SbA
sistema bibliotecario di ateneo

Datarepository Unive
GUIDELINES

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani PIANO NAZIONALE DI RIPRESA E RESILIENZA
Università Ca'Foscari Venezia

Before reading these guidelines, remember :
PRESERVING DATA DOES NOT NECESSARILY MEAN

1. WHAT IS THE REPOSITORY?

Datarepository Unive is the institutional repository that Ca' Foscari University has adopted for managing research data.

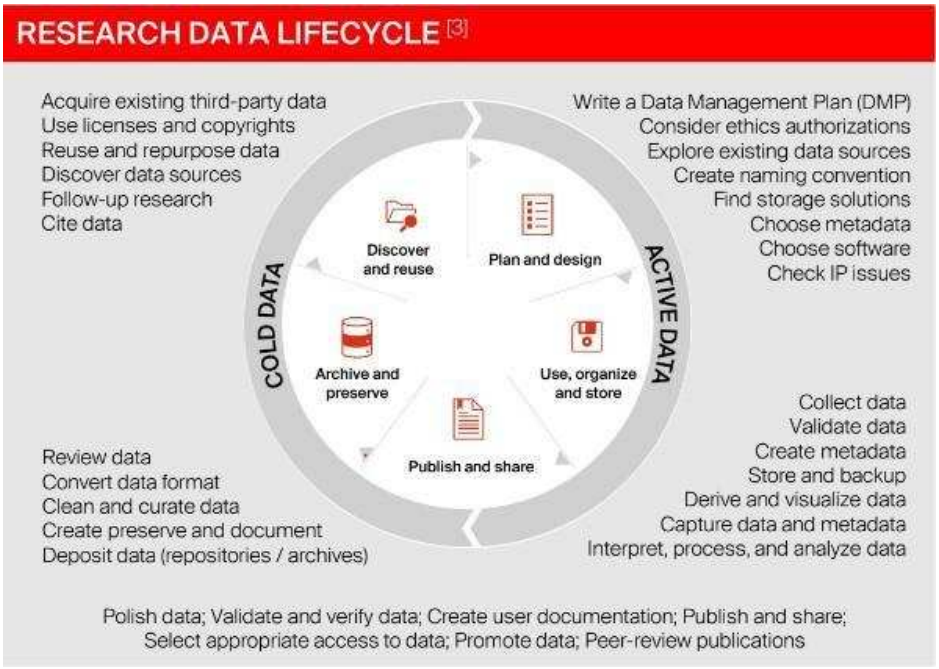
As a researcher, you must preserve the data created or collected during your research, ensuring their security and integrity. Datarepository Unive is the platform that our University provides for this purpose.

Preserving data does not necessarily mean publishing and making it open: even if you cannot share your data, it is still recommended to deposit it in Datarepository to ensure safe storage.

Datarepository Unive is not designed to deposit publications (which should be archived in ARCA).

N.B.: Datarepository Unive is not merely a backup archive for securing data but a platform that meets the criteria identified by the Confederation of Open Access Repositories (COAR) and endorsed by the European Commission for defining a "trusted repository" (see Annotated Model Grant Agreement).

2. RESEARCH DATA LIFECYCLE



(source: https://zenodo.org/records/7551315)

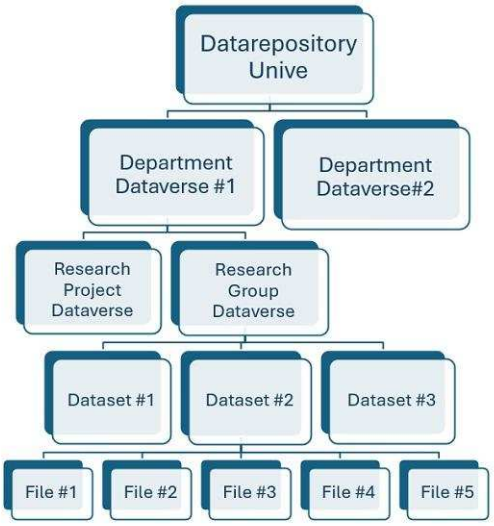
If you are the principal investigator of a research project:

- from the initial and analysis phases, you must ensure the storage, integrity, and safety of the data (active data) by using a system for storing and backing digital objects;
- In the final stages of reviewing, refining, and curating the processed data, consider transferring the data you will need to illustrate your research (cold data) to Datarepository Unive

3. *STRUCTURE OF DATAREPOSITORY UNIVE*

Datarepository Unive has a hierarchical structure and can be defined as a large box containing:

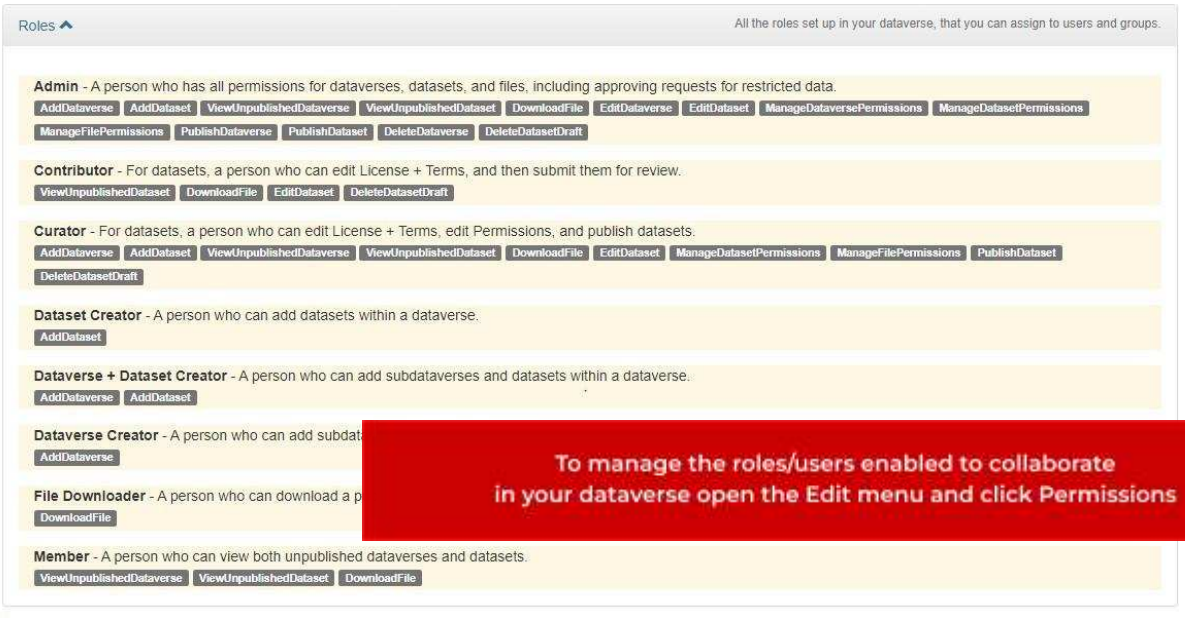
- 8 departmental dataverses (one for each Ca' Foscari department)
- each departmental dataverse can contain many dataverses, ideally one for each research project or research group
- every single dataverse is, in turn, a container for one or more **datasets**
- each dataset - endowed with a **unique DOI** - contains the the research data files.



A dataverse is a container of datasets created around the same research project or having the same broad topic. Each dataset has a unique DOI and contains research data files.

If you are the **admin** of a dataverse, i.e. you are responsible for the collection of datasets uploaded into it, you can assign roles and permissions to the users of that collection, determining who may contribute to the upload of datasets and in what role:

- contributor
- curator
- dataset creator
- dataverse+dataset creator (use of this role is not recommended)
- file downloader
- member



4. HOW TO DEPOSIT DATA

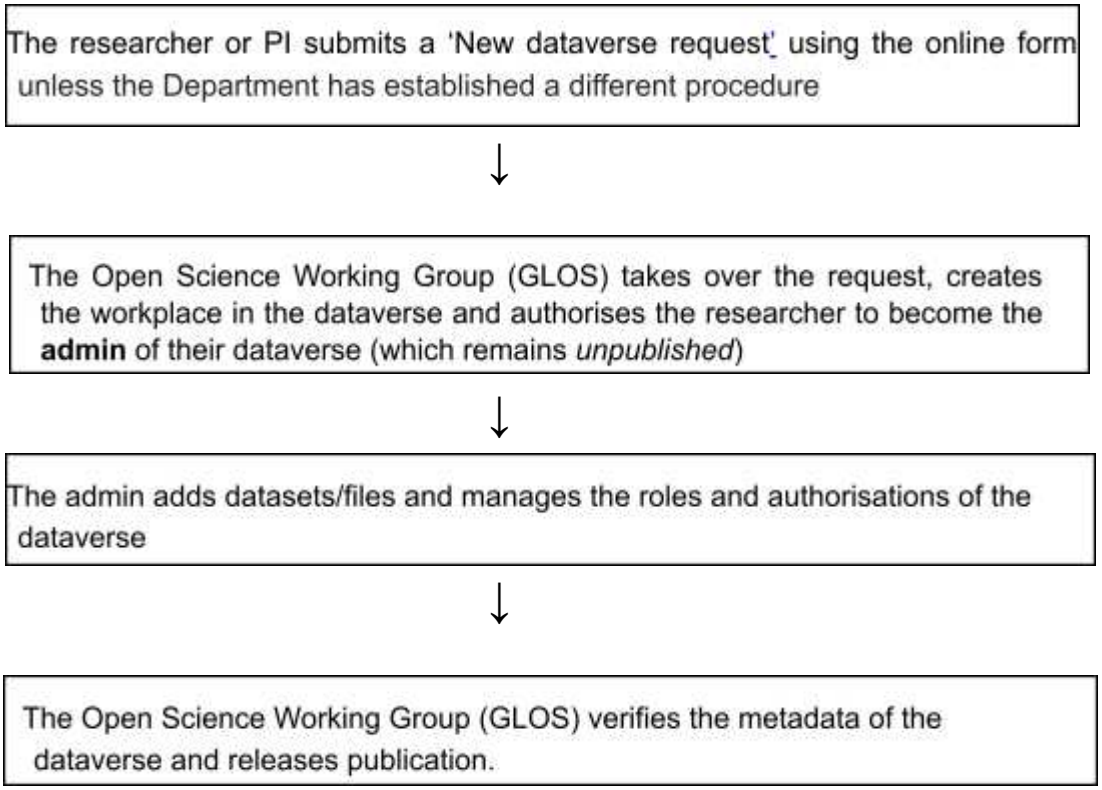
To deposit data, use the online form and submit a New dataverse request to the Open Science Working Group (Gruppo di lavoro su Open Science - GLOS)*. They will take over the request, create the workplace in Dataverse and authorise the researcher to become administrator of their (unpublished) dataverse unless the Department has established a different procedure.

To access it, you must log in with your University credentials using this link: https://datarepository.unive.it.

The dataverse administrator can manage the roles of other contributors - if any - and manage to upload one or more datasets with metadata and licences.

The Open Science Working Group (GLOS) verifies the compliance of metadata and licences before releasing the publication.

Workflow



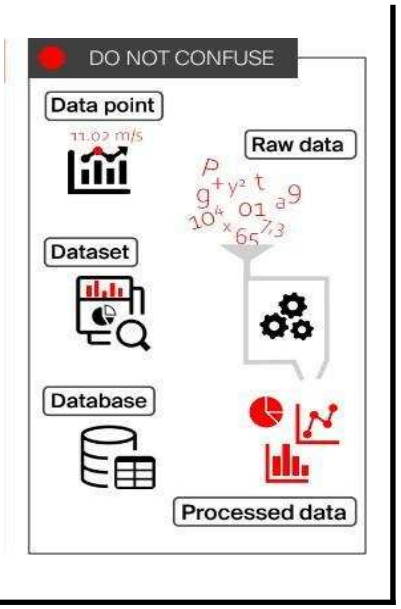
(*) The Open Science Working Group (Gruppo di lavoro su Open Science - GLOS) will be in charge of implementation until the University Library System/Digital University Library Open Science service is set up.

Don't forget to review the checklist before depositing data.

5. DATA: TYPES AND FORMATS

If you plan to use the platform, evaluate the types of data you want to store and consider the format, accessibility, and purpose of your data.

Raw data original data collected during research, but not yet processed	acceptable
FAIR data: data derived from the cleaning, anonymisation and checking of original data	appropriate
Analysed data Data resulting from the analysis and/or integration of processed data. It can take the form of tables, texts or graphs to facilitate understanding and communication	appropriate



Note that the starting data for your research may be reused data, that is, raw or processed data from third parties In this case, make sure you have the owner's consent and consider the need to refer to the original dataset if it is already available in the Open Science ecosystem.

	Type	Descrption	*Examples*
	Observational data	Data acquired in real-time in situ and which cannot be retrieved, recreated or replaced	*Sensor surveys; sensory (human) observations; survey results; notes/transcripts of interviews*
	Experimental data	Data collected under controlled conditions, in situ or laboratory; it is reproducible data, albeit at a high cost.	*Gene sequencing; chromatograms; spectroscopy; microscopy*
	Simulation data	Data generated by imitating the operation of a real-world process or system using computerised test models	*Climate models; economic models; biogeochemical models*
	Derived/compiled data	Data generated from existing data, often with different sources	*Derived variables; compiled database; 3D models*
	Reference/ canonical data	Data sets from static or organic [peer-reviewed] collections, generally published	*Databases of genetic sequences, chemical structures, census data, spatial data portals*
	Metadata	Structured information associated with data and essential for discovering, describing, using, managing, and preserving the data.	*README files; keywords; file and folder names*

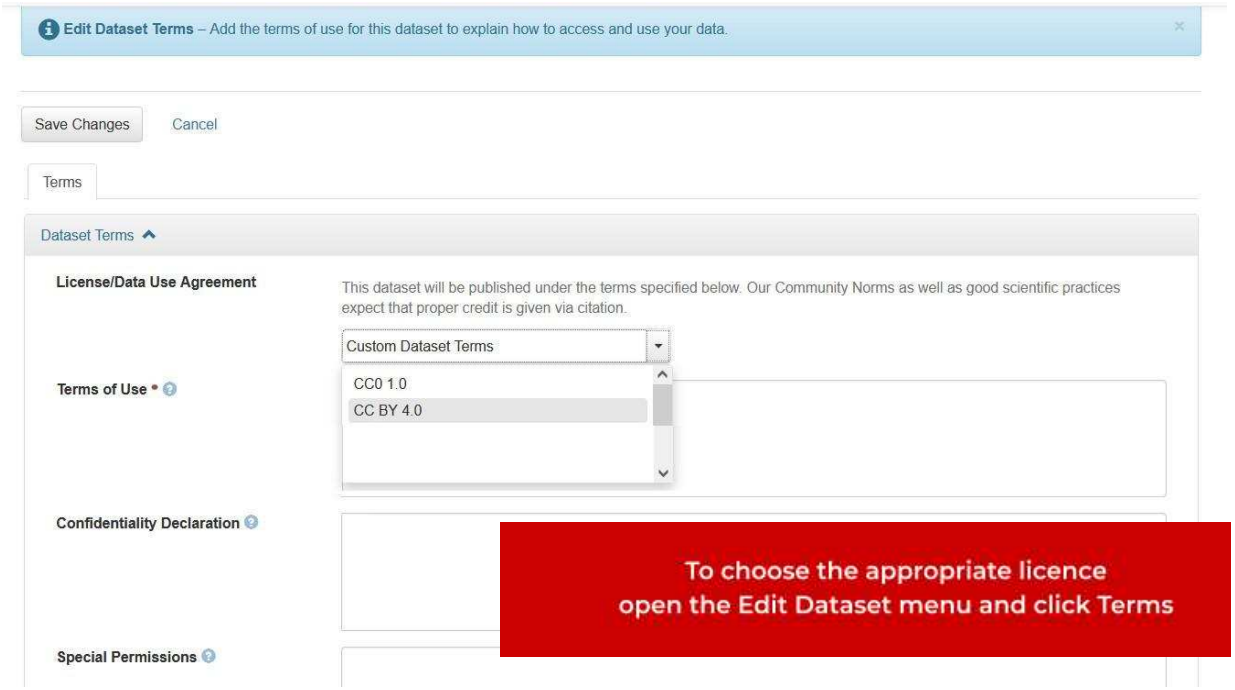
In line with FAIR principles, the recommended data formats for depositing on Dataverse are:

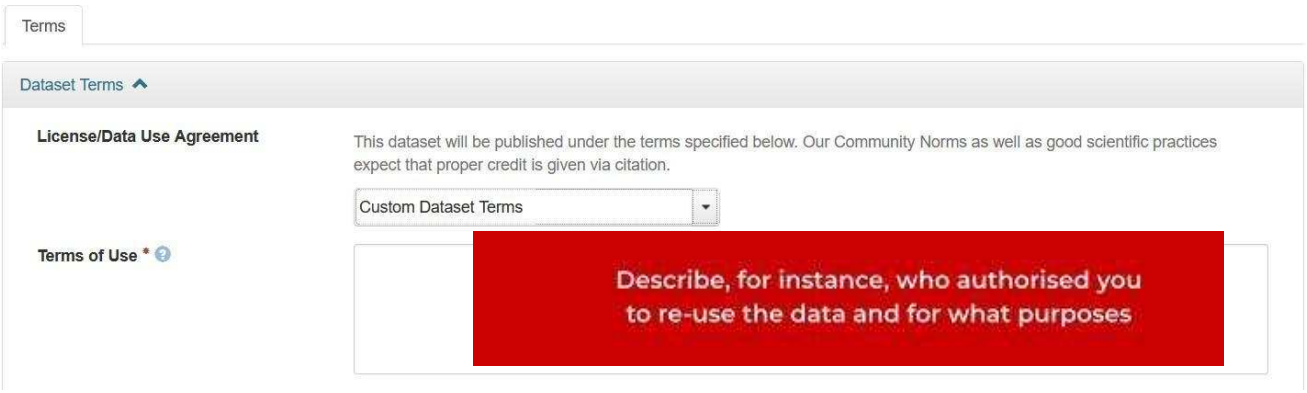
TYPE OF DATA	APPROPRIATE	ACCEPTABLE	DEPRECATED
Tabular (extensive metadata)	csv – hdf5	txt – html – tex – fastq – por	
Tabular (minimal metadata)	csv – tab – ods – sql – tsv	xml (if appropriate dtd) – xlsx	xls – xlsb
Textual / Presentation	txt – pdf – odt – odm – tex – md – htm – xml – extxyz – odf	pptx – rtf – docx – pdf (with embedded forms) – eps – ipf	doc – ppt – dvi – ps
Code / Computation	m – r – py – iypnb – rstudio – rmd – NetCDF – aiml	sdd	mat – rdata
Image / Spectroscopy	tif – png – svg – jpeg – fits	jcamp – jpg – jp2 – tif – tiff – pdf – gif – bmp – dm3 – oir – lsm	indd – ait – psd – spc
Audio	flac – wav – ogg – mxl – midi – mei – humdrum	mp3 – aif	
Video	mp4 – mj2 – avi – mkv	ogm – mp4 – WebM	wmv – mov – qt
Geospatial	NetCDF – tabular gis attribute data – shp – shx – dbf – prj – sbx – sbn – postGis – tif – tfw – geoJson	mdb – mif	
3D structures & images	x3d – x3dv – x3db – pdf3D – pov – pdbml	dwg – dxf – pdb	pxp
Other	xml – json – rdf		

6. DATA: ACCESSIBILITY AND LICENCES FOR USE

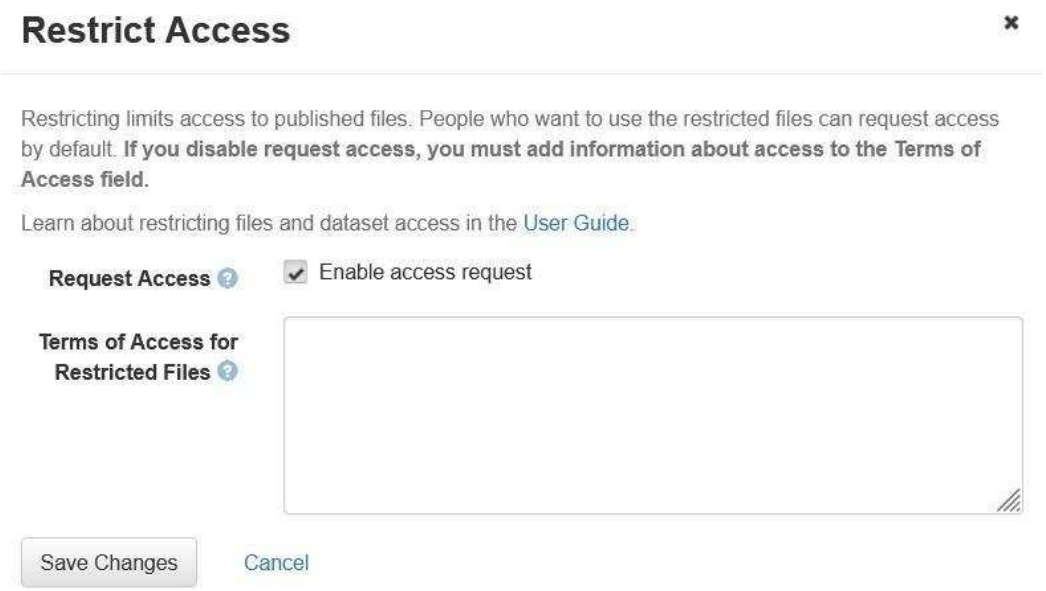
According to the principles of Open Science (see Research Data Management Policy), research results should be made accessible under a free-use licence that guarantees the transparency of uses and verifiability of sources. In this sense, where applicable, it is recommended to use **Creative Commons** licenses, specifically the **CC-BY license for data** and the **CC0 (public domain) license for metadata**.

As an administrator, you can establish terms of use for each dataset.
You can customise the terms of use, e.g. to protect data granted by third parties, by choosing 'Custom Dataset Terms' in the field 'License/Data Use Agreement'.





If needed, you can restrict access to specific files in the dataset.



7. CONTACT

The Open Science Working Group can be contacted by e-mail at this address: openscience@unive.it

Annex: CHECKLIST

Given the above considerations, please review the following checklist before creating a new dataset:

- Ensure that you own the data or have obtained permission from relevant third parties
- Verify data integrity and accessibility (uncorrupted data, non-proprietary formats, etc.)
- Associate the dataset with accurate descriptions and metadata to facilitate reuse and understanding and document how the data was acquired
- Specify the conditions for re-use under the appropriate licence (CC or other...)
- Assess the benefits of depositing the dataset, including its significance, uniqueness, acquisition costs, links to other datasets, and potential for re-use.
- Verify compliance with data protection laws, including anonymisation and restricted access measures.
- Consider all ethical aspects. Scientific ethics encourages transparency, sharing, and data publication, but it is essential to consider potential misuse of the data.

(this checklist was inspired by the EPFL walkthrough)