a Home for Your Science” (K. Andrew DeSoto, 2016, Observer, Volume 29, Issue 5, <https://www.psychologicalscience.org/observer/finding-a-home-for-your-science>).

### *Demography*

- Data Sharing for Demographic Research (DSDR):  
<http://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 data sets.

## **Metadata and data preparation**

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

- Digital Curation Centre (DCC):  
<http://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, ICPSRC:  
<http://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>

---

<sup>2</sup> <https://www.zpid.de>

## Open research data and data deposition in the Physical sciences and Engineering domain

The PE domain has a large number of data depositories. In the following section a number of areas are mentioned in 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 depositories

#### *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 (<http://simbad.u-strasbg.fr/simbad/>), the world reference database for the identification of astronomical objects; VizieR (<http://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 (<http://aladin.u-strasbg.fr/aladin.gml>).

#### *Chemistry*

The use of public depositories 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: <http://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.



## *Geophysics*

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) :  
<http://www.fdsn.org/webservices/datacenters/>

The data management center of IRIS – Incorporated Research Institutions for Seismology (<http://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: <http://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 (<http://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 as well as visualization and modelling tools. At present, EPOS includes ~300 research institutions from 25 European countries. It is presently in its implementation phase (2014 – 2019), which is funded under the Horizon 2020 EU Framework Programme for Research and Innovation.

## *Material sciences*

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): <http://www.crystallography.net/cod/>

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

## *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/>.

## *Software Engineering*

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 **Survivable fixed telecommunication Network Design** 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): <http://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 by clip descriptor and recommended usage guidance.

- QUALINET Databases: <http://dbq.multimediatech.cz/>

accepts and shares numerous datasets used in the field of Quality of Experience research in multimedia media systems.

## **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 - <http://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.