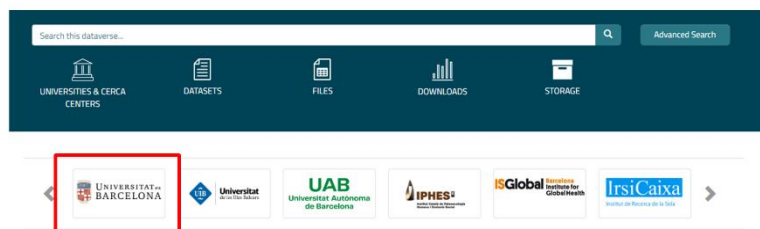


PROCEDURE FOR PUBLISHING RESEARCH DATA IN THE DATAVERSE FOR RESEARCHERS

1. First step when publishing data to the CORA Research Data Repository.

- a) Go to <https://dataverse.csuc.cat/> and create a user account with your UB email. You can follow the procedure created by the CSUC (<https://confluence.csuc.cat/pages/viewpage.action?pageId=97231135>)
- b) You will be assigned a role that allows you to create datasets with your research data, but you will not be able to publish them. This is the responsibility of the UB Data Management Technician. Therefore, once we have entered the necessary data and information, we will need to contact them. To create the dataset, we will have to go to the instance where we want to publish. If you do not have a specific instance (or dataverse) for your project/research group, you will have to go to the University of Barcelona's instance:



2. How do I correctly format the data?

- a) First, we always look for an appropriate format. At the following CSUC URL, you'll find a table showing the correct options for each case.
URL: <https://confluence.csuc.cat/spaces/RDM/pages/166789435/Format+recommendations>
- b) With the data, there must always be a legend or explanation of what each data point represents. For example, if we use numerical data in a CSV, we must explain what each column and row is, the units used, etc. This can be done in the same document or by adding a "Codebook" to the dataset with all the necessary explanations. This process facilitates data reuse.

3. How do I correctly format the metadata?

To make the data more easily findable, we will add a series of metadata (information about the data in question). The more information we provide about the data, the better.

This information must include the following sections:

- A simple title that describes the data (not the article for which it is used) and is easy to understand.
- Description of the data itself

- Authors (again, the people who participated in collecting the data, not the article, paper...) with their ORCID. At least one of them must be affiliated with UB.
- Contact person
- Keywords for the data
- Description of how this data is organized.
- Grant of the project, if available

To supplement the information about this data, we can add other metadata (Location, Date of collection...).

The CRAI is responsible for filling in other metadata such as the “Topic Classification.” For the “Topic Classification” we will use the UB Thesaurus website (<https://vocabulary.crai.ub.edu/ca/thub>). There we will find the general vocabulary for defining each data set. Again, to make this selection, we must focus on the data and not on the article or paper to which they are related. Between 2 and 6 words must be chosen.

4. Decision on the license for the use of research data

You need to decide which “Data Use Agreement” you prefer for the data. The CRAI is committed to open science, meaning it recommends using licenses that allow adapting and sharing the data, provided that credit is given to the data author.

A useful tool for deciding on the license is the following: <https://ufal.github.io/public-license-selector/>

5. Creating the Readme.txt

Once you have filled in the metadata and chosen the license, we proceed to create the README.txt. This is a document with general information, with the following structure:

```

Fixer Edició Format Visualització Ajuda
DD-MM-YYYY
Contacte
Mail Contacte

GENERAL INFORMATION
-----
Dataset title:
    Títol

Authorship:
    Nom Investigador/a 1
    Departament Investigador/a 1
    ORCID: 0000-xxxx-xxxx-xxxx

    Nom Investigador/a 2
    Departament Investigador/a 2
    ORCID: 0000-xxxx-xxxx-xxxx

DESCRIPTION
-----
Descripció de les dades
Keywords: Keyword1, keyword2...
Data collected: Month day & day, Year

ACCESS INFORMATION
-----
Dataset licensed under a "Llicència escollida" (text available at https://creativecommons.org/licenses/Llicència_escollida)
Dataset DOI: https://doi.org/xxxxxxxx

FILE OVERVIEW
-----
Descripció de l'organització de les dades
Descripció dels paràmetres amb els que es classifiquen les dades, si escau

```

You can also find templates in different languages at:

<https://confluence.csuc.cat/pages/viewpage.action?pageId=104529933>

The CSUC has scripts that automate the process of creating the README:

https://colab.research.google.com/github/CSUC/RDR-scripts/blob/main/create_Readme/create_Readme_script.ipynb

6. Creation of the private link

You can generate a link to share this unpublished version of the dataset with others, even if they don't have an account in the repository. The link will stop working when the dataset is published or if it is manually disabled.

You can only have one active preview URL per dataset:

- Full Preview: Creates a link for others to review this version before it is published. They will have access to all files and will be able to see all metadata, including authoring information.
- Anonymous Preview: Creates a link for others to access an anonymized version of this preview. Metadata that could reveal authorship or the space name will not be displayed. Files will only be accessible if they are not restricted.

7. Publication of the data

Once we are satisfied with our dataset, we will contact the CRAI research support office to have them review and proceed with publishing the data.

8. New Versions

Once published, if you want to make changes to the dataset, new versions will be created each time.