CEU Biblioteca

How to manage research data

By research data we mean all material that has been recorded and is used to demonstrate the results achieved. These data can be numerical, descriptive or visual, raw or analysed, experimental, or observational.

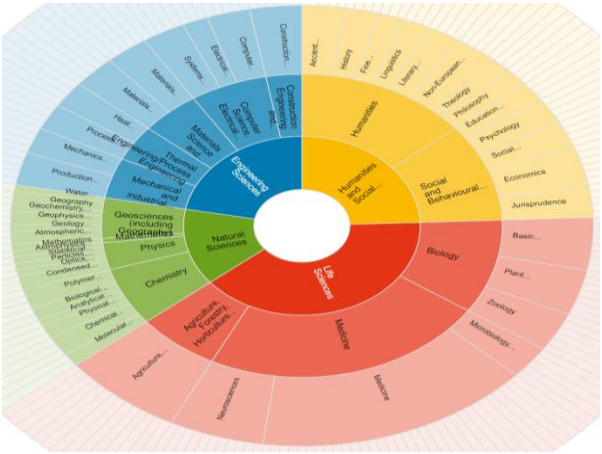
A) Means for the dissemination of research data:

1. **As material attached to the article**: publishers indicate how this supplementary material is to be attached. Depending on the type of data, formats will vary.

   Thus, protocols, sequences, structures, will be stored in a data repository according to the subject, such as Genbank; and the resulting article must contain the identification number assigned to it in this repository.

2. **As a data article in a data journal**: these are collected in Data Journals which are open access so that they can be reused and cited.

3. **Find a data repository to deposit them**: these can be institutional or thematic, it may be useful to consult re3data; also the Scientific Data has a guide to repositories



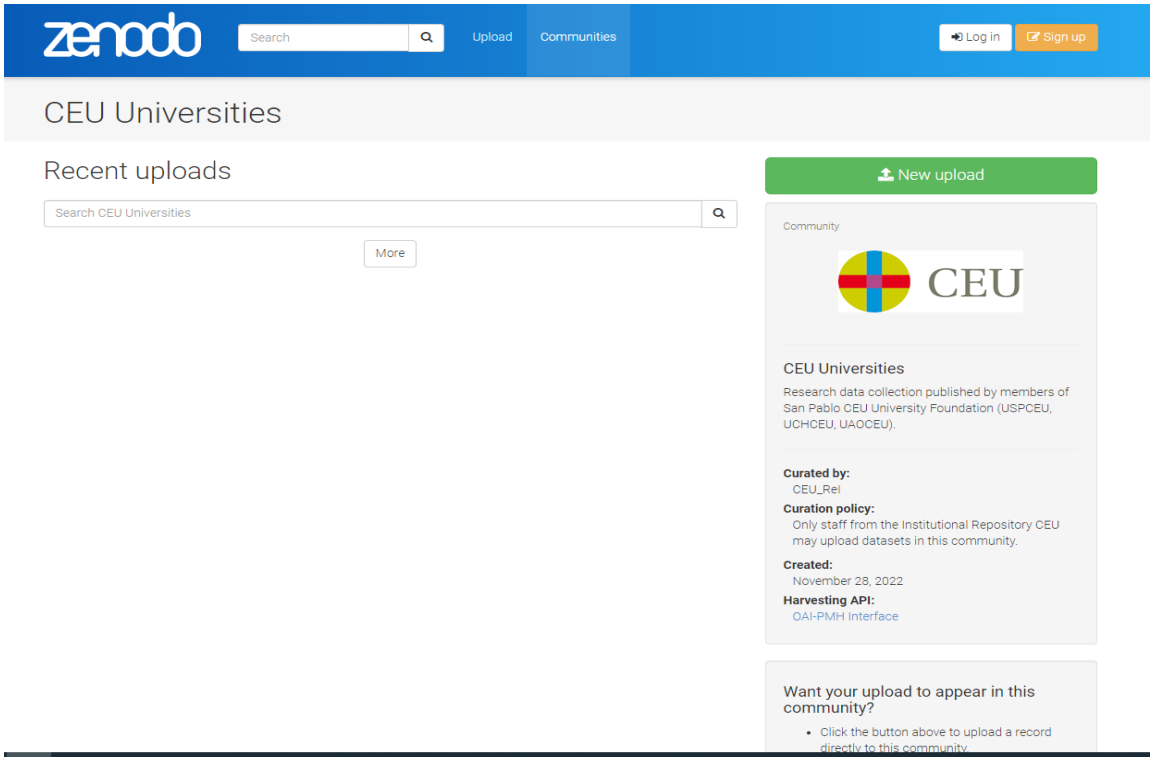
Source: re3data allows you to search the data repository by subject.

Zenodo is a multidisciplinary repository created by OpenAIRE and CERN, with the support of the European Commission:

- Supports multidisciplinary research data from any researcher, scientific community or R&D institution.
- Accepts all types of data and formats
- Allows large storage (1 GB)
- Encourages open data repository (although it offers all types of licences)
- Assigns a digital identifier (DOI) that proves its publication.

From CEU Repositorio Insitucional (*CEU Institutional Repository*; CEU RepI) we have created a community in **Zenodo** to host the research data of our universities:

CEU Universities: https://zenodo.org/communities/ceu_rei/?page=1&size=20



Submissions can be made to this Community and Collections will be formed; whenever the datasets are uploaded the CEU RepI administrator will review them to ensure that they meet the membership requirements.

Furthermore, we invite all CEU researchers who have already created their Collection in Zenodo to transfer it to our new *CEU Universities* community.

We remind you that there are dataset search engines:

- **Google Dataset Search**: Google's dataset search engine for researchers

- **DataCiteSearch**: generalist search engine for open datasets with DOI.

B) A number of **good practices** are advised to ensure that the whole process is effective:

Processing: data files must be ordered and have a cohesive structure. It is advisable to:

- Name the files according to the same criteria: to be meaningful and short (with no special characters or accents). Indicate the version when naming the file.
- Make backup copies and try to make a definitive version.

Public and private data: select those sensitive data that may not appear in the open relating to intellectual property Law 23/2006, to the protection of personal data and to industrial property.

Data obtained through interviews, questionnaires, etc. must be taken into account: on the one hand, requesting the consent of the persons involved by means of a form and, on the other hand, applying an anonymisation process in those files containing sensitive or confidential data.

OpenAire recommends the Amnesia tool that transforms relational and transactional data into anonymised datasets where formal privacy safeguards are maintained.

File formats: open and standard so that they can be read in the future.

There are international recommendations :

- **Containers:** TAR, GZIP, ZIP
- **Databases:** XML, CSV, JSON
- **Geospatial:** SHP, DBF, GeoTIFF, NetCDF
- **Video:** MPEG, AVI, MXF, MKV
- **Sound:** WAVE, AIFF, MP3, MXF FLAC
- **Statistics:** DTA, POR, SAS, SAV
- **Images:** TIFF, JPEG 2000, PDF, DNG, GIF, BMP, SVG
- **Tabular data:** CSV, TXT
- **Text:** XML, PDF / A, HTML, JSON, TXT, RTF
- **Internet Archive:** WARC

Licences for use and re-use: the licence we have chosen must appear attached to the "readme file" data, the most common are: Creative Commons and Open Data Commons

It is necessary to have a **Unique and Persistent Identifier** to locate and preserve the research data in the repository of choice to store them. There are several:

URI (persistent resource identifier), **DOI** (unique and permanent identifier for electronic publications), etc.

Cite each dataset independently, whether it is a data article published in a data journal or a dataset archived in a repository. It is recommended to include a persistent author identifier (e.g. ORCID).

C) **Data Management Plan** (DMP) is the report that encompasses all phases: creation or collection, processing and analysis, preservation, publication and re-use.

It sets out at the outset how data will be managed, what data will be shared, how it can be used and how it will be preserved, before it is created or collected. It is not a final document, but is modified during the project.

Horizon 2020 has given way to the new **Horizon Europe 2021-2027** programme, which is the European Union's Framework Programme for Research and Innovation. Therefore, the new Horizon Europe programme strengthens the open access mandate and extends it to open science. Beneficiaries of EU-funded projects are obliged to submit a Data Management Plan(DMP).

The template designed by Digital Curation Center to fill in the requested fields can be used to design the **DMP:**

- **Administrative information:** centre identifier, funder, funding identification number, project name, description, MR (Main Researcher) MR's Orcid identifier, date of first version, etc;
- **Data description:** typology, volume, formats, file structure and version management;
- **Documentation and metadata** to accompany the data;
- **Legal and ethical requirements:** consent of the persons involved, protection of identity, processing of sensitive data and management of intellectual property rights;
- **Storage and backup:** control of access to collaborators and management of backups;
- **Selection and preservation:** what data and how it will be preserved, where it will be preserved and restrictions on data exchange;
- **Data sharing**: under what conditions data will be shared, how users will find the data, when it will be available, will get a persistent identifier, etc..;
- **Responsibilities and resources:** allocation of responsibilities, costs associated with open access data, as they are eligible for reimbursement during the project period.

Several **tools** are available to guide us in the development of the **DMP**:

- PAGODA which is the exclusive data repository for the Madroño Consortium.
- ARGOS promoted by the European Commission, allows the creation of DMPs that can be freely exchanged between infrastructures.

FAIR (findable, accessible, interoperable and reusable) principles need to be followed in order to be able to share and reuse research data legally in future research.

1. **Findable:** describe the metadata as fully as possible together with the assigned permanent identifier, another metadata is the DOI or URI which is given by the data journal and the data repository respectively.
2. **Accesible**: use standard protocols that facilitate their open and free communication.
3. **Interoperable**: metadata must use formats, languages, and vocabularies supported by the scientific community.
4. **Reusable**: must use an open, machine-readable licence, provide information on provenance.

The **CEU Repositorio Institucional** of **Universidad San Pablo-CEU** will be pleased to help you by mail. sdigital.bib@ceu.es