

Datarepository Unive

LINEE GUIDA



1. COSA È IL REPOSITORY
2. CICLO DI VITA DEI DATI DELLA RICERCA
3. STRUTTURA DEL DATAREPOSITORY UNIVE
4. COME DEPOSITARE
5. DATI: TIPOLOGIE E FORMATI
6. DATI: ACCESSIBILITÀ E LICENZE D'USO
7. CONTATTI

Appendice: CHECK LIST

Prima di leggere queste linee guida, tieni sempre presente che
PRESERVARE I DATI NON SIGNIFICA NECESSARIAMENTE PUBBLICARLI

1. COSA È IL REPOSITORY

Datarepository Unive è il repository istituzionale che l'Università Ca' Foscari ha adottato per la gestione dei dati della ricerca.

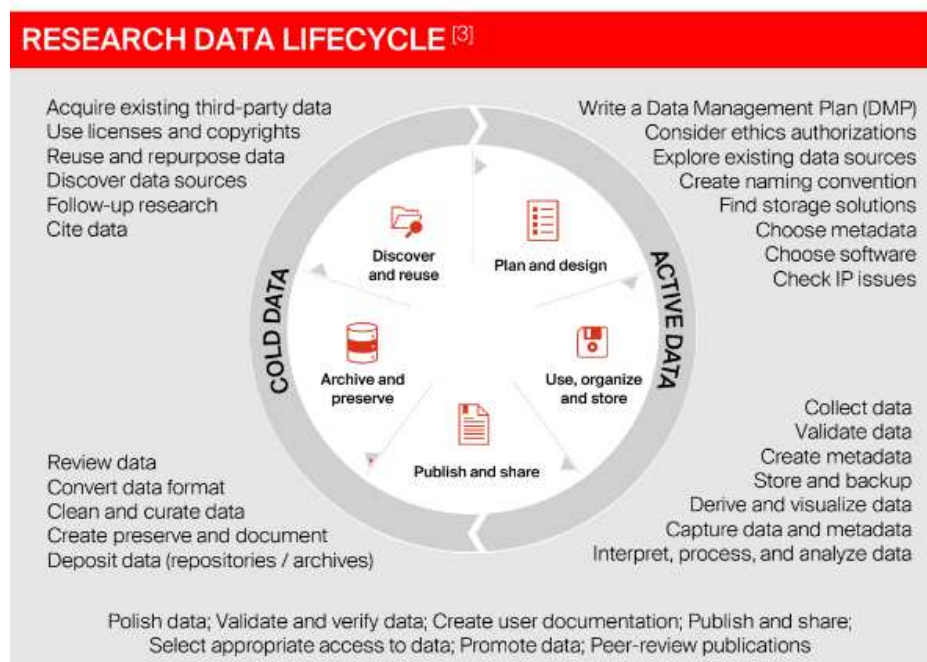
Come ricercatrice/ricercatore hai la necessità di preservare i dati creati o raccolti dalla tua ricerca garantendone sicurezza e integrità. Datarepository Unive è la piattaforma che il nostro Ateneo ti mette a disposizione.

Preservare i dati non significa necessariamente pubblicarli e renderli aperti: anche se non puoi condividere i tuoi dati, è opportuno depositarli in Datarepository, per garantire loro un'archiviazione sicura.

Datarepository Unive non prevede invece il deposito delle pubblicazioni (che vanno archiviate in [ARCA](#)).

N.B.: Datarepository Unive non è un semplice archivio di back-up per mettere in sicurezza i dati, ma una piattaforma che risponde ai criteri individuati dalla Confederation of Open Access Repositories - [COAR](#) e accolti anche dalla Commissione Europea per la definizione di "trusted repository" (cfr. [Annotated Model Grant Agreement](#)).

2. CICLO DI VITA DEI DATI DELLA RICERCA



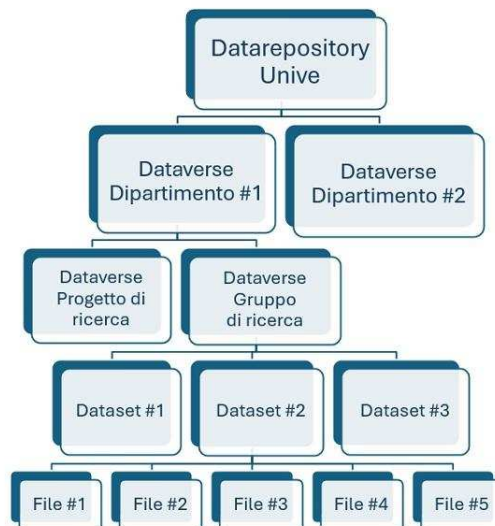
(fonte: <https://zenodo.org/records/7551315>)

Se sei responsabile di un progetto di ricerca, durante le fasi iniziali e di analisi, devi assicurare l'archiviazione, l'integrità e la sicurezza dei dati (**active data**) utilizzando un sistema di storage e backup degli oggetti digitali. Nelle fasi finali di revisione e affinamento dei dati elaborati considera di trasferire i dati utili a illustrare la tua ricerca (**cold data**) nel Datarepository Unive.

3. STRUTTURA DEL DATAREPOSITORY UNIVE

Datarepository Unive ha una struttura gerarchica e può essere definito come un grande contenitore che ha al suo interno:

- **8 dataverse dipartimentali** (uno per ognuno degli 8 dipartimenti di Ca' Foscari)
- ogni dataverse dipartimentale può contenere molti dataverse, idealmente uno per ogni progetto di ricerca o gruppo di ricerca
- ogni singolo dataverse è a sua volta un contenitore di uno o più **dataset**
- ogni dataset - dotato di un **DOI univoco** - contiene infine i file con i dati della ricerca.



Un dataverse è un contenitore di dataset che hanno in comune il progetto di ricerca, oppure l'ambito disciplinare nel caso di dataset creati, ad esempio, per motivi didattici. Ogni dataset è dotato di un DOI univoco e contiene a sua volta i file con i dati della ricerca.

Se sei l'**admin** di un dataverse, ovvero sei responsabile della collezione di dataset caricati al suo interno, puoi assegnare ruoli e permessi agli utenti di quella raccolta, stabilendo chi può contribuire al caricamento dei dataset e con quale ruolo:

- contributor
- curator
- dataset creator
- dataverse+dataset creator (si sconsiglia l'utilizzo di questo ruolo)
- file downloader
- member

Roles [^](#) All the roles set up in your dataverse, that you can assign to users and groups.

Admin - A person who has all permissions for dataverses, datasets, and files, including approving requests for restricted data.
AddDataverse AddDataset ViewUnpublishedDataverse ViewUnpublishedDataset DownloadFile EditDataverse EditDataset ManageDataversePermissions ManageDatasetPermissions ManageFilePermissions PublishDataverse PublishDataset DeleteDataverse DeleteDatasetDraft

Contributor - For datasets, a person who can edit License + Terms, and then submit them for review.
ViewUnpublishedDataset DownloadFile EditDataset DeleteDatasetDraft

Curator - For datasets, a person who can edit License + Terms, edit Permissions, and publish datasets.
AddDataverse AddDataset ViewUnpublishedDataverse ViewUnpublishedDataset DownloadFile EditDataset ManageDatasetPermissions ManageFilePermissions PublishDataset DeleteDatasetDraft

Dataset Creator - A person who can add datasets within a dataverse.
AddDataset

Dataverse + Dataset Creator - A person who can add subdataverses and datasets within a dataverse.
AddDataverse AddDataset

Dataverse Creator - A person who can add subdataverses within a dataverse.
AddDataverse

File Downloader - A person who can download a published file.
DownloadFile

Member - A person who can view both unpublished dataverses and datasets.
ViewUnpublishedDataverse ViewUnpublishedDataset DownloadFile

Per gestire i ruoli / user abilitati a collaborare al tuo dataverse apri il menu Edit e clicca Permissions

4. COME DEPOSITARE

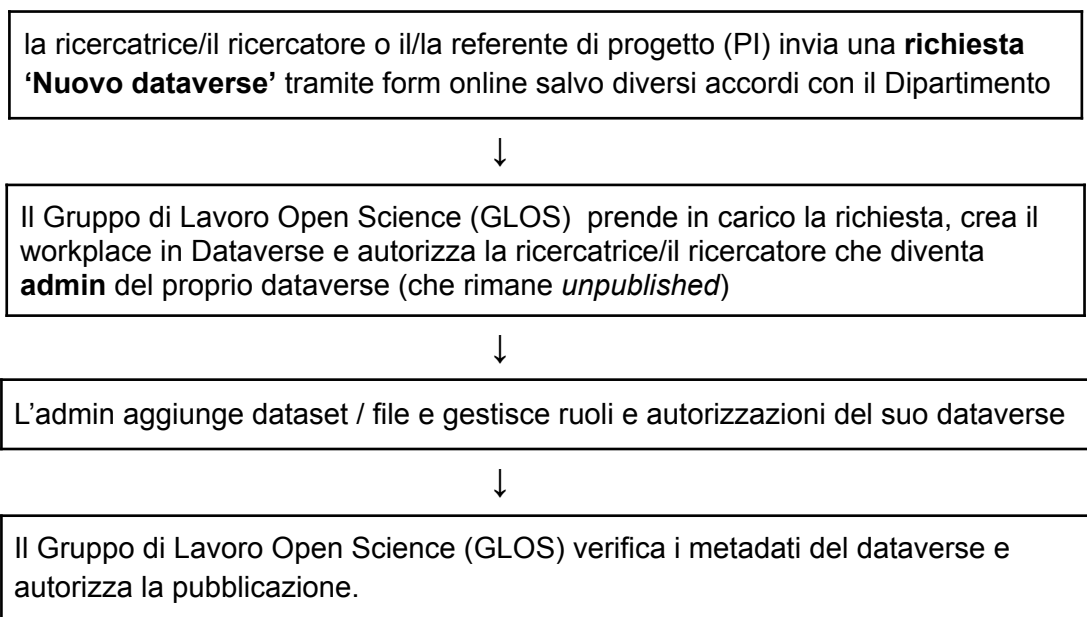
Per depositare i dati invia una [richiesta 'Nuovo dataverse'](#) tramite form online a Gruppo di Lavoro Open Science (GLOS)* che prende in carico la richiesta, crea il workplace in Dataverse e autorizza la ricercatrice/ricercatore che diventa **amministratrice/amministratore** del proprio dataverse (unpublished), salvo diversi accordi con il Dipartimento.

Per accedere sarà necessario fare il login con le credenziali d'Ateneo a questo link: <https://datarepository.unive.it>.

L'amministratrice/amministratore del dataverse potrà gestire - se previsti - i ruoli di altri collaboratori e gestire il caricamento di uno o più dataset corredandoli di metadati e licenze.

Il Gruppo di lavoro su Open Science (GLOS) verifica la conformità di metadati e licenze prima di autorizzare la pubblicazione.

Workflow



(*) Il Gruppo di Lavoro Open Science (GLOS) accompagnerà questo periodo di implementazione fino alla strutturazione del servizio Open Science SBA-BDA.

Prima di depositare controlla la [check list](#).

5. DATI: TIPOLOGIE E FORMATI

Se decidi di usare la piattaforma, valuta i tipi di dati che intendi conservare e rifletti sul tipo di formato, accesso e finalità ultima che vuoi i tuoi dati assumano.

| | |
|---|-------------------------------|
| Dati grezzi / Raw data dati originali raccolti durante il processo di ricerca, ma non ancora elaborati | non sempre appropriati |
| Dati elaborati / FAIR: dati derivati dalla pulizia, anonimizzazione e controllo dei dati originali | appropriati |
| Dati analizzati derivano dall'analisi e/o dall'integrazione dei dati elaborati, e; assumono la forma di rappresentazione tabellari, testuali, grafiche per facilitare la comprensione e comunicazione | appropriati |

DO NOT CONFUSE

N.B. è possibile che i dati di partenza della tua ricerca siano dati di riuso ovvero dati grezzi o elaborati provenienti da **terze parti**. In questo caso assicurati di avere il consenso del proprietario ed eventualmente valuta la necessità di rimandare al dataset originario se già disponibile nell'ecosistema Open Science.

| | Tipologia | Descrizione | Esempi |
|--|---------------------------------------|--|---|
| | Dati osservazionali | Dati acquisiti in tempo reale in situ e che non possono essere recuperati, ricreati o sostituiti | <i>Rilevazioni di sensori; osservazioni sensoriali (umane); risultati di sondaggi; appunti / trascrizioni di interviste</i> |
| | Dati sperimentali | Dati raccolti in condizioni controllate, in situ o in laboratorio; sono dati riproducibili, anche se a costi elevati | <i>Sequenziamento genico; cromatogrammi; spettroscopia; microscopia</i> |
| | Dati di simulazione | Dati generati imitando il funzionamento di un processo o di un sistema del mondo reale utilizzando modelli di test computerizzati | <i>Modelli climatici; modelli economici; modelli biogeochimici</i> |
| | Dati derivati / compilati | Dati generati dall'utilizzo di dati già esistenti, spesso provenienti da origini diverse | <i>Variabili derivate; banca dati compilata; modelli 3D</i> |
| | Dati di riferimento / canonici | Set di dati derivanti da raccolte statiche o organiche [peer-reviewed], generalmente pubblicati | <i>Banche dati di sequenze genetiche; strutture chimiche; dati di censimenti; portali di dati spaziali</i> |
| | Metadati | Informazioni strutturate associate ai dati funzionali alla scoperta, descrizione, utilizzo, gestione e preservazione dei dati stessi | <i>README files ; keywords; nomi di file e cartelle</i> |

In linea coi principi FAIR, per il deposito su Dataverse i formati di dato raccomandati sono:

| 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 - ism | 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. DATI: ACCESSIBILITÀ E LICENZE D'USO

Secondo i principi dell'Open Science (vedi la [Policy per la gestione dei dati della ricerca](#)) i risultati della ricerca vanno resi accessibili con una licenza di libero utilizzo che garantisca la tracciabilità degli usi e il riconoscimento della fonte originaria. In tal senso, dove applicabile, si raccomanda l'adozione delle licenze **Creative Commons** e in particolare della licenza **CC-BY per i dati** e della licenza **CC0 (pubblico dominio) per i metadati**.

Come admin devi stabilire i termini di utilizzo di ciascun dataset.

Puoi personalizzare i termini d'uso, per es. per proteggere dati concessi da terze parti scegliendo nel campo "License/Data Use Agreement" la voce "Custom Dataset Terms"

i **Edit Dataset Terms** – Add the terms of use for this dataset to explain how to access and use your data. x

Save Changes Cancel

Terms

Dataset Terms ^

License/Data Use Agreement

This dataset will be published under the terms specified below. Our Community Norms as well as good scientific practices expect that proper credit is given via citation.

Terms of Use ?

Custom Dataset Terms

CC0 1.0

CC BY 4.0

Confidentiality Declaration ?

Special Permissions ?


Per scegliere la licenza appropriata, apri il menu Edit Dataset e clicca Terms

Terms

Dataset Terms ^

License/Data Use Agreement This dataset will be published under the terms specified below. Our Community Norms as well as good scientific practices expect that proper credit is given via citation.

Custom Dataset Terms

Terms of Use * 

Descrivi, per esempio, da chi sei stato autorizzato a riutilizzare i dati e per quali finalità

In caso tu abbia necessità di riservare l'accesso ad alcuni specifici file presenti nel dataset, puoi scegliere di riservare l'accesso.

Restrict Access

Restricting limits access to published files. People who want to use the restricted files can request access by default. **If you disable request access, you must add information about access to the Terms of Access field.**

Learn about restricting files and dataset access in the [User Guide](#).

Request Access  Enable access request

Terms of Access for Restricted Files 

Save Changes

Cancel

7. CONTATTI

Il Gruppo di Lavoro Open Science (GLOS) può essere contattato via mail a questo indirizzo:

openscience@unive.it

Appendice : CHECKLIST

Prima di creare un nuovo dataset considera questa check-list:

- sei la proprietaria / il proprietario dei dati (o hai il permesso delle terze parti coinvolte)
- hai assicurato l'integrità e accessibilità ai dati (dati non corrotti, formati non proprietari ecc...)
- hai associato al tuo dataset le corrette descrizioni per garantire il riuso e comprensione (metadati) nonché le modalità di acquisizione dei dati
- hai specificato le condizioni di riuso con una licenza adeguata (CC o altre...)
- hai valutato i vantaggi del deposito di quello specifico dataset (significatività, unicità/singolarità, costi di acquisizione dei dati, collegamenti con altri dataset, potenziale riuso ecc)
- hai adottato le corrette misure previste dalla normativa sulla protezione dei dati (anonimizzazione, accesso ristretto ecc...)
- hai considerato gli aspetti etici. Se da un lato l'etica scientifica incoraggia la trasparenza dei dati, la condivisione e la pubblicazione, è necessario interrogarsi se esiste la possibilità di un uso distorto o improprio dei dati

(checklist ispirata a [EPFL walkthrough](#))