



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



Linee Guida di Ateneo per la Gestione dei Dati della Ricerca

Linee guida a cura di:

Alma Mater Studiorum – Università di Bologna

ARIC – Area Ricerca, Settore Coordinamento Servizi Ricerca e Progetti di Area,
Data Steward

Con il contributo di:

ARIN – Area Innovazione, Settore Knowledge Transfer Office

ARPAC – Area Patrimonio Culturale, Settore Gestione e sviluppo della biblioteca
digitale d'Ateneo – AlmaDL; Settore Gestione e sviluppo della Biblioteca
delle risorse elettroniche – AlmaRE

GLOS – Gruppo di Lavoro Open Science

Referenti Open Science dei Dipartimenti

SSRD – Staff Rettore e Direttore Staff Rettore e Direttore Generale,
Unità Professionale “Protezione dei dati personali”

APPC – Area Pianificazione, Programmazione e Comunicazione
- Settore Comunicazione - Ufficio Graphic design per la comunicazione

Queste linee guida sono pubblicate con licenza Creative Commons Attribution 4.0
International: <https://creativecommons.org/licenses/by/4.0/>



Queste linee guida accompagnano la *Policy di Ateneo per la Gestione dei Dati della Ricerca*, che stabilisce criteri e principi da seguire per trattare i dati in modo corretto e consapevole, in linea con gli standard internazionali e le peculiarità proprie dell'ambito disciplinare.

I destinatari di questo documento sono **tutti i ricercatori e le ricercatrici dell'Università di Bologna**, a qualsiasi stadio della propria carriera e in qualsiasi disciplina.

Gestire i dati della ricerca significa **curarli e organizzarli in modo consapevole durante l'intero ciclo di ricerca** con l'obiettivo di:

- rendere il processo di ricerca il più efficiente possibile;
- rendere i dati stessi interpretabili, comprensibili e rintracciabili nel tempo;
- favorire l'integrità della ricerca;
- stimolare la collaborazione con altri ricercatori.

La gestione dei dati va **pianificata attentamente** all'inizio della ricerca, accompagna tutte le fasi operative di **produzione, raccolta e analisi dati** e si conclude con la preservazione (archiviazione a lungo termine dei dati stessi) e preferibilmente la **condivisione degli stessi**.



In questa guida, vedremo a uno a uno gli aspetti più rilevanti di questo processo, fornendo indicazioni procedurali e strumenti utili ad affrontare, in modo puntuale, ciascuna delle azioni necessarie ad una corretta gestione dei dati della ricerca.

Maggiori dettagli sugli argomenti presentati in queste linee guida si possono trovare nelle schede di approfondimento indipendenti che le accompagnano. I riferimenti alle schede di approfondimento sono presentati nel testo come segue **Nome scheda**.

PIANIFICAZIONE DEL FLUSSO DATI	5
Principali azioni di questa fase	6
Identificare i tipi di dati	7
Identificare i metadati fondamentali	8
Pianificare l'organizzazione dei dati in dataset	9
Redigere un Data Management Plan	10
PRODUZIONE, RACCOLTA E ANALISI	11
Principali azioni di questa fase	12
Conservare i dati in appositi spazi di storage	13
Assicurare la qualità dei dati	14
Raccogliere la documentazione	15
PRESERVAZIONE E CONDIVISIONE	16
Principali azioni di questa fase	17
Valutare quali dati è necessario depositare	18
Scegliere il repository più appropriato	19
Depositare i dati secondo i principi FAIR	20
Associare ai dati una licenza	21

Link utili

Policy di Ateneo sulla Gestione dei Dati della Ricerca:

<https://www.unibo.it/it/ateneo/chi-siamo/open-access-e-open-science>

Rassegne video dell'Università di Bologna:

- “Dati: conoscerli e gestirli per valorizzare la ricerca”
<https://www.youtube.com/playlist?list=PLaUmBQ7P5K-AyDDnv1f8upAyEOtAF2gj3>
- “Open Access e Open Science”
https://www.youtube.com/playlist?list=PLaUmBQ7P5K-A83TIY96DyUI6t3rCryRK_

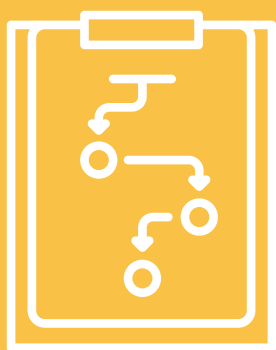
Video “Dati: conoscerli e gestirli per valorizzare la ricerca. Il Research Data Management”

<https://www.youtube.com/watch?v=5WZF00pVSrY&list=PLaUmBQ7P5K-AyDDnv1f8upAyEOtAF2gj3&index=2>

Research Data Management Decision Tree

<https://doi.org/10.5281/zenodo.7190005>

PIANIFICAZIONE DEL FLUSSO DATI






Principali azioni di questa fase

- **Identificare** i tipi di **dati**,
 - **decidendo se generarne di nuovi** e/o **riutilizzare** quelli già disponibili, prodotti da fonti esistenti;
 - avendo consapevolezza dei principi etici e delle norme privacy e di proprietà intellettuale da rispettare.
- **Identificare i metadati** fondamentali – per poter descrivere in modo significativo i dati generati o prestare attenzione ai metadati associati quando si riusano dati esistenti.
- **Pianificare** l'organizzazione strutturata dei dati in **dataset**.
- **Redigere un Data Management Plan** per tenere traccia delle proprie scelte.


Identificare i tipi di dati

I dati della ricerca sono **informazioni**, in qualsiasi formato, utilizzate nell'ambito di una specifica **attività di ricerca** e necessarie per **validarne i risultati**. Esiste un'enorme varietà di dati, e la loro identificazione e classificazione dipendono spesso dal dominio disciplinare di appartenenza  **Dati della ricerca**.



In ogni caso, già all'inizio della tua ricerca:

- **Individua i tipi di dati** con cui lavorerai: per farlo, esamina ogni fase della tua ricerca per valutare **quali tipi di informazioni dovrai raccogliere e/o utilizzare** per rispondere alle tue domande di ricerca.
- Verifica sempre se **esistono dati già pubblicati** da altri ricercatori che possano essere utili per rispondere alle tue domande di ricerca.

Se produci dati nuovi:

- Annota sempre anche i metadati relativi ai dati stessi e al loro processo di creazione [vedi [Identificare i metadati fondamentali](#)].
- Se la ricerca coinvolge persone vulnerabili, animali, o tecnologie particolari (intelligenza artificiale, possibili usi militari, ...), assicurati di gestire i dati in modo etico e di possedere le necessarie autorizzazioni per la loro raccolta, contattando gli appositi comitati etici quando opportuno.
- Nel caso i tuoi dati includano dati personali, assicurati di prendere le precauzioni necessarie a gestirli  **Il rispetto della privacy**.
- Valuta se esistono altre ragioni per tenere i tuoi dati riservati, ad esempio la valorizzazione commerciale di un risultato collegato o vincoli assunti nei confronti di soggetti terzi.

Se riutilizzi dati esistenti:

- Controlla bene i metadati loro associati, che contengono informazioni fondamentali per il loro corretto riutilizzo [vedi [Identificare i metadati fondamentali](#)].
- Attraverso i metadati scopri anche se puoi riutilizzare i dati per i tuoi scopi, o se ci sono limitazioni imposte dagli autori  **Diritto d'autore**.
- Assicurati sempre di riutilizzare i dati nel rispetto dei principi etici e delle norme privacy  **Il rispetto della privacy** e di proprietà intellettuale.


Link utili



"Data steward all'Università di Bologna" <https://youtu.be/6lc0isyefs8?si=JqCFzHlh2Hz1U2ee>

Materiali di approfondimento sui dati:

- The Turing Way Guide for Reproducible Research <https://the-turing-way.netlify.app/reproducible-research/rdm/rdm-data>
- Research Data Alliance <https://www.rd-alliance.org/>
- Digital Curation Centre (DCC) <https://www.dcc.ac.uk/>

Identificare i metadati fondamentali

I metadati sono **informazioni strutturate** che accompagnano i dati della ricerca [vedi [Identificare i tipi di dati](#)]. Questi “dati sui dati” servono a facilitarne la comprensione e il riutilizzo, a favorire l’identificazione e l’indicizzazione dei dati da parte di motori di ricerca e portali di aggregazione  **Metadati e documentazione**.

I metadati sono in genere strutturati secondo **schemi standard**, in molti casi specifici per **ambito disciplinare**, consolidati a livello delle comunità di ricerca e implementati dalle infrastrutture di archiviazione e accesso a lungo termine  **I repository**. Uno schema di metadati che adotta un vocabolario controllato  **Utilizzare diversi tipi di standard** garantisce l’intelligibilità e la comprensione delle informazioni.

Già all’inizio della tua ricerca:

- Individua il prima possibile lo **schema di metadati** più adatto alla tua ricerca. Ricorda che questa scelta può essere condizionata dal repository dove archiverai a lungo termine i tuoi dataset.
- Identifica **le informazioni** richieste dallo schema di metadati che utilizzerai e assicurati di raccoglierle durante tutto il corso della ricerca.
- Dove possibile, ricordati di utilizzare **vocabolari controllati**.
- Al momento del deposito nel repository, non dimenticare di inserire tra i metadati queste informazioni fondamentali: **i nomi e le affiliazioni** (i.e., Università di Bologna) di tutti coloro che hanno collaborato alla creazione del dataset (per documentare adeguatamente l’autorialità), **il Persistent Identifier**, spesso abbreviato in PID (i.e., il Digital Object Identifier, DOI) e **il tipo di licenza** associati al dataset (vedi anche [Depositare i dati secondo i principi FAIR](#)).
- Ricorda che tutte le informazioni necessarie agli altri per comprendere i tuoi dati, se non possono essere strutturate nei metadati, devono essere associate ai dataset sotto forma di **documentazione** aggiuntiva [vedi [Raccogliere la documentazione](#)].


Link utili

Videopillola “5 Minute Metadata- What is metadata?” <https://www.youtube.com/watch?v=L0vOg18ncWE>

Risorse per la ricerca di schemi di metadati:

- FAIR Sharing standard registry <https://fairsharing.org/search?fairsharingRegistry=Standard>
- RDA metadata standards directory <https://rd-alliance.github.io/metadata-directory/standards/>
- DCC guidance on disciplinary metadata <https://www.dcc.ac.uk/guidance/standards/metadata>

Pianificare l'organizzazione dei dati in dataset

Un dataset è un **insieme strutturato di dati** (vedi [Identificare i tipi di dati](#)) in **relazione** tra loro, in qualsiasi formato, creati e/o raccolti con uno scopo comune (i.e., rispondere alla stessa domanda di ricerca) e **organizzati** per riflettere i risultati di un'attività di ricerca  **Approfondimento sui dataset**. L'attenzione e la cura nella gestione dei dataset sono cruciali per garantire la qualità e l'utilità dei risultati della ricerca. Un dataset ben organizzato e gestito correttamente deve essere sempre accompagnato da metadati informativi (vedi [Identificare i metadati fondamentali](#)).

Già all'inizio della tua ricerca:


- **Organizza** ed eventualmente suddividi i tuoi dati in modo strutturato.
- Adotta una **chiara nomenclatura** per file e cartelle che comporranno il tuo dataset, così da rendere chiari i contenuti e le relazioni tra elementi.
- Inizia a **raccogliere la documentazione** necessaria per rendere il dataset il più comprensibile e riutilizzabile possibile (vedi [Raccogliere la documentazione](#)).
- Comincia a riflettere su quale potrebbe essere il repository più adatto per **archiviare il tuo dataset** a lungo termine (vedi [Scegliere il repository più appropriato](#)). Il contenuto e le dimensioni del dataset potrebbero limitarti nella tua scelta.



Redigere un Data Management Plan

Spesso abbreviato in DMP, il Data Management Plan (o piano di gestione dei dati) è lo strumento fondamentale per **documentare tutte le scelte relative alla gestione dei dati** del progetto. Solitamente consiste in un documento di testo, ed è una buona pratica che accompagni e testimoni la gestione dei dati **fin dalle fasi iniziali** del processo. Molti enti finanziatori richiedono un DMP da contratto.

- Inizia a scrivere il DMP appena inizi a ragionare su come gestirai i tuoi dati.
- Ricorda che la gestione dei dati **evolve durante il corso della ricerca** e che le valutazioni fatte inizialmente potrebbero non essere definitive.
- Ricorda che, per essere davvero utile, il DMP dovrà **essere tenuto aggiornato**.

All'interno del documento:

- Fornisci una **panoramica dettagliata dei tuoi dati di ricerca**, sia nuovi che riutilizzati (vedi [Identificare i tipi di dati](#)).
- Specifica **metodologie, strumenti e software** utilizzati per raccogliere, creare o analizzare i tuoi dati  **Gestire il software**.

- Indica le strategie che utilizzerai per assicurare la **qualità dei dati** ed evitare possibili imprecisioni o incoerenze (vedi [Assicurare la qualità dei dati](#)).
- Riporta le **strategie di conservazione sicura** dei tuoi dati, ad esempio in termini di condivisione con i tuoi collaboratori o di creazione di copie di backup (vedi [Conservare i dati in appositi spazi di storage](#)).
- Descrivi le **strategie di archiviazione a lungo termine** (vedi [Valutare quali dati è necessario depositare](#) e [Scegliere il repository più appropriato](#)) e come hai applicato i **principi FAIR** ai tuoi dati (vedi [Depositare i dati secondo i principi FAIR](#)).
- Descrivi i **ruoli e le responsabilità** all'interno del team di ricerca.
- Documenta i **costi**, anche in termini di tempo, associati alla gestione dei tuoi dati.
- Tratta gli aspetti di gestione dei dati legati alla **privacy**  **Il rispetto della privacy**, ai **diritti di proprietà intellettuale**  **Diritto d'autore** e all'**etica**.

Link utili

Video "Dati conoscerli e gestirli per valorizzare la ricerca. Il Data Management Plan"

<https://www.youtube.com/watch?v=SIOsrQdrhtQ>

Video "Dati della ricerca: Il Data Management Plan" <https://www.youtube.com/watch?v=QnHzMpib2Hc>

Linee guida per compilare un DMP: Science Europe templates e linee guida

<https://scienceeurope.org/our-priorities/research-data/research-data-management/>

Strumenti per la compilazione online del DMP:

- Elixir Data Stewardship Wizard <https://ds-wizard.org/>
- DCC DMPonline <https://dmponline.dcc.ac.uk/>
- ARGOS <https://argos.openaire.eu/home>

Template di Data Management Plan:

- Horizon Europe Data Management Plan Template https://www.openaire.eu/images/Guides/HORIZON_EUROPE_Data-Management-Plan-Template.pdf
- Science Europe Data Management Plan Template <https://scienceeurope.org/media/411km040/se-rdm-template-3-researcher-guidance-for-data-management-plans.docx>

PRODUZIONE, RACCOLTA E ANALISI





Principali azioni di questa fase

- **Conservare** i dati in appositi **spazi di storage**, provvedendo ai necessari **backup**.
- **Assicurare** la **qualità** dei dati attraverso dei processi metodici e
 - **tenendo traccia** delle diverse **versioni** dei files;
 - **organizzando file e cartelle** in modo gerarchico, assegnando loro nomi coerenti;
 - **scegliendo i formati** più adatti per i dati, possibilmente standard e aperti, per favorire interoperabilità e riusabilità.
- **Raccogliere** come **documentazione** tutte le informazioni necessarie a spiegare e comprendere i dati.

Conservare i dati in appositi spazi di storage

Con **storage** intendiamo la conservazione dei dati (vedi [Identificare i tipi di dati](#)) a **breve o medio termine**, durante le fasi attive del processo di ricerca. Quando parliamo di **sistemi di storage** ci riferiamo ad esempio ad hard drive esterni, servizi cloud, oppure server.

Durante la tua ricerca:

- Pianifica precocemente in che modo conserverai i tuoi dati, decidendo **dove salvarli** e **prevenendo eventuali spese** da sostenere.
- Ricorda che la scelta del sistema di storage più adatta dipende dalla **natura dei dati, dal loro volume** e dalla frequenza con cui persone diverse si trovano a **collaborare** sugli stessi file.
- Crea periodicamente delle **copie di backup** dei dati, su supporti diversi, per evitare di perderli.
- Assicurati di proteggere i dati, utilizzando delle **password efficaci** e aggiornando periodicamente gli **antivirus** del computer.
- Ricorda che la **gestione e conservazione dei dati personali** richiede un livello ulteriore di protezione, per cui puoi utilizzare una password specifica o ricorrere a programmi che criptano file e cartelle 📄 **Il rispetto della privacy**.
- Se devi condividere dati col tuo gruppo di ricerca, utilizza piattaforme di storage che permettano l'accesso da remoto, stabilendo da subito i **diritti di accesso di ognuno a file e cartelle**.
- Assicurati che l'infrastruttura di storage che hai scelto ti permetta di verificare quali **modifiche** sono state apportate ai dati, da chi sono state apportate e di poter eventualmente recuperare le **versioni precedenti**. Questi aspetti sono cruciali in un contesto di ricerca collaborativa.

Link utili

Videopillola "Dati conoscerli e gestirli per valorizzare la ricerca. Salvare e condividere i dati"

<https://youtu.be/VQOyK0tQ1N4?feature=shared>.

Materiali di approfondimento sulla scelta del sistema di storage:

- The Research Data Management toolkit <https://rdmkit.elixir-europe.org/storage#what-features-do-you-need-in-a-storage-solution-when-collecting-data>
- The Turing Way Guide for Reproducible Research <https://the-turing-way.netlify.app/reproducible-research/rdm/rdm-storage>

Strumenti per la criptazione di file e cartelle:

- Veracrypt <https://www.veracrypt.fr/en/Home.html>
- BitLocker <https://docs.microsoft.com/it-it/windows/security/information-protection/bitlocker/bitlocker-overview>

Strumenti per la valutazione dei costi del data management:

- OpenAIRE RDM cost calculator <https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs>
- UK data service RDM cost calculator <https://ukdataservice.ac.uk/app/uploads/costingtool.pdf>

Assicurare la qualità dei dati

Al di là delle peculiarità di ogni disciplina, a un livello basilare la qualità dei dati (vedi [Identificare i tipi di dati](#)) è assicurata attraverso un **insieme di processi metodici** che ne permettono la rintracciabilità, l'uso corretto e il riuso da parte di altri.

Durante la tua ricerca:

- Scegli un'**organizzazione chiara delle cartelle e dei file** e tieni traccia delle loro diverse versioni.
- Scegli uno **standard per la nomenclatura** dei file che sia chiaro e leggibile. Ad esempio, è bene riportare il nome dell'autore o la provenienza dei dati, la data di creazione e il numero di versione, evitando spazi o caratteri speciali (es. FocusGroup1_20240502_v2.rtf).
- Scegli i **formati più adatti** per i dati, possibilmente standard e aperti, per favorire interoperabilità e riusabilità 📄 **Dati della ricerca: tipologie, formati, metodi.**
- **Valida e verifica i tuoi dati** per evitare che siano imprecisi, incompleti o incoerenti. Alcune strategie comprendono la convalida dell'immissione dei dati, il controllo dell'intervallo di dati, la rimozione/registrazione di variabili imprecise o mancanti, il controllo di scale di dati coerenti.
- Definisci una **metodologia e dei processi standard** per l'analisi e l'elaborazione dei dati, specialmente se fai ricerca in contesti collaborativi. Ad esempio, puoi definire quali dati salvare in quali cartelle, in quale fase documentare i dati, o con quali sistemi condividerli con i tuoi collaboratori (vedi [Conservare i dati in appositi spazi di storage](#)).
- Se i tuoi dati sono collegati ad altri risultati della ricerca, descrivi nella documentazione come sono collegati e fornisci la **citazione completa del materiale collegato** (es. un dataset derivato da un altro dataset già esistente) (vedi [Raccogliere la documentazione](#)).

Link utili

Materiali di approfondimento sulle strategie di controllo qualità dei dati:

- The Turing Way Guide for Reproducible Research <https://the-turing-way.netlify.app/reproducible-research/rdm/rdm-data-curation>
- The Research Data Management toolkit https://rdmkit.elixir-europe.org/data_quality

Strumenti per la definizione della nomenclatura:

- Bulk Rename Utility (Free File Renaming Utility for Windows) <https://www.bulkrenameutility.co.uk/>
- File naming conventions https://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf


Strumenti per la standardizzazione e condivisione delle metodologie:

- Protocol manager <https://protocols.io>

Raccogliere la documentazione

I dati e i dataset (vedi [Pianificare l'organizzazione dei dati in dataset](#)), per essere **intelligibili e interpretabili da persone diverse** da coloro che li hanno generati, devono essere accompagnati da documentazione a corollario. Un esempio di documentazione è il README file, un documento testuale, leggibile e interpretabile dagli individui ("human readable") che **spiega i dati e la loro organizzazione** all'interno del dataset.

Durante la tua ricerca:

- Cura la scrittura della documentazione durante tutte le fasi attive della raccolta e dell'analisi dei dati.
- Documenta i **dati** che compongono il dataset, le **relazioni** che li legano, la loro **provenienza**.
- Inserisci informazioni esaustive circa le **metodologie** (declinate sia in termini di protocolli che di specifiche tecniche e di eventuali strumenti utilizzati) **che sono state applicate** per la raccolta e/o riutilizzo e/o generazione dei dati.
- Documenta i **processi di garanzia della qualità** nella generazione/analisi dati (vedi [Assicurare la qualità dei dati](#)).
- Inserisci informazioni su eventuali **strumentazioni o software necessari** per aprire, leggere o interpretare i dati stessi  **Gestire il software**.
- **Archivia** la documentazione **insieme ai dati** nel momento in cui questi vengono depositati in un repository (vedi [Valutare quali dati è necessario depositare](#)).
- Salva il file di documentazione in un formato aperto e accessibile (es .rtf, .md).

Link utili

CESSDA Data Management Expert Guide. Documentation and metadata

<https://dmeg.CESSDA.eu/Data-Management-Expert-Guide/2.-Organise-Document/Documentation-and-metadata>

Utrecht University. Research data management support Guides. Metadata and documentation

<https://www.uu.nl/en/research/research-data-management/guides/during-research/metadata-and-documentation>

PRESERVAZIONE E CONDIVISIONE





Principali azioni di questa fase



- **Valutare quali dati è necessario depositare** per garantire la comprensione, verifica e riproducibilità della ricerca.
- **Scegliere il repository** più appropriato per il deposito dei propri dati, per garantirne la preservazione e la condivisione a lungo termine.
- **Depositare i dati**, organizzati in dataset, secondo **i principi FAIR**.
- **Associare** ai propri dati **una licenza** che ne garantisca, dove possibile, **il più ampio riutilizzo**, qualora non sussistano vincoli con terzi o strategie di valorizzazione che possano limitare questa scelta.

Valutare quali dati è necessario depositare

Il deposito indica l'archiviazione dei dati, **organizzati in dataset** (vedi [Pianificare l'organizzazione dei dati in dataset](#)), in un'infrastruttura digitale pensata per la loro preservazione a lungo termine, chiamata **repository** (vedi [Scegliere il repository più appropriato](#)). Può avvenire al termine di un'attività di ricerca, ma deve comunque precedere la pubblicazione dei risultati in un articolo scientifico.

Il deposito garantisce la **preservazione dei dati** e la loro **visibilità** ben oltre la fine del progetto di ricerca che li ha generati.

Durante la tua ricerca:

- Opera una **selezione** su quali dati depositare per permettere la validazione delle conclusioni e la riproducibilità della ricerca. Ricorda che la **responsabilità del deposito** dei dati è in capo al ricercatore che li genera.
- **Ricorda di depositare** dataset o software originali, dati grezzi ottenuti dall'analisi di campioni fisici, e dati osservativi che non possono essere riprodotti  **Gestire il software**.
- **Non depositare necessariamente** dati facilmente riottenibili, oppure che siano troppo voluminosi rispetto alla loro effettiva utilità.
- **Non depositare** dati che sono già disponibili, ad esempio dati che stai riutilizzando perché sono stati già depositati da qualcun altro.
- **Documenta** accuratamente in ogni caso la **provenienza** dei dati che depositi, insieme alle **metodologie** con cui sono stati prodotti e gestiti (vedi [Raccogliere la documentazione](#)).
- Puoi depositare anche i dati che devono rimanere inaccessibili a terzi per motivi di privacy, etica o proprietà intellettuale  **Diritto d'autore**, avendo cura di scegliere un repository adatto.

Link utili

Video "Dati conoscerli e gestirli per valorizzare la ricerca. Conservare i dati a lungo termine"

https://www.youtube.com/watch?v=J3VyrUzj_E

Materiali di approfondimento sulla preservazione a lungo termine dei dati:

- The Research Data Management toolkit https://rdmkit.elixir-europe.org/data_publication
- Stanford University Library guidelines on data management and sharing <https://laneguides.stanford.edu/DataManagement/>
- Digital Curation Centre, How to Appraise and Select Research Data for Curation <https://www.dcc.ac.uk/guidance/how-guides/appraise-select-data>

Scegliere il repository più appropriato

I repository sono **infrastrutture che garantiscono l'archiviazione a lungo termine** dei dataset (vedi [Pianificare l'organizzazione dei dati in dataset](#)). Possono essere di natura disciplinare, istituzionale o generalista. Possono essere ufficialmente certificati o meno, ma se sono affidabili assegnano un **PID** (come, ad esempio, un DOI) e permettono l'associazione di **metadati** (vedi [Identificare i metadati fondamentali](#)) e licenze [Diritto d'autore](#) a ciascun dataset caricato (vedi anche [Depositare i dati secondo i principi FAIR](#)).

Durante la tua ricerca:

- Usa dei **registri** (vedi link utili) per cercare uno o più repository che si adattino alle tue necessità e alle tipologie di dato che vuoi depositare.
- Scopri se c'è un repository **specifico per il tuo ambito di ricerca**. Un repository disciplinare ti permette di descrivere i tuoi dati usando uno schema di metadati specifico per la materia e li rende più visibili alla tua comunità scientifica di riferimento.
- Verifica se la tua **istituzione di appartenenza** mette uno o più repository a disposizione dei propri membri. L'Università di Bologna offre i repository istituzionali AMS Acta e AMS Historica, per i quali garantisce anche un servizio di supporto, validazione e curatela dei dati [I repository](#).
- Ricorda che esistono anche repository **generalisti** (come ad esempio Zenodo) che tendono a raccogliere dati e materiali eterogenei.
- Scegli un repository che abbia un profilo di sicurezza adeguato e consenta un **accesso controllato** se i tuoi dati devono rimanere inaccessibili a terzi per motivi di privacy [Il rispetto della privacy](#), etica o proprietà intellettuale.
- Ricorda che i servizi di cloud storage (vedi [Conservare i dati in appositi spazi di storage](#)), i siti web personali o di progetto e le piattaforme di social networking come ResearchGate e Academia.edu non sono repository perché non garantiscono la preservazione dei dati nel tempo.
- Presta attenzione a editori o singole riviste che, soprattutto in alcune discipline, si stanno dotando di repository su cui consigliano (o, in alcuni casi, obbligano) gli autori a depositare i propri dati. In queste situazioni, è sempre consigliabile pubblicare preventivamente i propri dati (anche) su un altro repository, disciplinare, istituzionale o generalista.

Link utili

Video "Dati conoscerli e gestirli per valorizzare la ricerca. Conservare i dati a lungo termine"

https://www.youtube.com/watch?v=J3VyrUzj_E

Registri per l'individuazione dei repository:

- Re3data <https://www.re3data.org/search?query=>
- OpenAIRE explore <https://www.openaire.eu/find-trustworthy-data-repository>
- FAIRsharing repository database <https://fairsharing.org/search?fairsharingRegistry=Database>

Repository di ateneo:


AMS Acta <https://amsacta.unibo.it/> | AMS Historica <https://historica.unibo.it/>

Per maggiori informazioni sui repository di Ateneo: "Preservare e disseminare i dati della ricerca in AMS Acta" (<https://sba.unibo.it/it/almadl/servizi-almadl/preservare-disseminare-dati-della-ricerca-in-ams-acta>); "Preservare e valorizzare il patrimonio culturale digitale" (<https://sba.unibo.it/it/almadl/servizi-almadl/preservare-valorizzare-patrimonio-culturale-digitale>).





Repository generalisti: Zenodo <https://zenodo.org/> | Figshare <https://figshare.com/> | Open Science Framework <https://osf.io/>

Depositare i dati secondo i principi FAIR

Il deposito o archiviazione a lungo termine dei dati è parte integrante di una gestione responsabile dei dati di ricerca in linea coi principi FAIR.

I principi FAIR, pubblicati per la prima volta nel 2016, sono dei suggerimenti improntati a migliorare la **riusabilità dei dati di ricerca**, da parte degli individui e dei sistemi informatici. Gestire i dati in linea coi principi FAIR significa renderli **Rintracciabili** (“Findable” in inglese), **Accessibili, Interoperabili e Riutilizzabili**  **I principi FAIR**.

Durante la tua ricerca:

- Scegli **formati standard e aperti**  **Dati della ricerca: tipologie, formati, metodi** e utilizza, dove possibile, vocabolari, ontologie e tassonomie  **Utilizzare diversi tipi di standard** per rendere i tuoi dati comprensibili, interoperabili e riutilizzabili.
- Se pubblichi i dati seguendo i principi FAIR rendi **maggiormente citabili e valorizzabili** sia i dati in sé che le analisi e le pubblicazioni che ne derivano e rendi la tua ricerca più **trasparente e verificabile**, in linea con quanto richiesto da sempre più enti finanziatori, come ad esempio l’Unione Europea.
- I dati FAIR non sono sempre liberamente accessibili a chiunque mentre i metadati, solitamente, lo sono (vedi [Associare ai dati una licenza](#)).
- Depositare i tuoi dati in un **repository**  **I repository** è il primo passo per renderli FAIR: in questo modo, ogni dataset (vedi [Pianificare l’organizzazione dei dati in dataset](#)) è accompagnato da metadati (vedi [Identificare i metadati fondamentali](#)), tra cui un identificativo persistente (PID) e una licenza  **Diritto d'autore**.

Link utili

Wilkinson et al, The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>


Video “Dati della ricerca: la European Open Science Cloud e i principi FAIR” <https://www.youtube.com/watch?v=eNiHNaU6MrQ>

Materiali di approfondimento sui principi FAIR:

- GOFAIR. FAIR Principles <https://www.go-fair.org/fair-principles>
- FAIRsFAIR Fostering Fair Data Practices in Europe <https://www.fairsfair.eu/>
- How to FAIR <https://howtofair.dk/>

Associare ai dati una licenza

Scegliere una licenza da associare ai dataset depositati in linea con i principi FAIR permette di specificare quali sono le modalità con cui questi dataset possono essere riutilizzati.


La pubblicazione dei dati ad accesso aperto favorisce una **ricerca aperta e collaborativa** ed è una pratica che si colloca all'interno del movimento dell'Open Science (Scienza Aperta). I dati ad accesso aperto (Open Data) sono **distribuiti con licenze che ne garantiscono la libertà di accesso, utilizzo, modifica e condivisione** da parte di chiunque, prevedendo al massimo restrizioni che riconoscano l'attribuzione di paternità e limitino alcune utilizzazioni in casi specifici, preservandone l'apertura  **Diritto d'autore**.

Open Science è un movimento che ha per obiettivo **l'accesso senza barriere al sapere scientifico da parte della comunità scientifica e dei cittadini**, si basa sui principi di trasparenza, inclusione, correttezza, equità e condivisione. È obiettivo strategico dell'Unione Europea dal 2015, dell'UNESCO dal 2021 e del Ministero della Ricerca italiano dal 2022 col *Piano nazionale della Scienza Aperta*.

È sempre fondamentale ricordare che la scelta della licenza deve essere guidata dal principio "as open as possible, as closed as necessary", avendo consapevolezza

che in alcuni casi è opportuno limitare l'accesso ai dati qualora ciò risulti funzionale alla valorizzazione della ricerca per finalità di natura commerciale, in linea con la missione dell'Ateneo di favorire il trasferimento dei risultati di ricerca per ottenere un impatto sull'economia e sulla società.

Durante la tua ricerca:

- Gestisci i tuoi dati secondo i **principi FAIR**  **I principi FAIR** durante tutto il loro ciclo di vita e scrivi un **Data Management Plan** (vedi [Redigere un Data Management Plan](#)).
- Distribuisci i tuoi dati ad accesso aperto **se non sussistono vincoli** derivanti da diritti di terze parti o altri divieti di legge, e se questo non pregiudica le opportunità di valorizzazione commerciale dei risultati di ricerca.
- Per fare Scienza Aperta e distribuire i tuoi dati apertamente, scegli **licenze molto permissive**, ossia che ne consentano qualsiasi uso, con qualsiasi mezzo e formato e per qualsiasi fine, anche di natura commerciale. Esempi di licenze di questo tipo sono CC0 1.0, CC BY 4.0, CC BY-SA 4.0.

Link utili

Video "I Principi dell'Open Science" <https://www.youtube.com/watch?v=qzktH6YlOf8>

Pagina di Ateneo su Open Science <https://www.unibo.it/it/ricerca/open-science>

The Turing Way Guide for Reproducible Research <https://the-turing-way.netlify.app/reproducible-research/open/open-data>

Open Definition "Defining Open in Open Data, Open Content and Open Knowledge" <https://opendefinition.org/od/2.1/en/>

Per maggiori informazioni sul diritto d'autore e sulla tutela e valorizzazione del patrimonio culturale: <https://sba.unibo.it/it/almadl/servizi-almadl/supporto-giuridico-per-la-gestione-del-diritto-dautore-e-la-tutela-del-patrimonio-culturale>

Conclusioni

La gestione dei dati della ricerca è un insieme di buone pratiche per valorizzare i dati durante tutto il loro ciclo di vita, dalle fasi iniziali fino al deposito e alla condivisione.

Gestire correttamente i dati presenta una serie di vantaggi per il ricercatore in termini di valorizzazione, qualità e impatto della ricerca.

Il processo di gestione dei dati della ricerca si articola, come abbiamo visto, su azioni ben definite: affrontare ciascun punto con consapevolezza e al momento giusto è una condizione fondamentale per assicurare un processo corretto e di qualità.

Punto di partenza imprescindibile per percorrere queste azioni è l'identificazione consapevole dei tipi di dati con cui si lavora. Ogni ricerca si basa su dati, diversi per tipologie, origine e utilizzo. Nel rispetto dunque delle diversità disciplinari, il processo di gestione del dato consta di fasi e azioni che sono standardizzabili. Ogni area disciplinare ha le sue specificità nella definizione della tipologia di dati prodotti e nelle strategie di gestione.

Contatti

All'interno dell'Università di Bologna sono presenti diversi tipi di supporto alla gestione corretta dei dati della ricerca.

Per supporto riguardo la gestione dei dati potete rivolgervi ai data stewards che lavorano in ARIC – area della ricerca: aric.datasteward@unibo.it.

Per supporto all'utilizzo dei repository di ateneo (AMS Acta o AMS Historica) potete rivolgervi a: almadl@unibo.it.

Per supporto in materia di diritto d'autore e di altri diritti connessi al suo esercizio e di tutela e valorizzazione del patrimonio culturale, potete rivolgervi a: almadl@unibo.it.

Per supporto riguardo la valorizzazione commerciale dei risultati di ricerca UniBo potete rivolgervi al Knowledge Transfer Office: kto@unibo.it.

Per supporto sulle questioni relative alla privacy potete scrivere all'indirizzo: privacy@unibo.it.

 **Checklist:** i punti fondamentali per la corretta gestione del dato di ricerca.

FASE DI PIANIFICAZIONE DEL FLUSSO DATI

- Identificare i tipi di dati.**
 - Decidere se **generare dati nuovi o riutilizzare** quelli già disponibili.
 - Avere consapevolezza **della normativa vigente**, ad esempio quella relativa al trattamento dei dati personali.
- Identificare i **metadati fondamentali**.
- Pianificare l'**organizzazione dei dati in dataset**.
- Redigere un **Data Management Plan**.

FASE DI PRODUZIONE, RACCOLTA E ANALISI DEI DATI

- Conservare i dati in **adatti spazi di storage**, con **backup**.
- Assicurare la **qualità** dei dati.
 - Tenere traccia delle diverse **versioni**.
 - Organizzare file e cartelle** in modo coerente.
 - Scegliere i formati** più adatti per i dati.
- Documentare** la raccolta dei dati.

FASE DI ARCHIVIAZIONE E CONDIVISIONE DEI DATI

- Identificare i **dati da preservare a lungo termine**.
- Scegliere il **repository più appropriato** per il deposito.
- Depositare** i dati, organizzati in dataset, secondo i **principi FAIR**.
- Associare la **licenza più appropriata** ai dati da depositare, tenendo a mente possibili vincoli esistenti.



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA