

CORA.Repositori de Dades de Recerca Regulations

(Doc.CO22/X) (4 RDM\RDR\Acords i Procediments (Reglament)\ReglamentiProcedimentsCORA_EN.docx)

Criteria for depositing

1.1 Who can deposit?

- To deposit and publish a dataset, at least one of the authors of the dataset must be affiliated to the university or research center at the time of obtaining and/or generating the data. In the event that no author is affiliated, those datasets of authors that have the authorization of the institution will also be accepted.
- At least one of the authors is required to have a unique identifier, such as ORCID, and that this information be included in the metadata.
- Publication in the repository will require validation of the technical dialogue of the institution according to the needs or operation of the entity responsible for the data (university or research center).
- It is up to the author to inform the principal investigator and/or director of the research project and have their prior approval. In the same way, doctoral students will follow the procedures of their institution.
- Institutions may approve exceptional cases where there is a collective delivery of the institution itself or one of its units where individual authorship does not have to be recorded. Even so, it will be necessary to indicate in the metadata the name of a person as depositor and contact of the dataset.

1.2 What is allowed to deposit?

- The deposit and publication of research data is allowed. Research data is understood as the non-theoretical evidence that supports the conclusions derived from a research work and that can be used to validate these conclusions. It can be quantitative or qualitative information collected during the development of an investigation through experimentation, observation, modeling, interview, information extracted from other sources (from which rights may be retained) or from other methods. Data can be primary or derived from primary data.

- Any file format is accepted. However, it is advisable to provide data in non-proprietary formats in order to ensure wider use and for long-term sustainability reasons. The formats indicated in the table are recommended and preserved: "[Recommended and accepted formats](#)".

1.3 What is not allowed to deposit?

Research management systems (CRIS) are responsible for including information from datasets made at the institution, regardless of where they have been deposited. The Research Portal of Catalonia draws on the information provided by the CRIS systems of the different participating institutions and is, therefore, the instrument that catalogues the datasets of the research system of Catalonia.

- Datasets that do not comply with the requirements of 1.4 of this regulation and for which we only want to have metadata information in order to have proof of it.

1.4 What should a dataset contain?

- A dataset must include the files containing the data, the associated metadata, the descriptive documentation, the [rights and the license of use](#).
 - The type of data and formats that the files must contain is described in section 1.2 of the [Regulation](#).
 - The [minimum mandatory metadata must be](#) entered following the established guidelines.
 - The mandatory minimum [descriptive documentation](#) is a plain text file (README) that describes the content of the dataset.
 - The licence must follow the requirements described in section 2 of the [Regulations](#).
- Each dataset must have at least one designated contact in order to guarantee, at least during the established preservation period, any request related to it.

1.5 Size of datasets

- The deposit of datasets up to 10 GB by default is allowed.
- For datasets between 10GB and up to 100GB, it is necessary to make a request following the "[Procedure to deposit datasets of more than 10GB and up to 100GB](#)".
- Despite the limit of 100GB, in these initial stages and, in order to gain experience, larger datasets will be accepted. These cases must have an express request signed by the

university by the vice-rector for research of the university, or whoever delegates, or by the research centers by the scientific director, or whoever performs their functions; as well as, the approval of the General Directorate of CSUC following the "[Procedure for depositing datasets of more than 100GB](#)".

- In relation to points 2 and 3, in the event that any of the files contained in the dataset exceeds 10 GB, it must be communicated to CSUC so that the limit can be modified during a period agreed between institutions. After this time has elapsed, the RDR will return to the standard limit for uploading files up to 10GB.

2. Licensing

When a *dataset is published* in CORA. RDR, a license must be defined in order to inform users of the conditions of reuse of the data. The use of [Creative Commons](#) (CC) CC0 and CC-BY licenses is recommended following the European Union motto "As open as possible and as closed as necessary" to improve the discovery, understanding and reuse of research data.

- The person making the deposit must select a license from a predefined list. CORA. RDR offers a full range of Creative Commons licenses, adopting CC0 (Creative Commons Zero) as the default defined license.
- The person making the deposit can define a different license to the Creative Commons, selecting "Custom Dataset Terms". In this case, you must describe the conditions of use of the data, according to the license chosen.
 - Specific licenses can be used for research data and databases, such as those offered by [Open Data Commons](#): ODC-BY, ODbL and PDDL.
 - Specific licenses may also apply for software, approved by the [Open Source Initiative](#) (OSI): [GNU General Public Licenses](#), [European Union PublicLicense](#) and [MIT license](#), among others.
- The metadata of the *datasets* published in CORA. RDR are accessible through the OAI-PMH protocol and subject to the [CC0 license](#).

3. Level of opening of datasets

- **Open data.** There are no restrictions on their access, allowing anyone to view and download them. This is the most appropriate option for publishing data that is not confidential, sensitive, commercial or with other legal restrictions.

- **Data seized.** There are temporary restrictions on access. This embargo may apply to all or part of the dataset, so that they are not accessible for a maximum period of two years. However, dataset metadata will always be visible.
- **Restricted data.** There are restrictions on access. The researcher and/or institution may consider restricting research data as long as it is for reasons of copyright, ethical or confidentiality issues. It should be emphasized that the basic principle of publicly funded research data is that data should be as accessible as possible and as protected as necessary. However, dataset metadata will always be visible.

4. Unsubscribe a dataset

To remove a dataset published in the repository it must be canceled. This will mean that it will no longer be findable, but it will be possible to consult, through the DOI, the citation of the dataset and the reason why it has been unsubscribed. Under no circumstances can the DOI be removed, as it is an identifier registered in [DataCite](#) and must be permanent over time.

- It is not allowed to delete a published dataset but it can be unsubscribed. This action should only occur if there is a legal and/or valid reason.
- To cancel a dataset, you must [contact the](#) person responsible for the instance indicating the reasons. The institution, following its own considerations, will or will not enforce the withdrawal.
- Draft datasets that have been deposited in the repository for more than 12 months are susceptible to deletion. Before this deadline, the person responsible for the instance will contact the person in the contact section of the register to assess the status of the dataset and carry out the relevant action.

5. Personal data

CORA. RDR does not currently accept:

- **Personal data.** Personal data that directly or indirectly identifies specific persons cannot be entered into CORA. RDR. These data need special protection and specific rules to avoid the risks derived from their treatment. Anonymised personal data can be deposited there.

6. Definitions

(under development)