



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA



24 gennaio 2024

Open Science Corner

Dai principi alla pratica: gestire i dati della ricerca

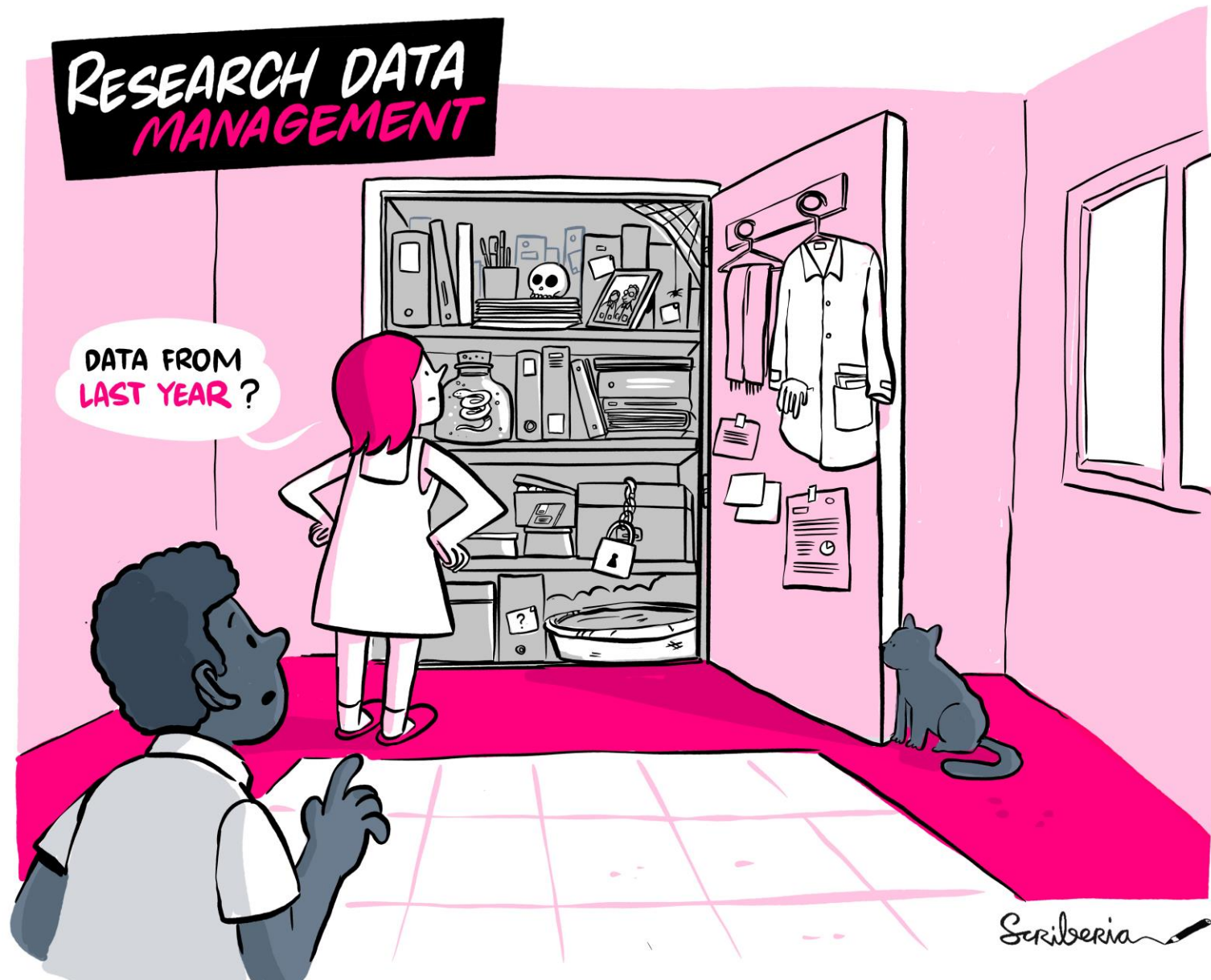
24/01/2024

Giulia Caldoni

Data Steward

Area della Ricerca (ARIC)

Alma Mater Studiorum – Università di Bologna



RESEARCH DATA MANAGEMENT

1. Gestire i dati correttamente: cosa vuol dire e perché è bene farlo

2. I dati della ricerca: qualche definizione

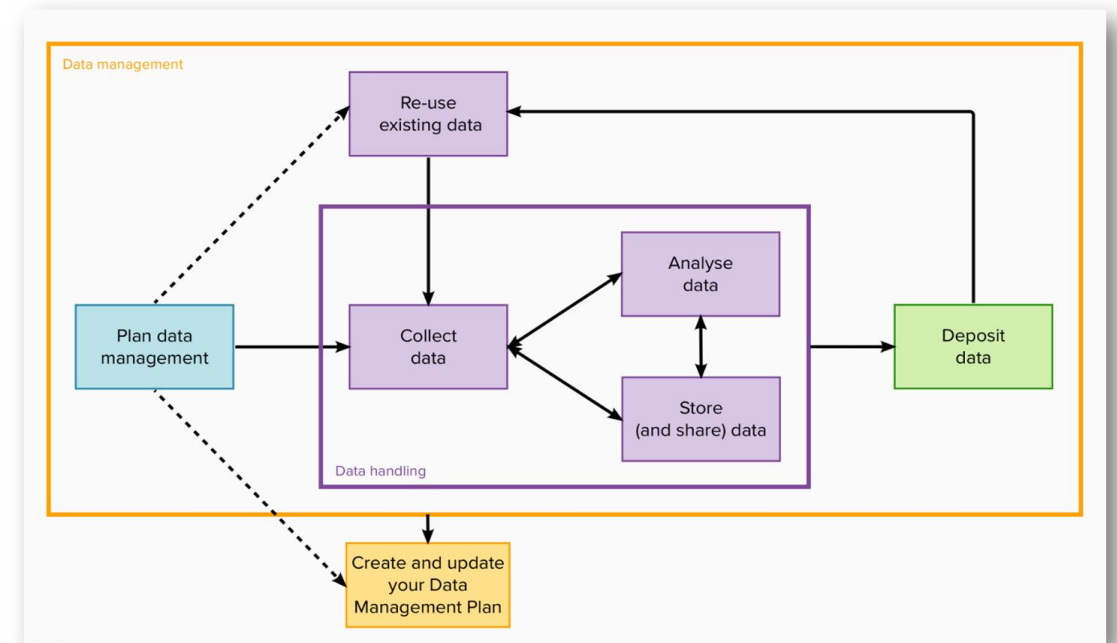
3. Gestire i dati della ricerca: come farlo in modo FAIR

Scriberia



Gestire i dati correttamente: cosa vuol dire

Quando parliamo di gestione dei dati della ricerca (o Research Data Management, RDM) ci riferiamo **alla cura e all'organizzazione consapevole dei dati** durante l'intero ciclo di ricerca...



...con l'obiettivo di rendere il processo di ricerca il più **efficiente** possibile e di **facilitare la cooperazione** con gli altri.



Gestire i dati correttamente: perché è bene farlo



Organizzare i dati rende il lavoro più efficiente.

In termini di costi/tempo: i dati gestiti una volta restano interpretabili, comprensibili e rintracciabili.



Una corretta gestione aiuta a prevenire la perdita dei dati.

L'archiviazione corretta dei dati e il backup regolare sono buone pratiche da implementare.



Alcuni dati potrebbero essere unici e non riproducibili.

Questo li rende preziosissimi per la comunità scientifica e gestirli correttamente ne favorisce la condivisione.



Un dato correttamente gestito favorisce l'integrità della ricerca.

Questo perchè sono facilitate la validazione e il controllo.



La corretta gestione dei dati stimola la collaborazione con altri ricercatori.

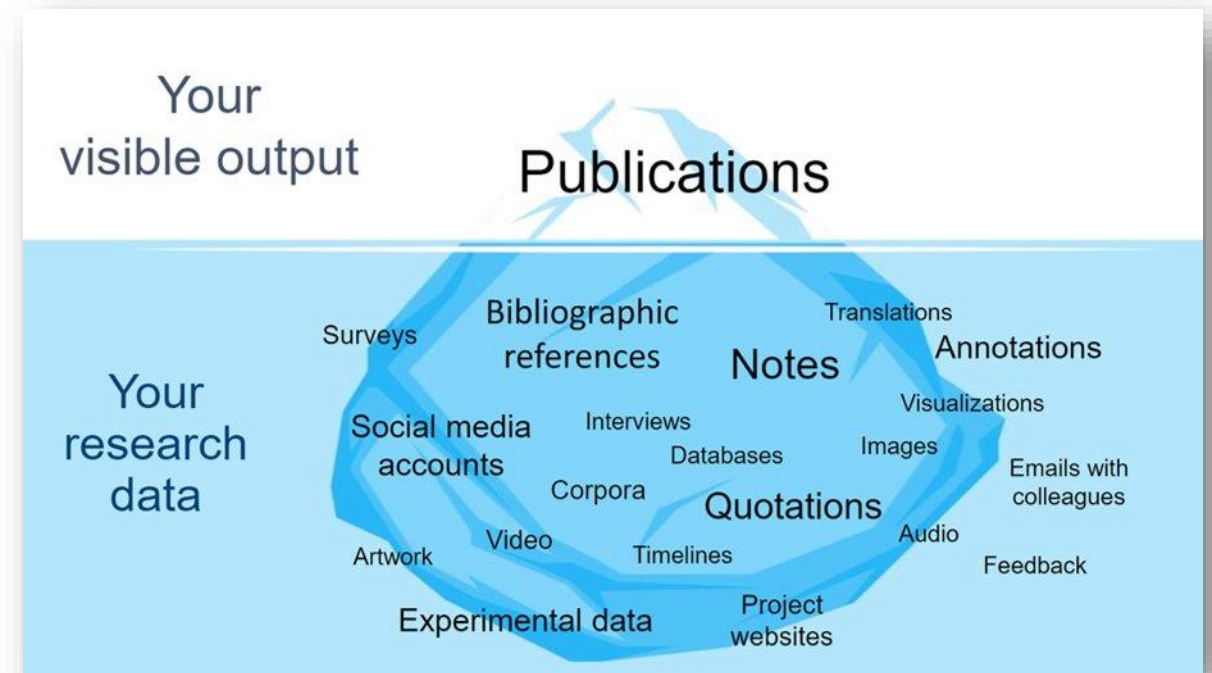
Questo perchè dati curati e interoperabili sono più facili da comprendere e riutilizzare.

I dati che non sono gestiti hanno un'appetibilità bassa per lo stesso ricercatore che li produce.



Perché l'interesse per i dati

L'Open Science passa attraverso la gestione consapevole di tutti gli output di ricerca (es. **pubblicazioni, dati, software, modelli, algoritmi e metodologie**).



La **valutazione della ricerca** si sta muovendo verso un modello meno legato alle sole pubblicazioni ma **più attento a tutti questi diversi output** (vedi es. [CoARA](#), [VQR](#), [EC report 2021](#)).

L'accento è posto sui dati e gli **strumenti necessari per validare le conclusioni** di un processo di ricerca (semplificando: dati + metodologie).



Perché l'interesse per i dati

National Agency for the Evaluation of
Universities and Research Institutes



Agenzia Nazionale di Valutazione del
sistema Universitario e della Ricerca

Evaluation of Research Quality



Valutazione Qualità della Ricerca

- b) *metodologia*, da intendersi come la capacità del prodotto di presentare in modo chiaro gli obiettivi della ricerca e il loro valore scientifico, la letteratura utilizzata e i risultati ottenuti, favorendo altresì, ove applicabile, la riproducibilità dei risultati, la trasparenza rispetto a metodi e procedure adottate e l'accesso ai dati utilizzati, nella logica di valorizzare l'intero processo che ha portato alla realizzazione del prodotto della ricerca;

Bando VQR 2020-2024, criterio metodologia (art. 7): i **dati e le metodologie della ricerca** rientrano per la prima volta nei criteri di valutazione dei prodotti della ricerca → valorizzare l'intero processo che ha portato alla realizzazione del prodotto della ricerca.

Riconosciuti come premianti: la **riproducibilità** della ricerca, la trasparenza rispetto a **metodi e procedure**, l'accesso ai **dati** utilizzati.



Gestire i dati correttamente in Unibo

Policy per la gestione dei dati di ricerca:

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

<https://pianostrategico.unibo.it/it/>

OBIETTIVO 31

Favorire lo sviluppo dell'open science

- Favorire la gestione FINDABLE, ACCESSIBLE, INTEROPERABLE, REUSABLE (FAIR) dei dati della ricerca, anche attraverso servizi e infrastrutture di supporto



Policy di Ateneo per la Gestione dei Dati della Ricerca

1) Finalità e Obiettivi

Alma Mater Studiorum - Università di Bologna (di seguito "Ateneo" o "Università") riconosce l'importanza fondamentale dei Dati creati e/o raccolti durante l'attività di ricerca come un valido risultato scientifico per il progresso delle conoscenze. I Dati della Ricerca costituiscono un patrimonio da valorizzare e una risorsa per la ricerca scientifica, la didattica e la società tutta. Pertanto, l'Università ne favorisce la gestione secondo gli standard e le migliori pratiche internazionali a garanzia dei valori di qualità e integrità della ricerca scientifica.

Attraverso la presente Policy, l'Ateneo intende definire i principi guida per la corretta Gestione dei Dati della Ricerca, in accordo con i Principi FAIR e gli standard internazionali e disciplinari di riferimento, nonché delineare ruoli e responsabilità nel processo di Gestione dei Dati stessi.

2) Definizioni

Per "Dati della Ricerca" (o "Dati") si intendono le informazioni, in qualsiasi formato, utilizzate secondo un protocollo definito nell'ambito di una specifica attività di ricerca condotta dal Ricercatore e necessarie per validare i risultati della ricerca stessa, inclusi quelli utilizzati nelle pubblicazioni scientifiche. A titolo meramente esemplificativo sono da intendersi Dati della Ricerca: risultati, fatti, osservazioni, esperienze, fonti edite e inedite, riferimenti bibliografici, testi, immagini, creati e/o raccolti in formato digitale, nonché altri output digitali della ricerca come, ad esempio, modelli 3D e codice sorgente.

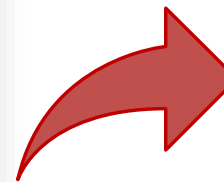
Per "Dataset" si intende un insieme organizzato di Dati posti in relazione tra loro, creati e/o raccolti con uno scopo comune e organizzati per riflettere i risultati di un'attività di ricerca.

Per "Gestione dei Dati della Ricerca" (o "Research Data Management" o "RDM") si intende l'adozione di buone pratiche, linee guida e standard per assicurare la qualità, l'integrità, l'intelligibilità, la conservazione sicura dei Dati nel corso delle ricerche, nonché la loro preservazione e l'accessibilità a lungo termine, nel rispetto delle normative vigenti e tenendo conto delle politiche di accesso aperto.

Per "Data Management Plan" (o "Piano di Gestione dei Dati della Ricerca" o "DMP") si intende un documento che illustra la Gestione dei Dati della Ricerca creati e/o raccolti oppure riutilizzati durante l'attività di ricerca. Nel DMP sono descritti la natura e l'origine dei Dati della Ricerca, le misure adottate per garantirne la qualità e la tutela, i soggetti coinvolti nella loro produzione ed elaborazione, l'organizzazione e Gestione dei Dati durante e dopo il termine dell'attività di ricerca e le misure adottate per rendere i Dati della Ricerca in linea con i Principi FAIR, nel rispetto dei vincoli di legge e degli accordi stipulati con terze parti.

Per "Principi FAIR"¹ si intendono i principi guida definiti e riconosciuti a livello internazionale per facilitare la scoperta e l'indicizzazione dei Dati della Ricerca, garantirne la preservazione e l'accessibilità, l'integrazione e il riuso anche in modo automatizzato. L'acronimo FAIR sintetizza le caratteristiche che i

¹ FAIR principles definition: Wilkinson, M., Dumontier, M., Aalbersberg, I. 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>



- Finalità e obiettivi
- Definizioni
- Ambito di applicazione
- **Gestione dei dati della ricerca**
- Ruoli e responsabilità

2024 → linee guida

4) Gestione dei Dati della Ricerca

Ogni Ricercatore gestisce i Dati della Ricerca secondo i Principi FAIR.

Nelle fasi iniziali della ricerca è opportuno che il Ricercatore delinea quali Dati intende trattare, siano essi creati e/o raccolti *ex novo* oppure riutilizzati in quanto prodotti da terzi.

Nel caso di Dati creati e/o raccolti *ex novo*, il Ricercatore agisce nel pieno rispetto della normativa privacy e di diritto d'autore, in coerenza con i principi etici. Il Ricercatore pianifica fin dalle prime fasi di ricerca ogni azione volta a tutelare e valorizzare i risultati della ricerca anche per finalità commerciali, al fine di promuoverne il trasferimento per sviluppare nuovi prodotti, processi e servizi sul mercato. Il Ricercatore si astiene dal compiere azioni, come ad esempio la divulgazione, che possano pregiudicare le opportunità di valorizzazione dei Dati della Ricerca.

Nel caso di riutilizzo, è importante verificarne a monte la fattibilità. Il Ricercatore è tenuto a controllare la sussistenza di una licenza d'uso associata e gli usi consentiti dalla licenza stessa, e ad assicurarsi che il riutilizzo avvenga nel rispetto del diritto d'autore e nel rispetto di eventuali vincoli determinati dalla normativa privacy.

La redazione di un DMP, anche se non viene esplicitamente richiesta dall'ente finanziatore della ricerca, è una buona pratica sempre consigliata per accompagnare e garantire la corretta Gestione dei Dati della Ricerca fin dalle fasi iniziali.

Durante l'attività di ricerca i Dati sono trattati in modo da garantirne la qualità e l'integrità, e in coerenza con quanto stabilito dai regolamenti di Ateneo o da eventuali accordi con terzi in termini di accessibilità e confidenzialità. La corretta conservazione (storage) del Dato durante le fasi attive della ricerca si avvale di infrastrutture atte a garantire almeno le procedure di backup e a proteggere, qualora necessario, la natura sensibile del Dato, secondo quanto previsto dalla normativa vigente. I Dati sono organizzati in Dataset, accompagnati dai Metadati e da una adeguata documentazione che ne illustri le metodologie di creazione e/o raccolta, nonché i protocolli e gli strumenti adottati per favorirne l'accessibilità e il riutilizzo.

Al termine dell'attività di ricerca, e comunque prima di pubblicare i risultati, i Dati rilevanti per la verificabilità, correttezza e riproducibilità della ricerca stessa sono archiviati e resi accessibili. Tali Dati sono archiviati in modo corretto, completo, affidabile, rispettandone l'integrità. Risultano inoltre accessibili, identificabili, tracciabili, interoperabili e, laddove possibile, disponibili per usi successivi.

I Dati depositati sono corredati delle informazioni necessarie alla loro identificazione e, possibilmente, ad Accesso Aperto, distribuiti cioè con licenze che ne garantiscano il libero utilizzo per favorire una ricerca aperta e collaborativa, a eccezione dei casi in cui sussistano opportunità di valorizzazione a fini commerciali, diritti di terzi o altri vincoli di legge che ne impediscano la disseminazione ad Accesso Aperto. In tali casi è opportuno che i Dati rilevanti per la riproducibilità della ricerca vengano comunque depositati in un Repository che consenta un accesso controllato ai Dati e garantisca invece l'Accesso Aperto ai Metadati e alla documentazione di supporto.



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Gestire i dati correttamente in Unibo

Servizi di supporto:

<https://www.unibo.it/it/ricerca/open-science>

OBIETTIVO 31

Favorire lo sviluppo dell'open science

- Favorire la gestione FINDABLE, ACCESSIBLE, INTEROPERABLE, REUSABLE (FAIR) dei dati della ricerca, anche attraverso servizi e infrastrutture di supporto

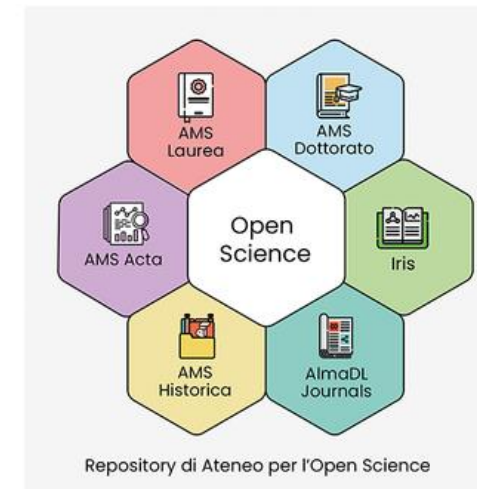
I dati della ricerca: il progetto "Data Steward@Unibo"



Data Stewards: supporto riguardo la gestione dei dati
(aric.datastewards@unibo.it)



Servizi di supporto: repository, consulenza, formazione



Digital Library di Ateneo (AlmaDL): supporto per
l'utilizzo dei repository di ateneo (almadl@unibo.it)



Gestire i dati correttamente in Unibo

Materiali utili:

Pagina internet/intranet:

- <https://intranet.unibo.it/Ricerca/Pagine/OpenScience.aspx>
- <https://www.unibo.it/it/ricerca/open-science>

Video:

- Rassegna di video informativi «Dati: conoscerli e gestirli per valorizzare la ricerca»
<https://www.youtube.com/playlist?list=PLaUmBQ7P5K-AyDDnv1f8upAyEOtAF2gj3>

Eventi:

- Ciclo di incontri «Open Science Corner»
<https://intranet.unibo.it/Ricerca/Pagine/CicloIncontri2024DataManagement.aspx>
- Corso «Research data management e principi di Open Science» per dottorandi
<https://eventi.unibo.it/competenze-trasversali-dottorato/research-data-management-e-principi-di-open-science>

<https://pianostrategico.unibo.it/it/>

OBIETTIVO 31

Favorire lo sviluppo dell'open science

- Favorire la gestione FINDABLE, ACCESSIBLE, INTEROPERABLE, REUSABLE (FAIR) dei dati della ricerca, anche attraverso servizi e infrastrutture di supporto

I dati della ricerca: qualche definizione



DATI: informazioni, in qualsiasi formato, utilizzate secondo un protocollo definito nell'ambito di una specifica attività di ricerca condotta dal ricercatore e **necessarie per validare i risultati della ricerca** stessa, inclusi quelli utilizzati nelle pubblicazioni scientifiche.



DATASET: insieme organizzato di dati posti in relazione tra loro, creati e/o raccolti con uno scopo comune e organizzati per riflettere i risultati di un'attività di ricerca.



METADATI: informazioni strutturate (ad es. di natura bibliografica, amministrativa, gestionale, di provenienza) che riguardano i dati della ricerca e ne favoriscono l'identificazione e l'indicizzazione da parte di motori di ricerca e portali di aggregazione, ne facilitano l'organizzazione, la gestione e l'intelligibilità, ne certificano la provenienza e le responsabilità autoriali e definiscono le condizioni di accesso e di riutilizzo.



Dati, dataset e metadati: un esempio dalla vita quotidiana



Dati: una serie di fotografie disorganizzate



Dataset: un insieme di fotografie raccolte con uno scopo comune



Metadati: informazioni che descrivono ogni fotografia e l'intero album

Così che le foto possano essere capite dagli altri e facilmente rintracciabili tra altri album

Diversi tipi di dati

Esiste un'enorme varietà di tipi di dati, che possono essere classificati in modi diversi:

- **Contenuto:** numerico, testuale, audiovisivo, multimediale...
- **Formato:** fogli di calcolo, database, immagini, mappe, file audio, testo...
- **Modalità di raccolta:** sperimentale, osservativa, di simulazione, dati derivati/compilati da altre fonti
- **Natura digitale** (nata digitale o digitalizzata) **o non digitale** (es. indagini cartacee, appunti...)
- **Natura primaria** (generati dal ricercatore per un particolare scopo di ricerca o progetto) **o secondaria** (creati originariamente da qualcun altro per un altro scopo)
- **Natura grezza o elaborata**

Ogni ricerca produce o riusa dei dati, ma **ogni area disciplinare ha le sue specificità** nella tipologia di dati prodotti e nelle loro strategie di gestione.



Dataset

Una singola ricerca può produrre tanti tipi diversi di dati, che concorrono a rispondere alla stessa domanda scientifica.

Name	Size	Download all
Analysis.csv md5:be5c73f71f3dce272f8b7377457225ef	275.6 kB	Preview Download
Anatomy.csv md5:5c3aa00fc330360e89029c24fa4150c1	1.7 kB	Preview Download
Bern_ECochG.db md5:7d76b18a0f20c306943b0956719c65d3	32.8 kB	Download
Demographics.csv md5:d5a2e9bf5a5a40c08ff62c2195279c67	2.1 kB	Preview Download
ECochG.csv md5:5a17b11708832c2fd8fbc64ddefab363	23.6 GB	
Hearing_thresholds.csv md5:3286485b10b6ee162b80fe3f85b26439	12.4 kB	
Preprocessed.csv md5:695f1ecfecdfdb9c55a9a7e1c79bc73	964.8 MB	
README.txt md5:a1db94aa395e46c50086605c50733c88	6.1 kB	
Telemetry.csv md5:60708e4686b398026418a78a63fc9750	17.6 kB	

Strutturarli all'interno di un dataset permette di **organizzarli e relazionarli** tra loro, rendendo chiaro il processo che ha portato ad ottenere il risultato.

Name	Size	Download all
Predicting code generation.txt md5:a7a999c28d8fa5cb86d86d381a033192	5.1 kB	Preview Download
Regression code generation.txt md5:30924e34baba4f09aa56772ea3111df0	5.2 kB	Preview Download
Rivers's Dataset.xlsx md5:a9cf79c2af88b50a86ee1dbfa0303ed1	119.8 kB	Download

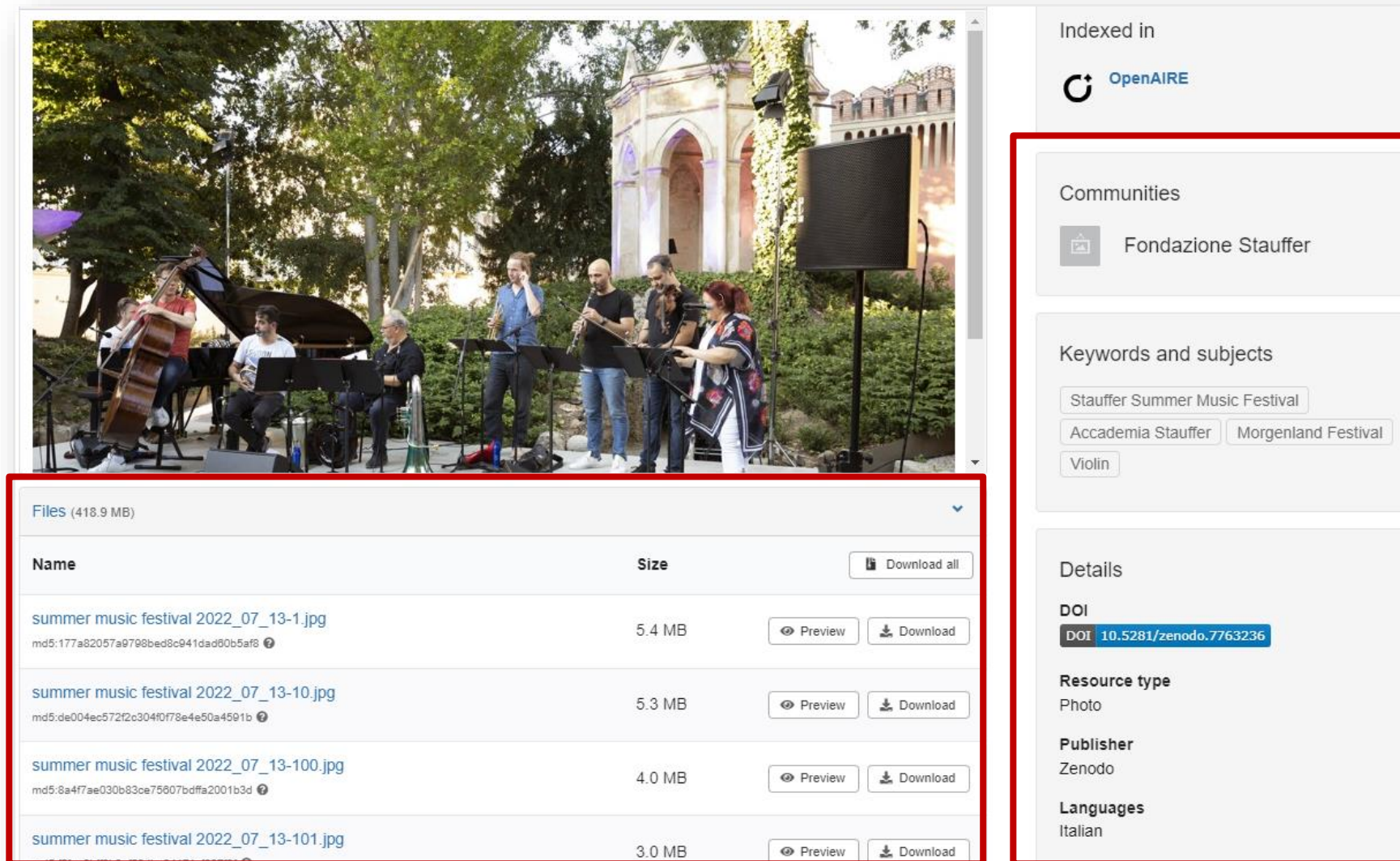


Metadati

Avere a disposizione dei dati non serve a nulla se non si riesce a comprenderli!

I metadati sono informazioni strutturate sui dati → si riferiscono a **informazioni strutturate che descrivono, spiegano, localizzano o facilitano il reperimento, l'uso o la gestione di un'altra risorsa** (o dati) - sia per le persone che per le macchine.

Metadati **apertamente disponibili** permettono di sapere quali sono i dati, anche se questi non sono accessibili.



The screenshot displays a digital repository interface. At the top, there is a photograph of a music festival performance outdoors. Below the photo, a table lists files with their names, sizes, and download options. To the right, a sidebar contains metadata information such as 'Indexed in OpenAIRE', 'Communities Fondazione Stauffer', 'Keywords and subjects' (including 'Stauffer Summer Music Festival', 'Accademia Stauffer', 'Morgenland Festival', and 'Violin'), and 'Details' (including DOI 10.5281/zenodo.7763236, Resource type Photo, Publisher Zenodo, and Languages Italian).

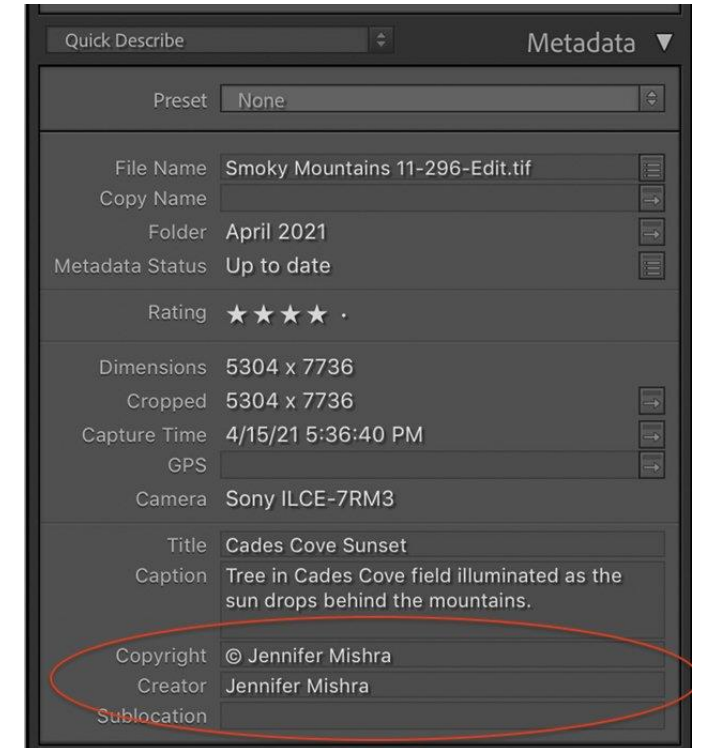
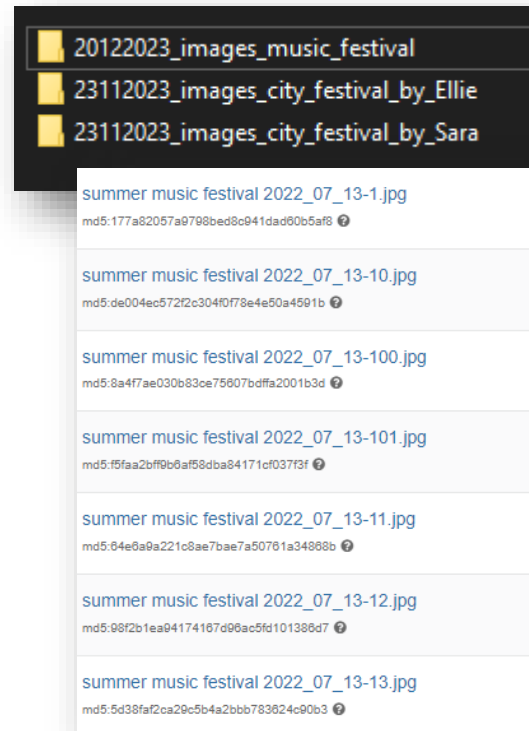
Name	Size	Download all
summer music festival 2022_07_13-1.jpg <small>md5:177a82057a9798bed8c941dad60b5af8</small>	5.4 MB	Preview Download
summer music festival 2022_07_13-10.jpg <small>md5:de004ec572f2c304f0f78e4e50a4591b</small>	5.3 MB	Preview Download
summer music festival 2022_07_13-100.jpg <small>md5:8a4f7ae030b83ce75807bdfaf2001b3d</small>	4.0 MB	Preview Download
summer music festival 2022_07_13-101.jpg	3.0 MB	Preview Download

Dati

Metadati



Dati, dataset e metadati: un esempio dalla ricerca

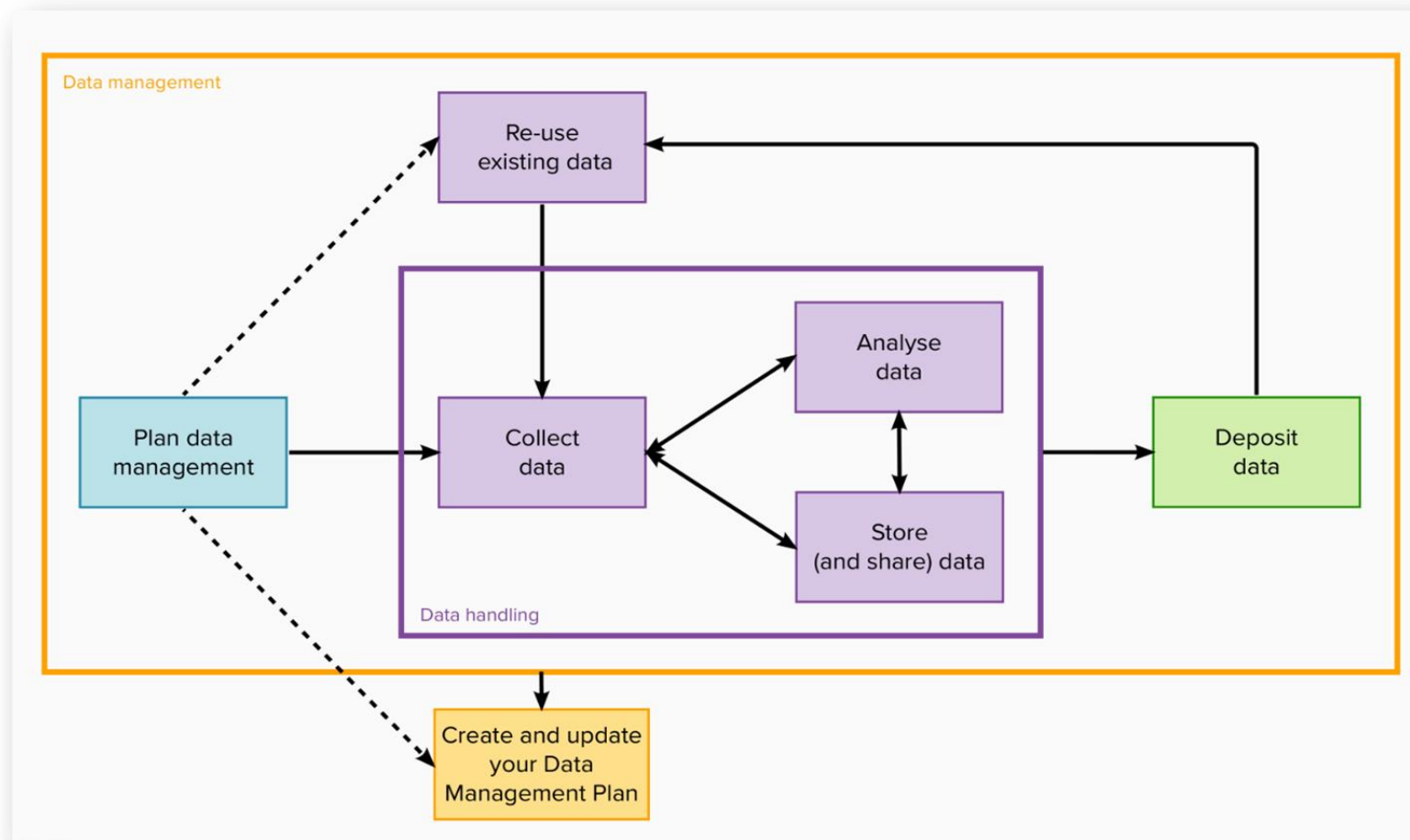


Dati: una serie di fotografie disorganizzate, scattate ad un evento

Dataset: un insieme di fotografie raccolte con uno scopo comune

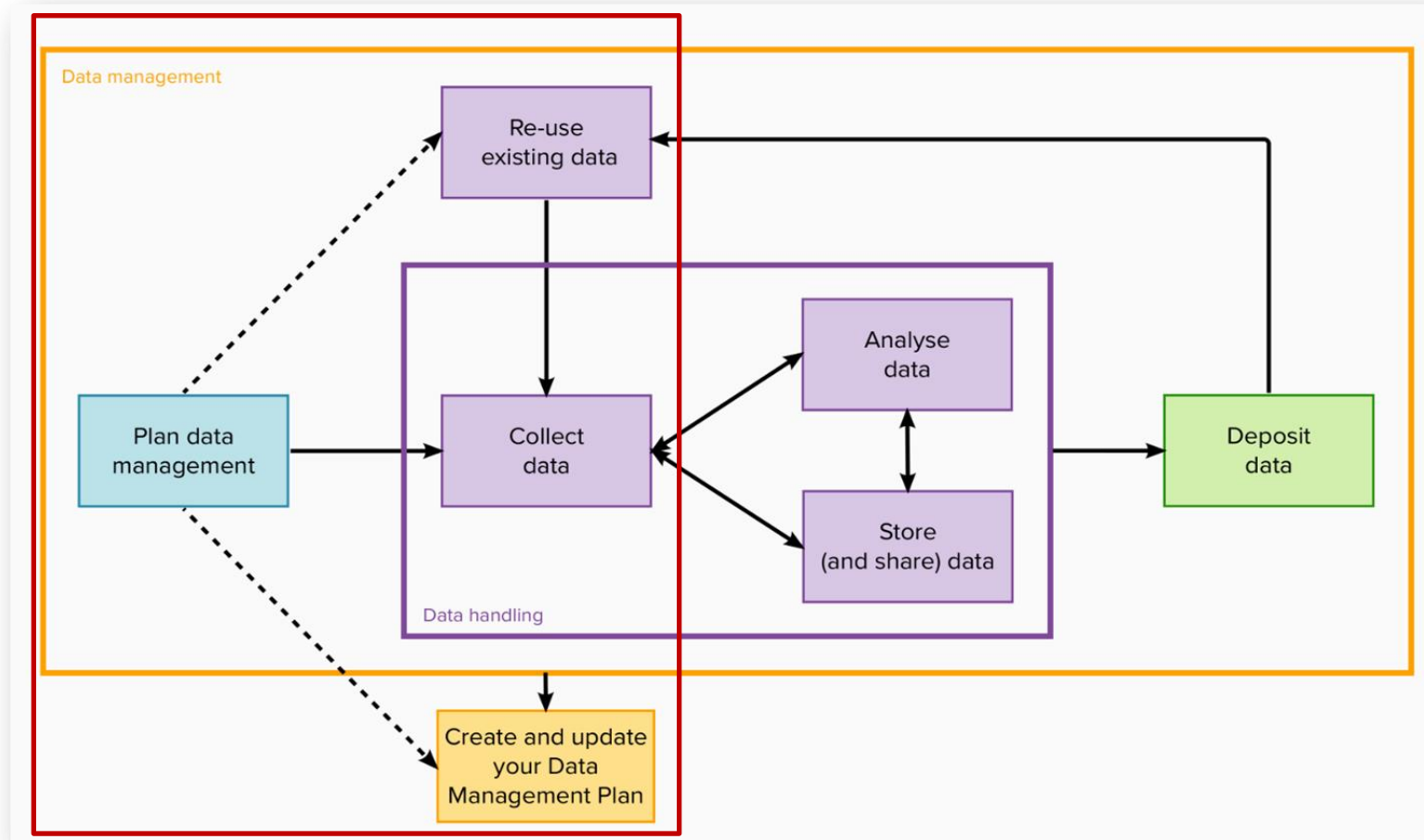
Metadati: informazioni strutturate che descrivono ogni immagine e l'intero dataset

Gestire i dati della ricerca: come fare



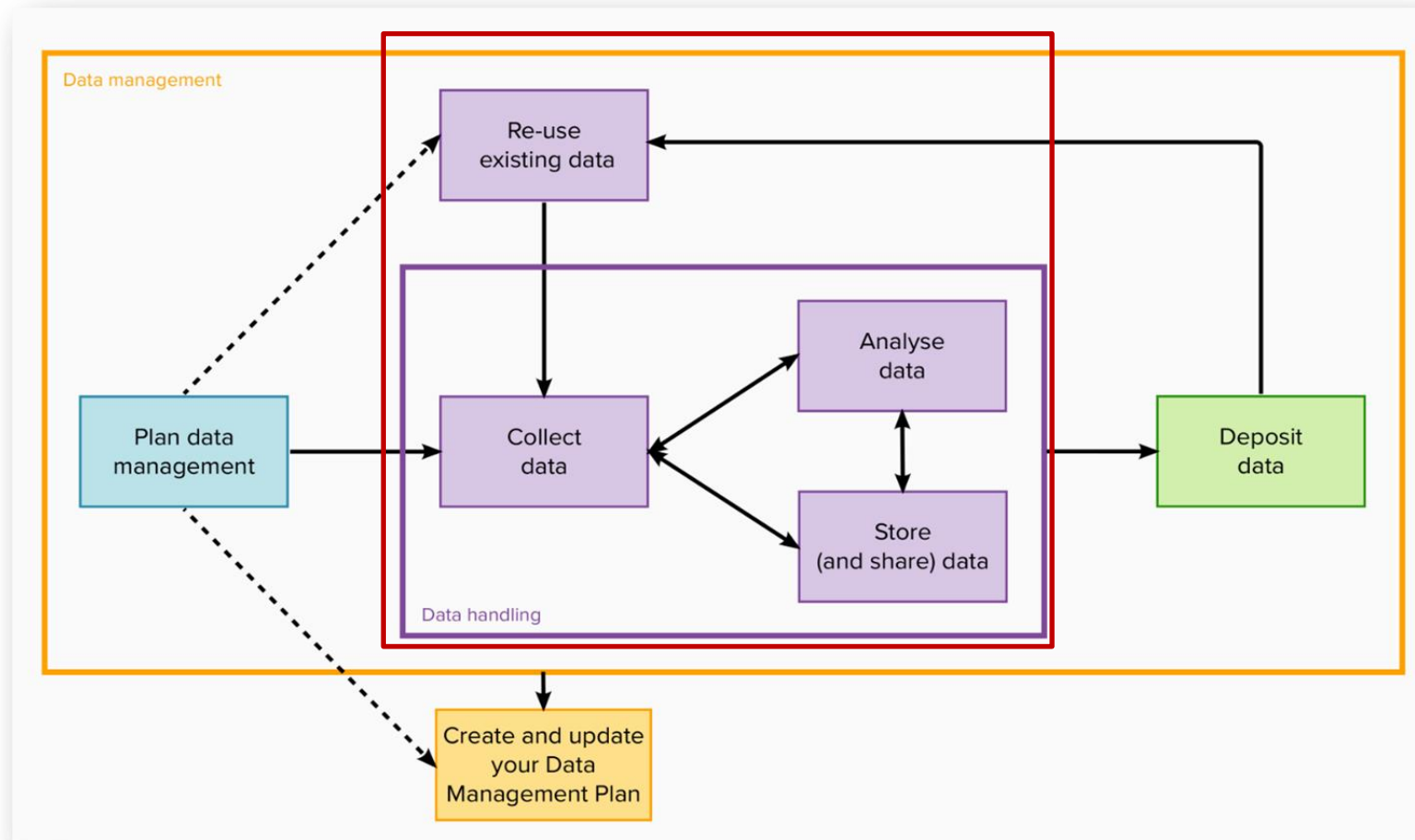
Gestire i dati della ricerca: come fare

FASE 1: pianificazione



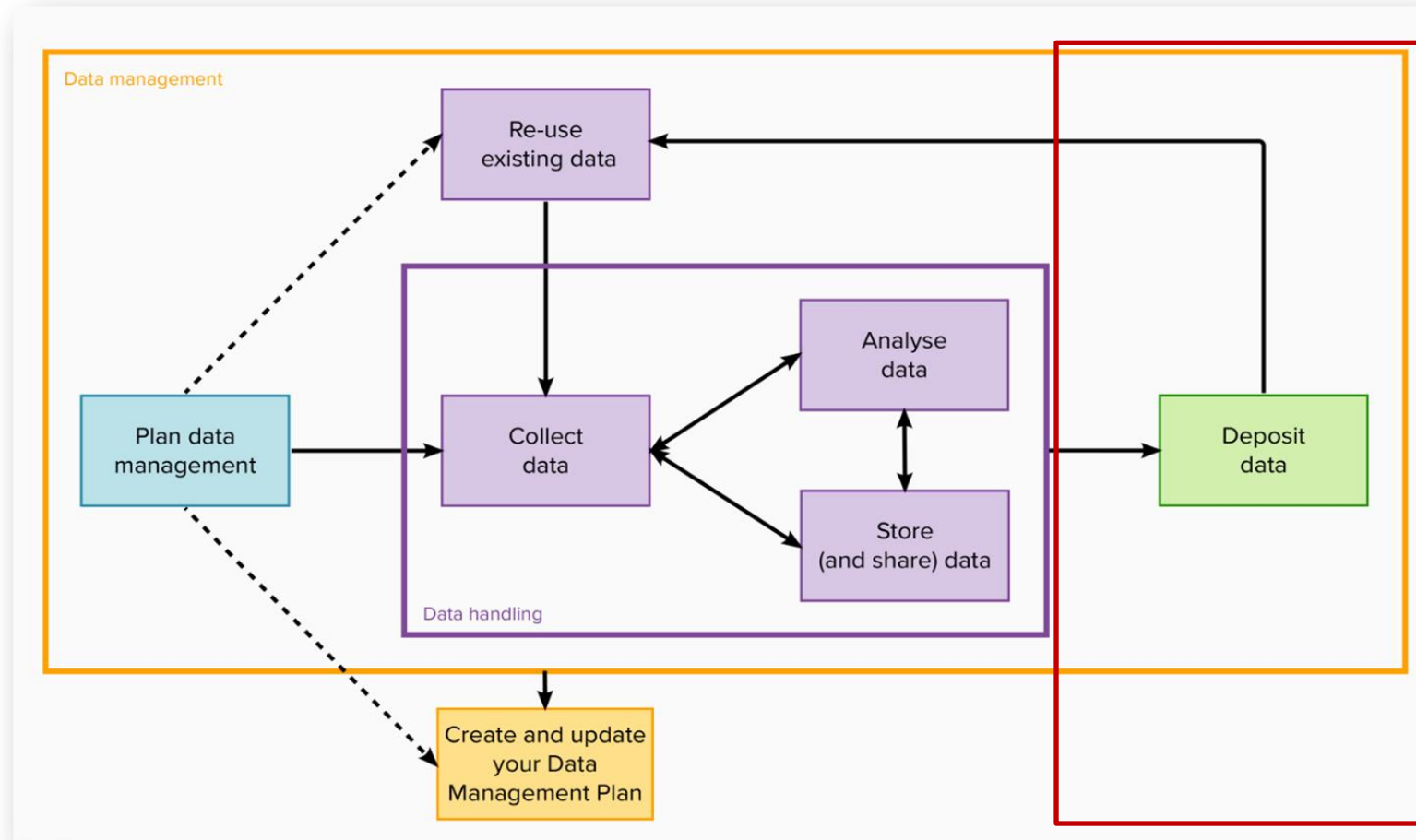
Gestire i dati della ricerca: come fare

FASE 2: gestione attiva



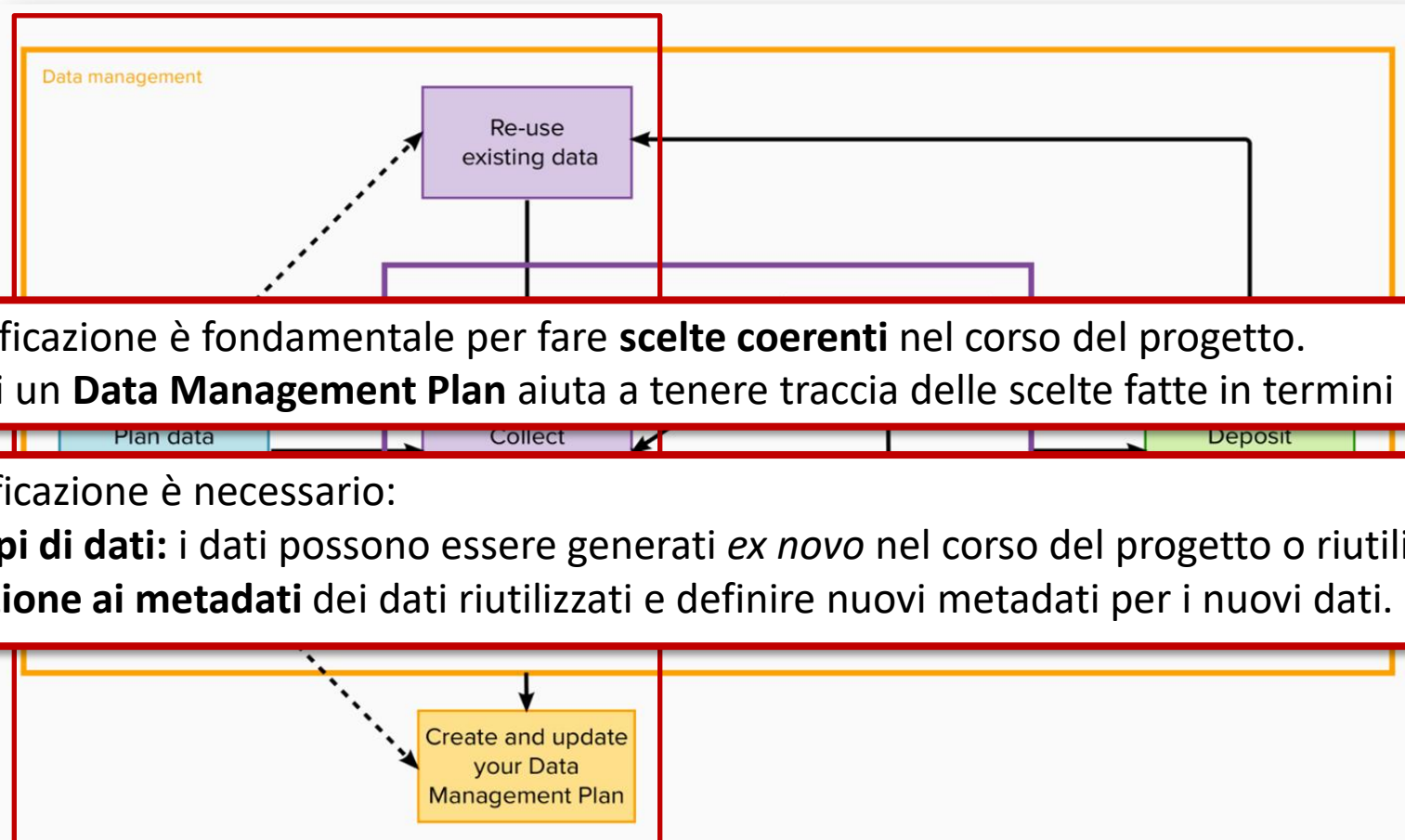
Gestire i dati della ricerca: come fare

FASE 3: preservazione



Gestire i dati della ricerca: come fare

FASE 1: pianificazione



La fase di pianificazione è fondamentale per fare **scelte coerenti** nel corso del progetto.
La redazione di un **Data Management Plan** aiuta a tenere traccia delle scelte fatte in termini di gestione dei dati.

Nella fase di pianificazione è necessario:

- **Identificare i tipi di dati:** i dati possono essere generati *ex novo* nel corso del progetto o riutilizzati da fonti esistenti.
- Prestare **attenzione ai metadati** dei dati riutilizzati e definire nuovi metadati per i nuovi dati.

Identificare i tipi di dati



Immagini: e.g. imaging medico, microscopia, digitalizzazioni..

<https://nbia.cancerimagingarchive.net/nbia-viewer/?thumbnailSeries=891235954&thumbnailDescription=CT%205.0MM%20STD&accessToken=c9300b3c-2cde-4575-b873-e121aea4bc60:6633f9b4-b07f-41e7-88f4-20d5279abf63:6262>
Mafatih al-jinan wa-masabih al-janan. Bologna, Biblioteca Universitaria, ms. 2970. CC BY-NC-ND 4.0. AMS Historica, <https://amshistorica.unibo.it/archivio/0001686/000010.jpg>

```
inputStream
// convert input record into output format
.map { buildNormalizedHostFinding(it) }
// collect consecutive items before moving to reduction step
.transform { bufferingFunction(it) }
// combine output into newline-separated string
.map { (groupingKey, groupedFindings) ->
  groupedFindings.reduce { acc, finding ->
    acc.output = "${acc.output}\n${finding.output}"
    acc
  }
}
// use s3 reference if final output size is too large
.doOnNext {
  if (it.output.toByteArray().size > maxOutputSize) {
    val s3Uri = saveToS3(it)
    it.output = s3Uri
  }
}
```

Codice: e.g. codice sorgente del software..

Video: e.g. registrazioni..



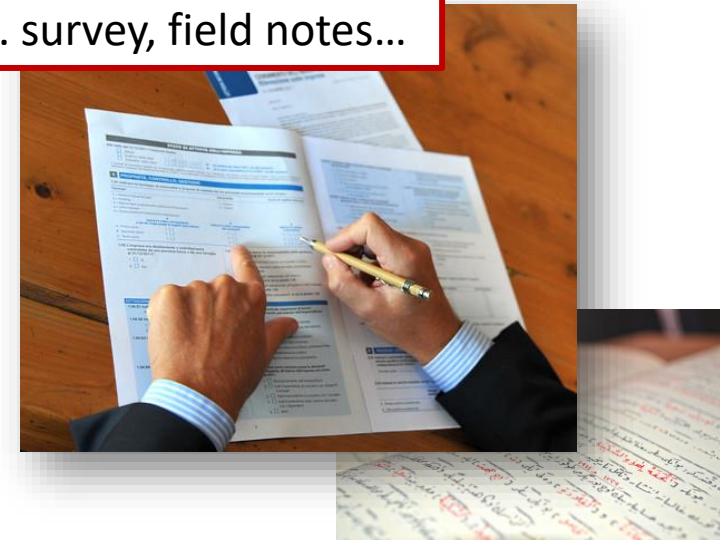
Screenshot da «Broke in China» un cortometraggio comico americano del 1927 diretto da Edward F. Cline e con Ben Turpin. [https://commons.wikimedia.org/wiki/File:Broke_in_China_\(1927\).webm](https://commons.wikimedia.org/wiki/File:Broke_in_China_(1927).webm)

Testi: e.g. survey, field notes..



Audio: e.g. registrazioni di interviste..

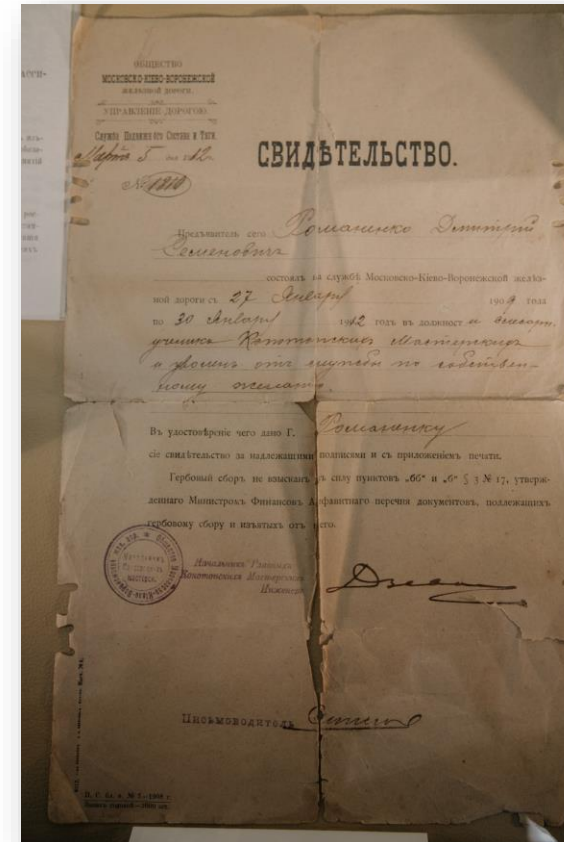
Elke Wetzig. Live interview with Deutschlandfunk at Wikimania. Wikimedia Commons. CC BY-SA 3.0, https://commons.wikimedia.org/wiki/File:Wikimania_interview_dlf3.jpg



Generare o riutilizzare i dati?



Generare nuovi dati
esempio: intervistare delle persone



Riutilizzare dati esistenti
esempio: riutilizzare documenti o
oggetti da un archivio locale

Generare o riutilizzare i dati?

Quando si **riutilizzano i dati**, l'elemento più importante da considerare è la **possibilità di riutilizzarli legalmente**.

Verificate se esistono dati già pubblicati da altri ricercatori che siano utili per la vostra ricerca.

Esempi di archivi in cui è possibile cercare dati da riutilizzare:

- Google Dataset Search:
<https://datasetsearch.research.google.com/>
- Dataverse dell'Università di Harvard:
<https://dataverse.harvard.edu/>
- ...o trovare un archivio disciplinare su re3data.org

The screenshot displays two data records. The top record is titled "Cambridge butterfly wing collection batch 3", published on May 9, 2019, with authors Jiggins, Chris; Montejo-Kovacevich, Gabriela; Warren, Ian; and Wiltshire, Eva. It features a "Rights" panel indicating a Creative Commons Attribution 4.0 International license. The bottom record is titled "Photos de l'association sportive le Goéland Boulonnais" by Florian Lebreton, with a "Files" section showing a "Restricted" status. Red arrows point to the license information in the top record and the restricted status in the bottom record.

Published May 9, 2019 | Version v1

Figure Open

Cambridge butterfly wing collection batch 3

Jiggins, Chris¹; Montejo-Kovacevich, Gabriela¹; Warren, Ian¹; Wiltshire, Eva¹

Rights

Creative Commons Attribution 4.0 International

Photos de l'association sportive le Goéland Boulonnais

Lebreton, Florian¹

Show affiliations

Jeu de données comprenant une photo du président de l'association Le Goéland Boulonnais et deux photos des trophées et prix reçues par l'association.

Les photos ont été prises en mai 2017.

Files

Restricted

The record is publicly accessible, but files are restricted to users with access.

Generare o riutilizzare i dati?

Quando si **generano nuovi dati**, è necessario considerare

- la possibilità di **valorizzazione e sfruttamento commerciale**



Volete brevettare?
Volete pubblicare (ad esempio, una monografia)
e temete che le idee possano essere rubate?



- Deposito di dati sotto embargo
- Accesso controllato



- la **presenza di dati personali/sensibili**, con necessità di gestione della privacy



Lavorate con dati personali?
Lavorate con dati sensibili?



- Consenso informato
- Anonimizzazione
- Accesso controllato
- Comitati di accesso
- Seguire il GDPR

- la sussistenza di **punti di attenzione di tipo etico**.

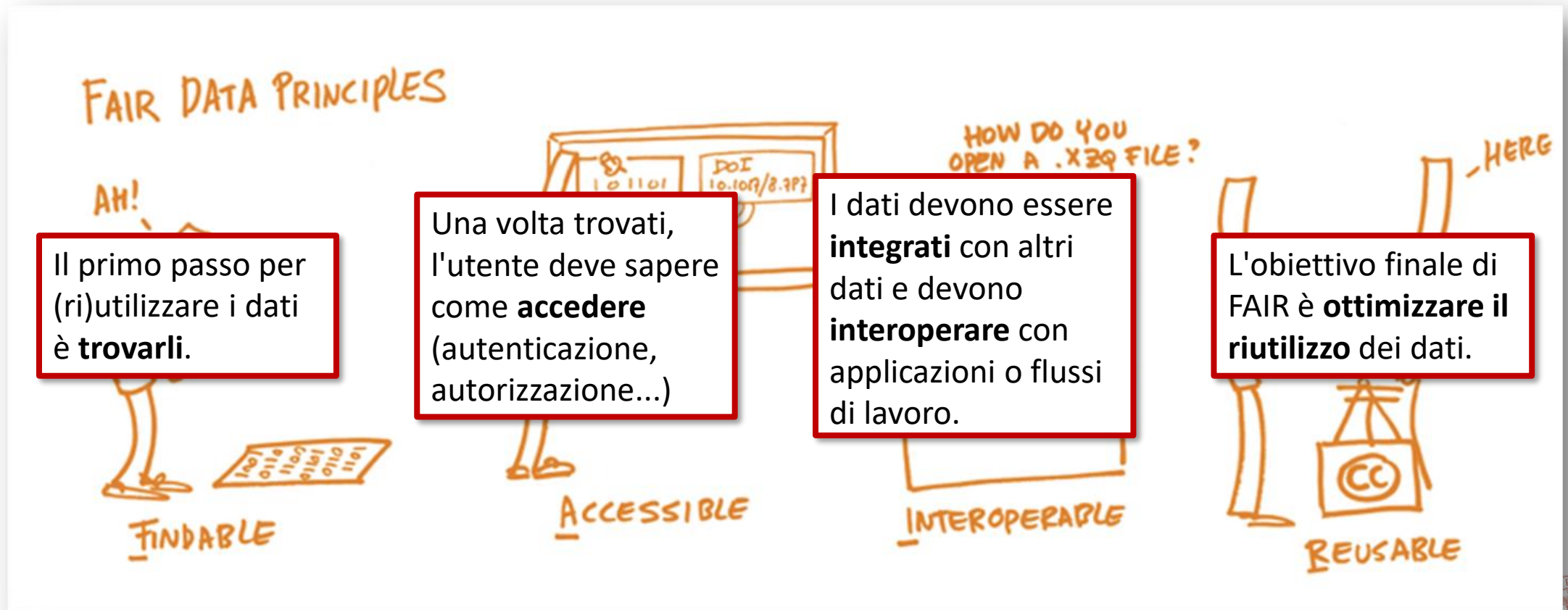


Fate ricerca utilizzando animali?
Fate ricerca sviluppando/utilizzando **Intelligenza Artificiale**?
La ricerca può **influenzare la vita dei soggetti reclutati** (e.g. scoperte accidentali, discriminazioni)?



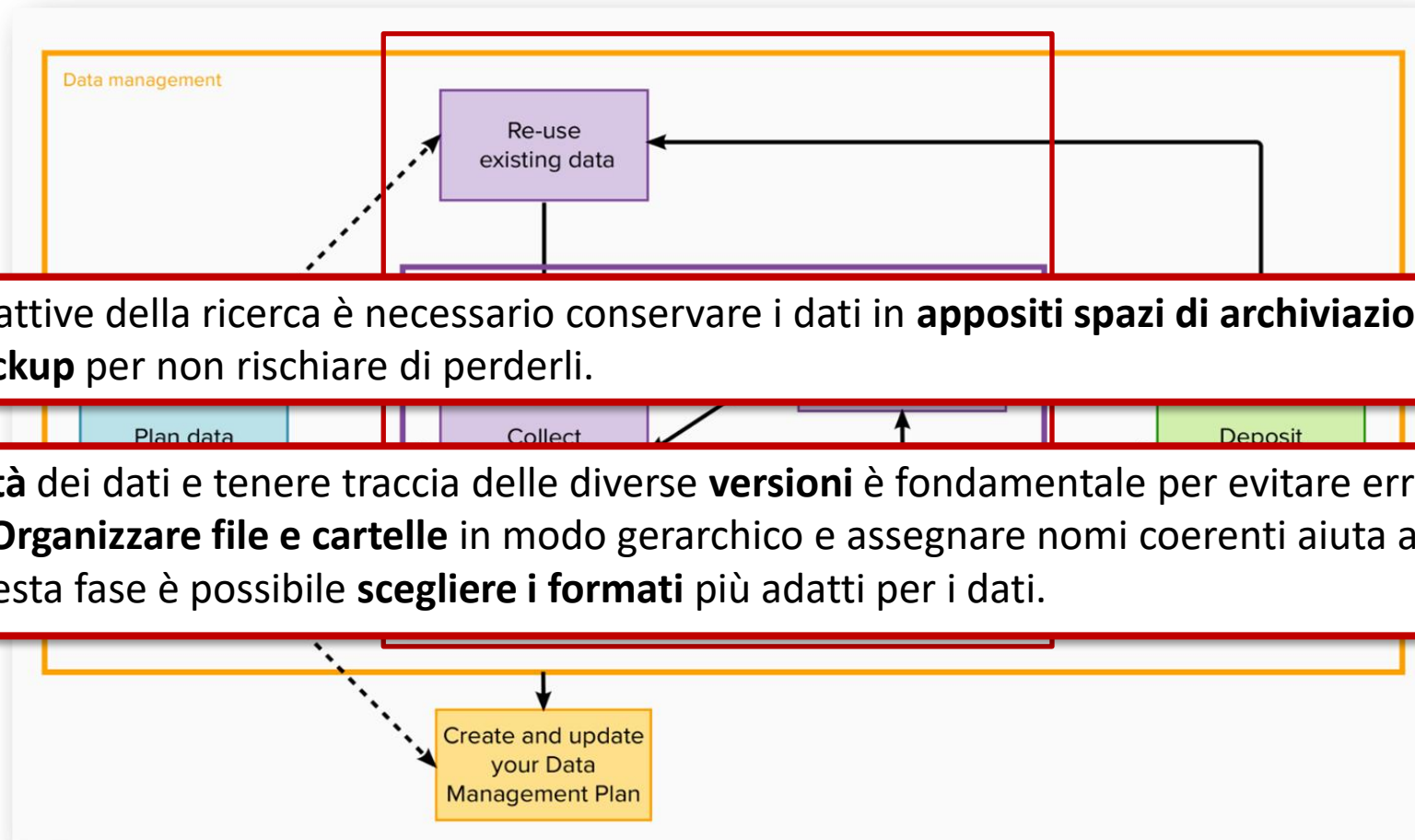
I principi FAIR

L'obiettivo principale è fornire alla comunità scientifica i dati, le metodologie e gli strumenti di ricerca necessari per convalidare (e in alcuni casi replicare) le conclusioni di una ricerca.



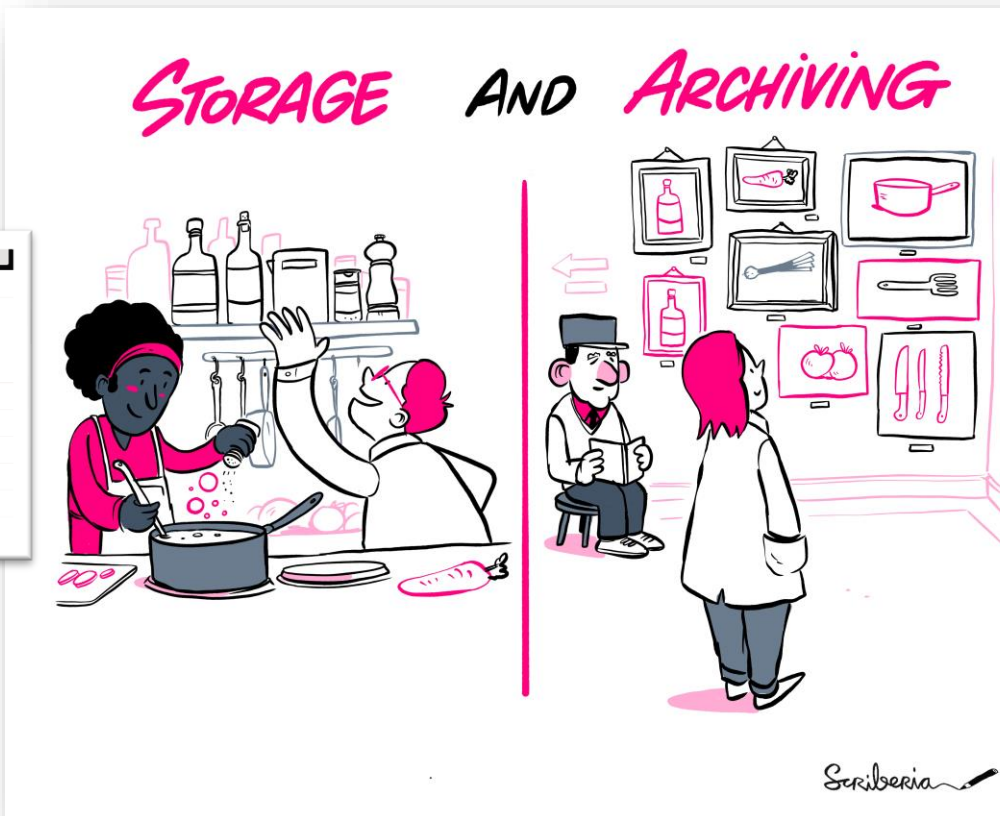
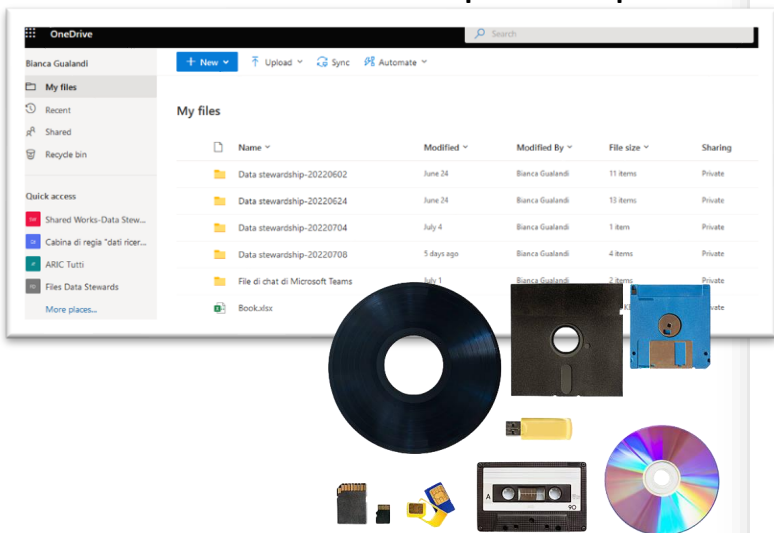
Gestire i dati della ricerca: come fare

FSE 2: gestione attiva

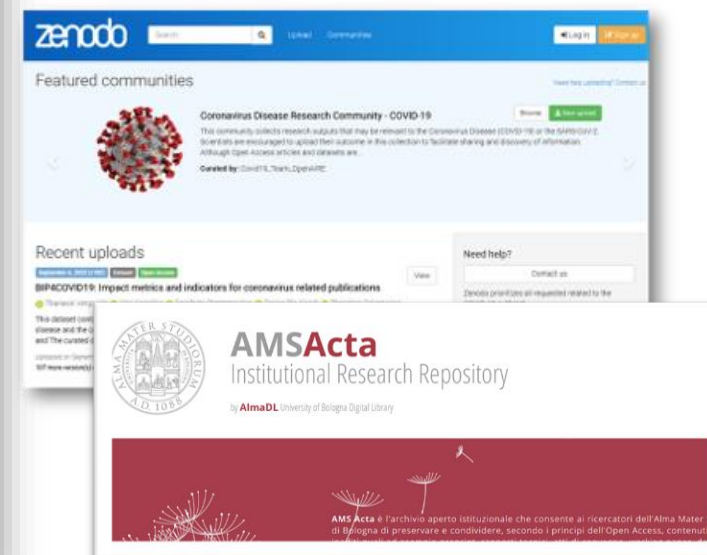


Conservare i dati a breve o a lungo termine

Nella fase di gestione del progetto, subito dopo aver raccolto i dati nuovi o riutilizzati, è necessario archivarli temporaneamente da qualche parte.



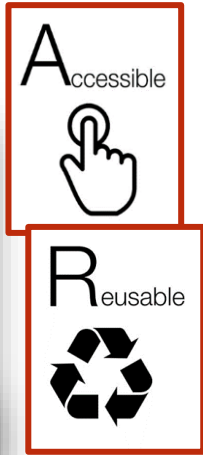
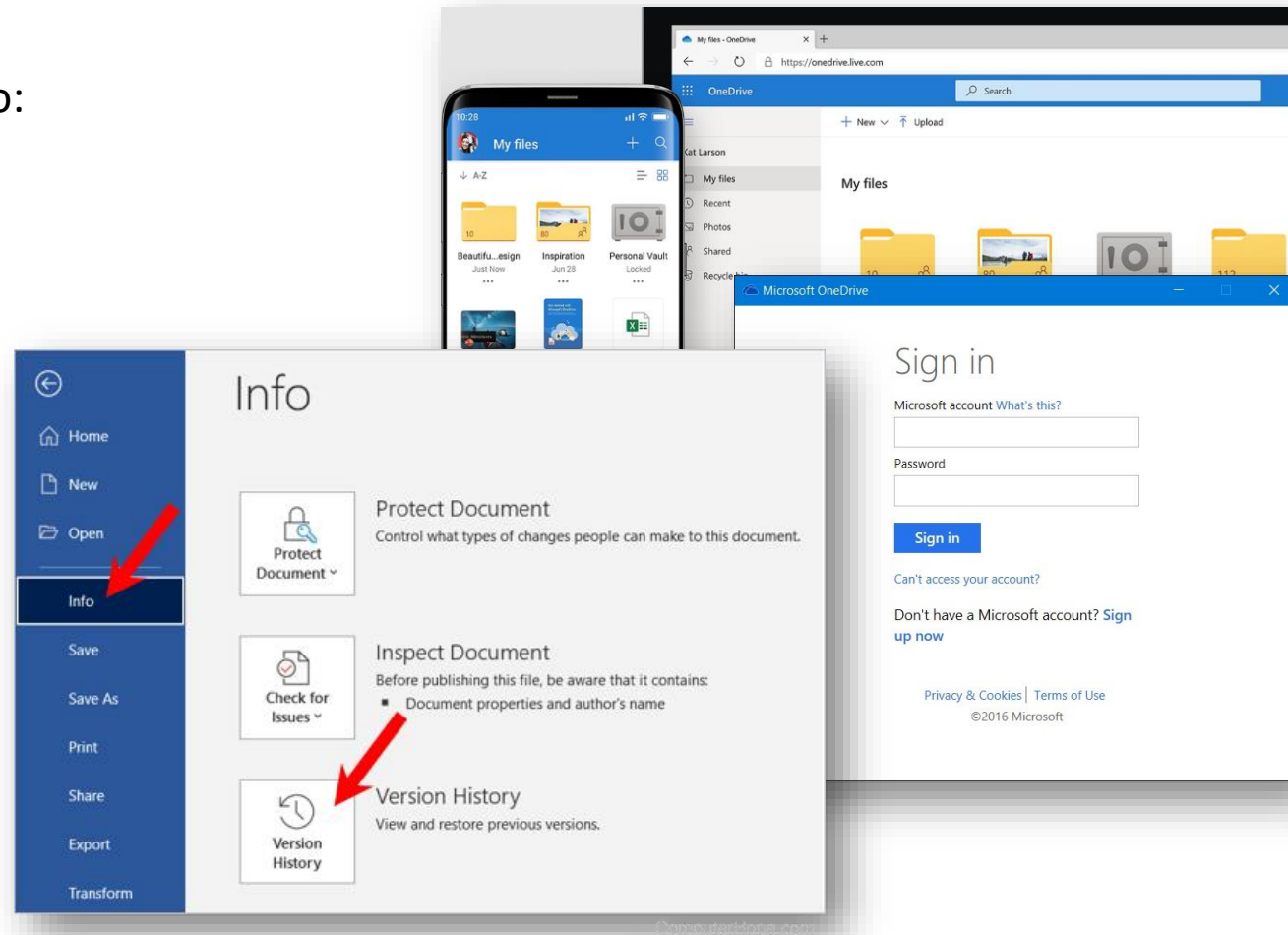
Alla fine del progetto, o quando i dati sono nella loro forma definitiva, è opportuno depositarli per la conservazione a lungo termine.



Storage/conservazione a breve termine

I sistemi di storage affidabili consentono:

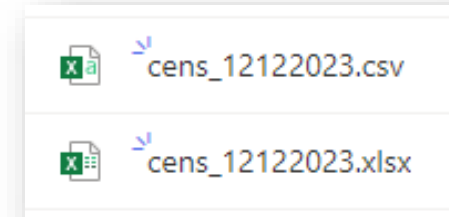
- procedure di **backup automatico**;
- modalità di **accesso controllato** anche in collaborazione (login necessario per accedere ai dati);
- accesso alla **cronologia delle versioni** dei file;
- la **denominazione dei file** e la **strutturazione delle cartelle**.



Curare la raccolta dei dati: organizzazione

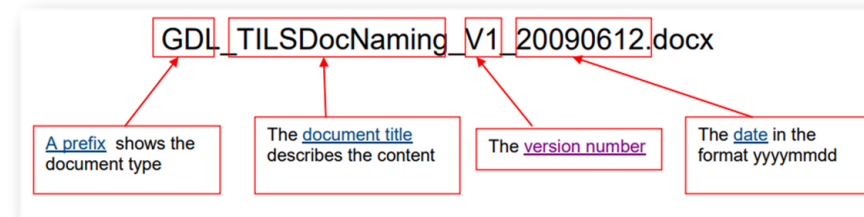
1) Conoscere l'esatto **formato tecnico** del file

È consigliato **esportare tutti i dati da formati proprietari a un formato aperto**.



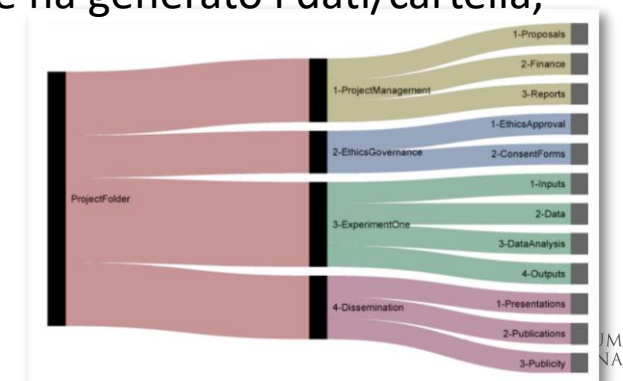
2) Stabilire nomi di file **leggibili/interpretabili**

È consigliato un nome **breve ma descrittivo**, senza caratteri speciali o spazi (usare maiuscole o trattini bassi), includendo **data e numero di versione**.



3) Creare una **struttura di cartelle chiara e organizzata**.

È possibile strutturare le cartelle **cronologicamente**, per **progetto**, in base alla **persona** che ha generato i dati/cartella, in base al **metodo/attrezzatura** di analisi o al **tipo di dati**.



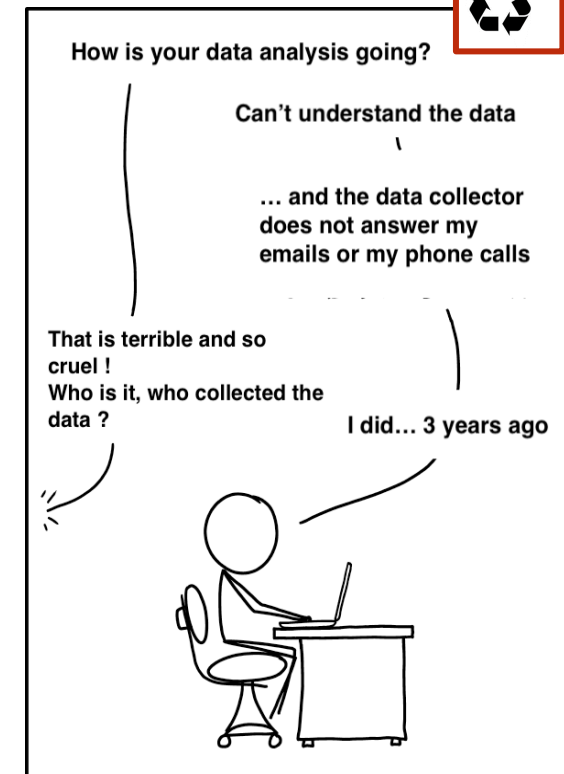
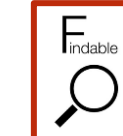
Curare la raccolta dei dati: descrizione

Descrivere adeguatamente i dati li rende più comprensibili, permettendo di capirli e interpretarli anche molto dopo la raccolta indipendentemente dall'aver o meno partecipato al processo di raccolta.

Per migliorare l'interoperabilità e il riutilizzo dei dati, è necessario **raccogliere tutte le informazioni pertinenti di cui gli utenti hanno bisogno** per comprendere e riutilizzare i dati.

Documentazione (in formato human-readable, come un READ ME file) che fornisca informazioni sulla **provenienza** dei dati, sui **metodi di raccolta, riutilizzo o generazione** dei dati e sui **processi di garanzia della qualità** nella generazione dei dati.

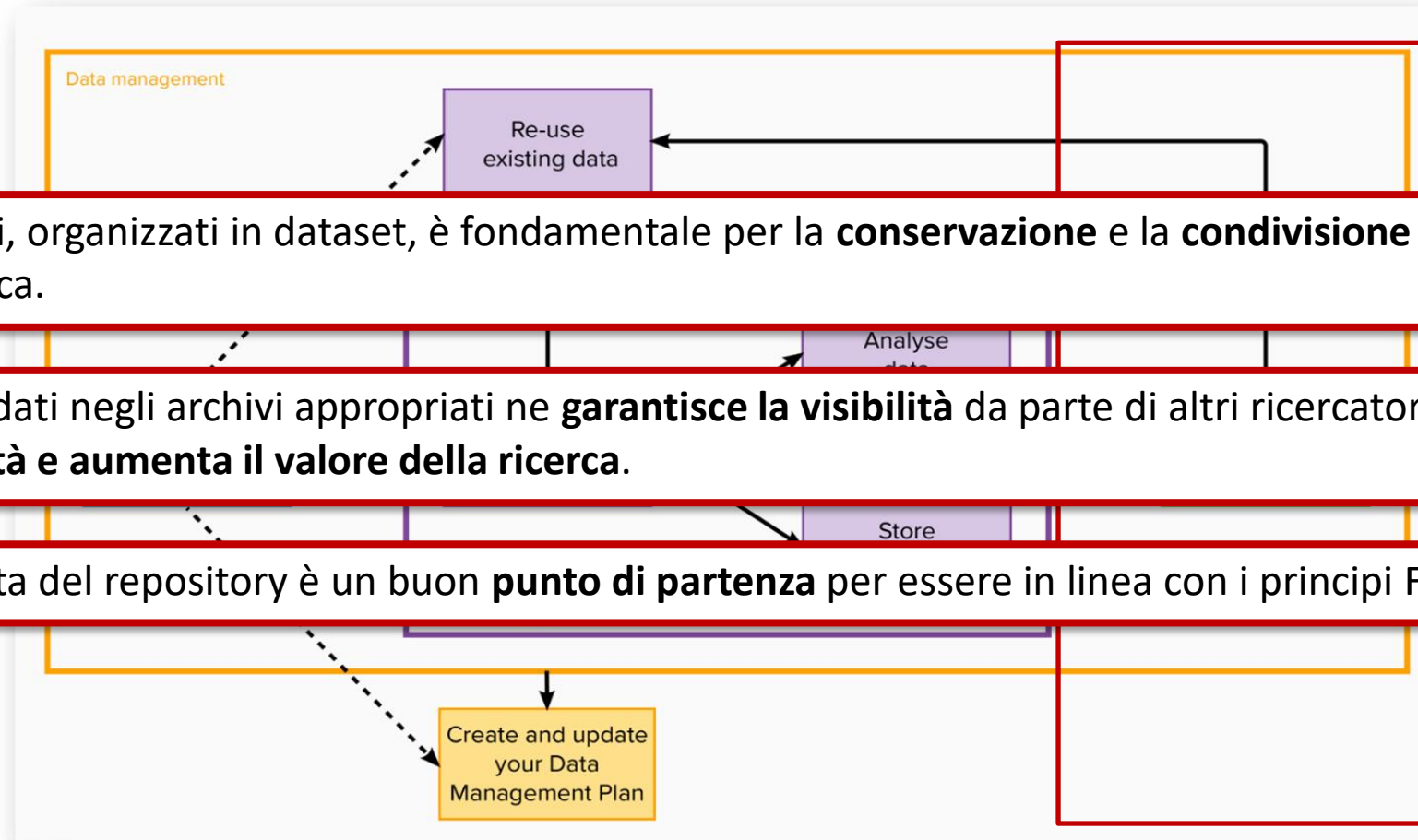
Metadati (in formato machine-readable) che aggiungono ai dati **informazioni** (e.g. le descrizioni delle variabili misurate) e **struttura** (e.g. le diverse variabili misurate), seguendo schemi standard e vocabolari controllati.



**Your first collaborators
are your future selves,
be nice to them !**

Gestire i dati della ricerca: come fare

FASE 3: preservazione



Il deposito dei dati, organizzati in dataset, è fondamentale per la **conservazione** e la **condivisione a lungo termine** dei dati della ricerca.

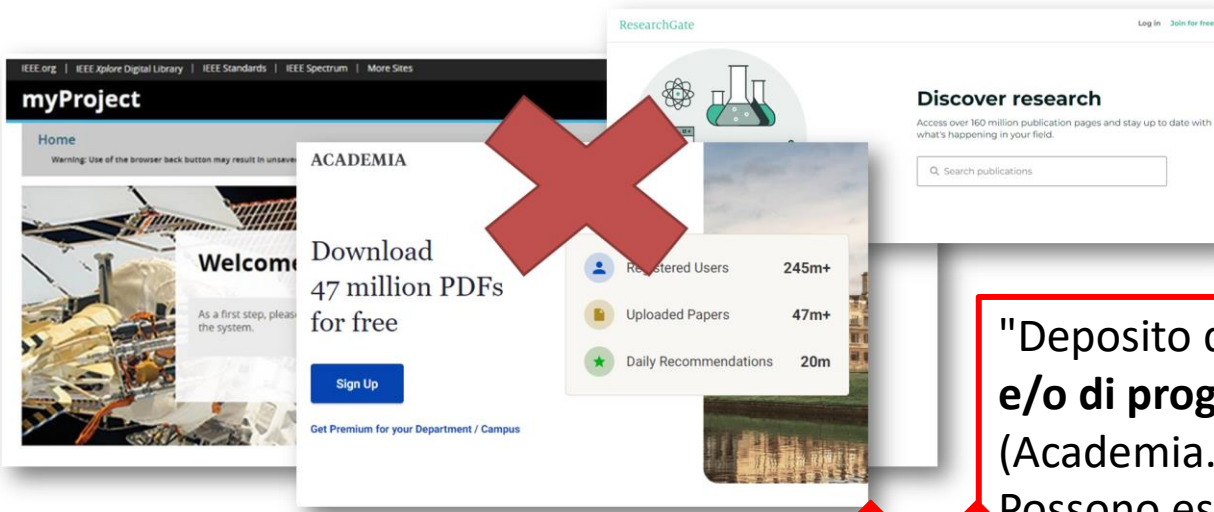
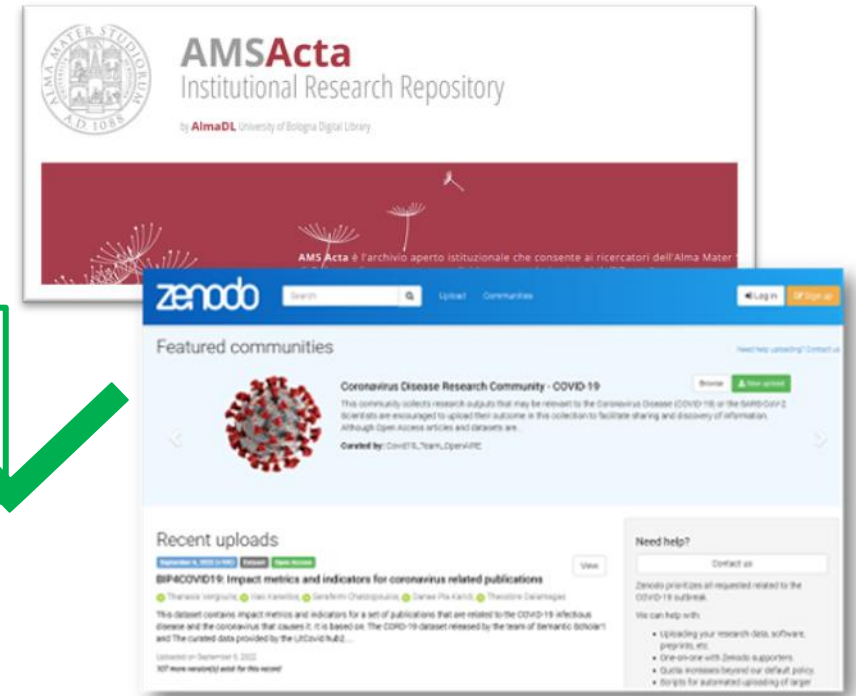
Il deposito dei dati negli archivi appropriati ne **garantisce la visibilità** da parte di altri ricercatori, **promuove la riproducibilità e aumenta il valore della ricerca**.

La scelta del repository è un buon **punto di partenza** per essere in linea con i principi FAIR.



Cos'è il deposito dei dati?

Il "deposito dei dati" è la loro **archiviazione per la conservazione a lungo termine in repositories affidabili**, ossia archivi digitali strutturati per conservare i dati, garantendo che non vadano persi o distrutti nel tempo.



"Deposito dei dati" **non significa caricare i dati su siti web (personali e/o di progetto), sui siti web degli editori o le piattaforme di ricerca (Academia.edu, ResearchGate)!**

Possono essere una buona vetrina per i vostri dati e le vostre pubblicazioni, ma non garantiscono la conservazione a lungo termine.

Come scegliere cosa depositare?

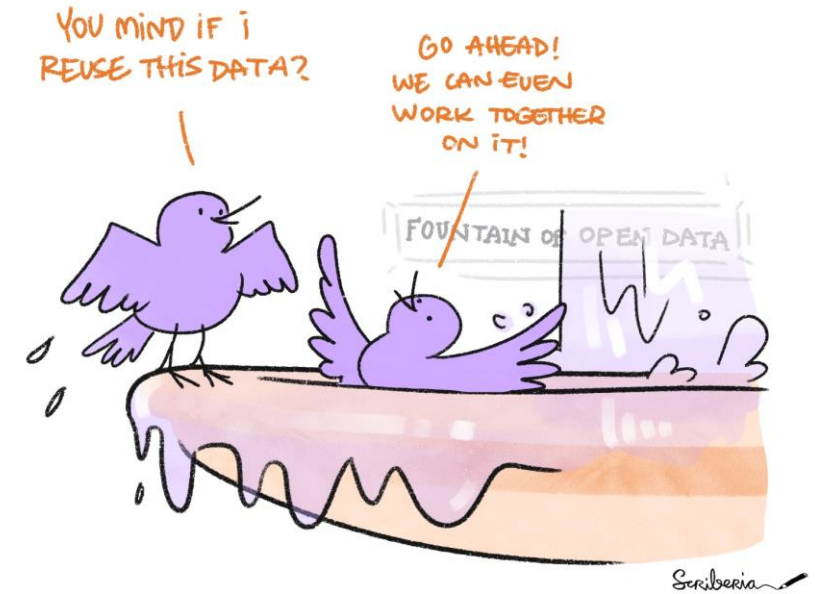
Pensare di depositare tutti i dati associati alla ricerca non è sostenibile: è necessario **valutare quali dati sono necessari** per garantire la comprensione, verifica e riproducibilità della ricerca.

È una valutazione che **può dipendere dall'ambito disciplinare**.

- Dataset e/o codice software originale.
- Dati grezzi ottenuti dall'analisi di campioni fisici.
- Dati osservativi che non possono essere rigenerati.
- Dataset non originali ma non facilmente disponibili
→ necessario avere il permesso di condividere!



- File di output delle analisi se 1) il set di dati e il codice utilizzato per generare l'output sono depositati e 2) la rigenerazione dell'output dai file depositati è abbastanza facile da realizzare.
- Grafici/diagrammi creati a partire dai dati originali e facili da rigenerare.
- Versioni incomplete, non funzionanti o intermedie di codice.
- Dataset conservati e accessibili tramite altre istituzioni o organizzazioni.



Data repository: quali scegliere e dove trovarli

Aggiungendo i vostri dati a un buon repository (archivio) di dati, li renderete facilmente **reperibili** sia per i ricercatori alla ricerca di dati utili sia per i motori di ricerca che mirano a mettere in contatto ricercatori e ricerca.



Utilizzare un registro come re3data per trovare un **repository disciplinare** OPPURE

Utilizzare uno dei **repositories istituzionali** (AMS Acta, AMS Historica) OPPURE

Utilizzare un repository **generalista** come Zenodo.

Data repository: quali scegliere e dove trovarli

The image shows two overlapping screenshots of the AMSActa website. The top screenshot highlights the 'Help & Documentation' link in the navigation menu with a red box and an arrow. The bottom screenshot shows the 'Help and Documentation' page with a list of links for authors and contact information for the support center.

AMSActa
Institutional Research Repository
by **AlmaDL** University of Bologna Digital Library

<https://amsacta.unibo.it/>

AlmaDL

it | en
Statistics Login
BROWSING TOOLS

Search..

i [Help & Documentation](#)

LATEST NEWS

AMSActa is the institutional open access repository which enables the researchers of the Alma Mater Studiorum - University of Bologna to share, preserve and showcase their scientific works making them easily available to the wider research community. The University of Bologna Open Access

i [Help & Documentation](#)

AlmaDL

it | en
Statistics Login
BROWSING TOOLS

Help and Documentation

- [Instructions for authors](#)

Support and Help Center
E-mail: almadl@unibo.it
Phone: (051 20)94318
Fax: (051 20)86041
Monday-Friday, 9.30am - 1.30pm

LATEST ADDITIONS

Masini, Francesca (2024) *NPN di...*
The dataset contains 1297 instances

@almadlunibo

Atom

Rss 1.0

Rss 2.0

AMS Acta
ISSN: 2038-7954
re3data.org - doi.org/10.17616/R3P19R
Managed by **AlmaDL**
[Cookie Settings](#)
[Privacy policy](#)
[Web site Terms](#)
[Repository mission and policies](#)

re3data.org
EUROPEAN UNION
OpenAIRE
OpenDOAR

ALMA MATER STUDIORUM
UNIVERSITY OF BOLOGNA

Data repository: quali scegliere e dove trovarli

The image shows two overlapping screenshots of the AMSActa website. The top screenshot displays the homepage with the URL `https://amsacta.unibo.it/` highlighted in a red box. A red arrow points to the 'Login' link in the top right navigation menu. The bottom screenshot shows the 'Modify the document: Dataset #7508' page. A breadcrumb trail includes 'Type', 'Funding program', 'Upload', 'Details', and 'Deposit'. Below this, there are buttons for 'Save and go back to your Inbox', 'Exit without saving', and 'Next >'. A 'Document type' dropdown menu is open, with 'Dataset' selected and highlighted by a red box. The 'Dataset' option is described as 'A bounded collection of quantitative or qualitative data (e.g. spreadsheet or XML data file) including relating documentation.' Other options include Conference or Workshop Item, Book Section, Monograph, Preprint, Software, Article, and Thesis.

AMSActa
Institutional Research Repository
by **AlmaDL** University of Bologna Digital Library

<https://amsacta.unibo.it/>

AlmaDL

it | en
Statistics Login
BROWSING TOOLS

AlmaDL

it | en
Manage documents Profile Saved searches Statistics giulia.caldoni2@unibo.it Logout
BROWSING TOOLS

Modify the document: **Dataset #7508**

Type → Funding program → Upload → Details → Deposit

Save and go back to your Inbox Exit without saving Next >

Document type

- Dataset**
A bounded collection of quantitative or qualitative data (e.g. spreadsheet or XML data file) including relating documentation.
- Conference or Workshop Item**
A paper, poster, speech, lecture or presentation given at a conference, workshop or other event. If the conference item has been published in a journal or book then please use "Book Section" or "Article" instead.
- Book Section**
A chapter or section in a book, an entry in an encyclopedia, indexes, concordances, a preface or an introduction to a volume.
- Monograph**
An essay, a volume of proceedings, a technical report, a project report, documentation, a manual, an anthology, a working paper or discussion paper, a catalogue, a translation or a critical edition.
- Preprint**
A not yet published contribution or manuscript in its pre-publication form.
- Software**
A computer program in source code (text) or compiled form
- Article**
[NO LONGER SELECTABLE] An article in a journal, magazine, newspaper. Not necessarily peer-reviewed. It may be in electronic-only format, such as an online journal or news website.
- Thesis**
[NO LONGER SELECTABLE] A thesis or dissertation.

LATEST AD
Masini, France
The dataset co

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
A.D. 1088

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
A.D. 1088

Caratteristiche di un repository: PID

Un identificatore persistente (PID) è un **referimento unico e univoco di lunga durata a una risorsa**.

`https://doi.org/10.5281/zenodo.5018216`

Resolver Service

Directory
indicator

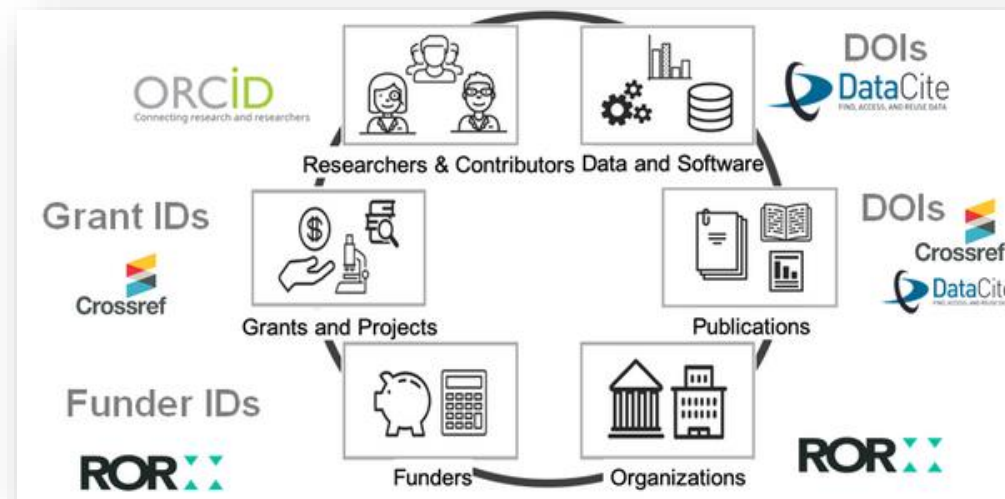
Prefix

Suffix

Molti repository assegnano un PID quando un oggetto viene depositato. Il PID può essere collegato a un insieme di metadati che descrivono un oggetto piuttosto che all'oggetto stesso.

La risorsa può essere:

- un **oggetto** (pubblicazioni, dati, software: URN, DOI, ARK, Handle);
- una **persona** (ricercatori, autori, collaboratori: ORCID, ISNI)
- un'**organizzazione** o un'**istituzione** (università, enti di finanziamento, organizzazioni di ricerca: ROR)



Lo scopo principale del PID è quello di fornire le informazioni necessarie per identificare, verificare e localizzare in modo affidabile l'oggetto.

Data repository: licenze

È opportuno depositare i dati (e le pubblicazioni) associandoli a **licenze chiare e permissive**, così che chi vuole riutilizzare i dati abbia chiaro **quali usi sono consentiti**.

Se i dati non sono accompagnati da una licenza, è necessario **ottenere l'autorizzazione** del proprietario per copiare, condividere o pubblicare in altro modo i dati.

! Anche se si depositano i dati con una licenza considerata "aperta" (CC BY o equivalenti), **rimane l'obbligo per chi li riutilizza di citare l'autore del set di dati**.

! In alcuni casi, la scelta della licenza da associare a dati e/o pubblicazioni è dettata da accordi presi con terze parti/enti di finanziamento → esempio sono le direttive di Horizon Europe

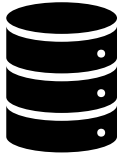
CREATIVE COMMONS LICENSES		COPY & PUBLISH	ATTRIBUTION REQUIRED	COMMERCIAL USE	MODIFY & ADAPT	CHANGE LICENSE
	PUBLIC DOMAIN	✓	✗	✓	✓	✓
	CC BY	✓	✓	✓	✓	✓
	CC BY-SA	✓	✓	✓	✓	✗
	CC BY-ND	✓	✓	✓	✗	✗
	CC BY-NC	✓	✓	✗	✓	✓
	CC BY-NC-SA	✓	✓	✗	✓	✗
	CC BY-NC-ND	✓	✓	✗	✗	✗

You can redistribute (copy, publish, display, communicate, etc.)
 You have to attribute the original work
 You can use the work commercially
 You can modify and adapt the original work
 You can choose license type for your adaptations of the work.

Per determinare la licenza più appropriata si può usare un selezionatore di licenze:

- per i dataset <https://creativecommons.org/choose/>
- per il software <http://ufal.github.io/public-license-selector/>

Riassumendo: alcuni punti chiave



Ogni tipo di ricerca genera o riutilizza dati.
Definire cos'è il dato dipende dall'area disciplinare di appartenenza.



Il dato è un asset di valore per il ricercatore e la comunità.
Per questo va curato attraverso una gestione responsabile e consapevole.



Pianificare la gestione del dato è strumentale per affrontare i necessari punti di attenzione al momento giusto. La strada al contrario non è sempre percorribile.



Etica, privacy e proprietà intellettuale devono essere affrontate in ogni fase del ciclo di vita dei dati e del progetto.



Scegliere cosa depositare è responsabilità del ricercatore, che sa quali dati sono fondamentali per la riproducibilità della sua ricerca.



La scelta del repository può influenzare la gestione dei dati.
Il deposito garantisce la conservazione a lungo termine, aumentata visibilità e citabilità.

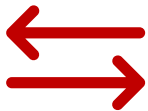
Il supporto disponibile a UniBo



Per supporto riguardo la **gestione dei dati di ricerca** potete rivolgervi a:
aric.datasteward@unibo.it



Per supporto all'utilizzo dei **repository di ateneo** potete rivolgervi a:
almadl@unibo.it, historica@unibo.it



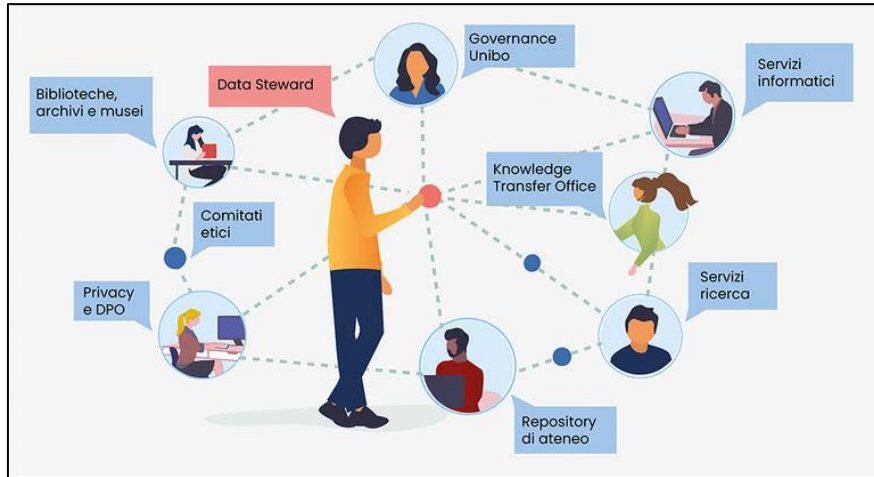
Per supporto riguardo la **proprietà intellettuale** dei risultati di ricerca UniBo potete rivolgervi a:
kto@unibo.it



Per supporto sulle questioni relative alla **privacy** potete rivolgervi a:
privacy@unibo.it



Il supporto disponibile a UniBo: i data stewards



Chi sono?

- **Esperti di dominio**
- Competenti nella **pianificazione e gestione responsabile dei dati**
- **Punto di contatto unico** per indicazioni sulla gestione dei dati

Cosa fanno?



Consulenza su pratiche di corretta gestione dei dati della ricerca durante tutto il processo di ricerca



Supporto alla stesura del Data Management Plan

→ livello di supporto differenziato in base agli obblighi di progetto, e.g. Horizon Europe

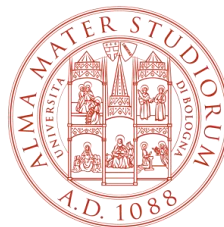


Eventi di (in)formazione per i ricercatori dell'Ateneo

→ Corso per dottorandi «Research data management e principi di Open Science» in partenza a marzo 2024



Partecipazione a network nazionali e internazionali per il posizionamento dell'Ateneo nel contesto Open Science.



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Giulia Caldoni

Area della Ricerca (ARIC)

Alma Mater Studiorum – Università di Bologna

aric.datasteward@unibo.it

giulia.caldoni2@unibo.it

www.unibo.it