

GESTIRE I DATI DELLA RICERCA: tutto ciò che c'è da sapere



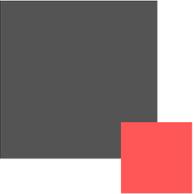
LEZIONE 1

Sara Coppini
sara.coppini2@unibo.it

Data Steward area tecnologica, Research Services Division (ARIC), Alma Mater Studiorum - Università di Bologna



06 Ottobre 2023



DURANTE L'INCONTRO DI OGGI...

- Per favore tenete i microfoni spenti durante la presentazione.
- Sentitevi liberi di accendere microfono e videocamera nei momenti di Q&A.
- Sentitevi liberi di scrivere le vostre domande in chat in qualsiasi momento, risponderemo durante il Q&A!
- Abbiamo previsto all'interno di questa lezione due pause, che seguiranno i momenti Q&A.

GESTIRE I DATI DELLA RICERCA: TUTTO CIÒ CHE C'È DA SAPERE



Modulo formativo diviso in tre incontri.

Unico pre-requisito è avere una propria esperienza di ricerca.



LE TEMATICHE CHE AFFRONTEREMO OGGI

1

I DATI DELLA RICERCA

Definizioni e esempi di dominio.

2

IL VALORE DEI DATI E DELLA LORO CORRETTA GESTIONE

Horror stories e vantaggi associati alla gestione del dato.

3

OPEN SCIENCE

Il contesto culturale e politico in cui si muove la ricerca.

4

PRINCIPI FAIR NELLA GESTIONE DEI DATI DELLA RICERCA

Cosa sono e come applicarli nella pratica di ricerca.

CHI SIAMO? DATA STEWARDS @UNIBO



Il progetto Data Steward @Unibo mira al rafforzamento del supporto alla gestione dei dati della ricerca:

- Chi sono i Data Stewards?

Figure di supporto per le tematiche di gestione FAIR dei dati della ricerca e stesura del Data Management Plan ai team ricerca di UNIBO

- Cosa fanno?

Supportano i ricercatori nella gestione FAIR dei dati (research data management) e nella stesura del Data Management Plan, principalmente nel contesto dei progetti Horizon Europe

Supportano la Governance di Ateneo nella promozione di Open Science

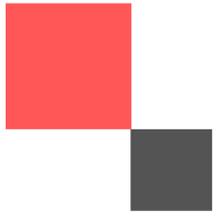
- Qual è il loro profilo?

Background scientifico (esperti di dominio)

Esperienza in data management

FAIR principles e conoscenza di pratiche di OS





SARA COPPINI

Area tecnologica

BIANCA GUALANDI

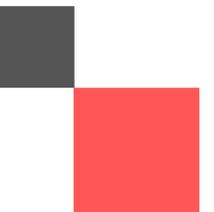
Area umanistica

MARIO MARINO

Area sociale

GIULIA CALDONI

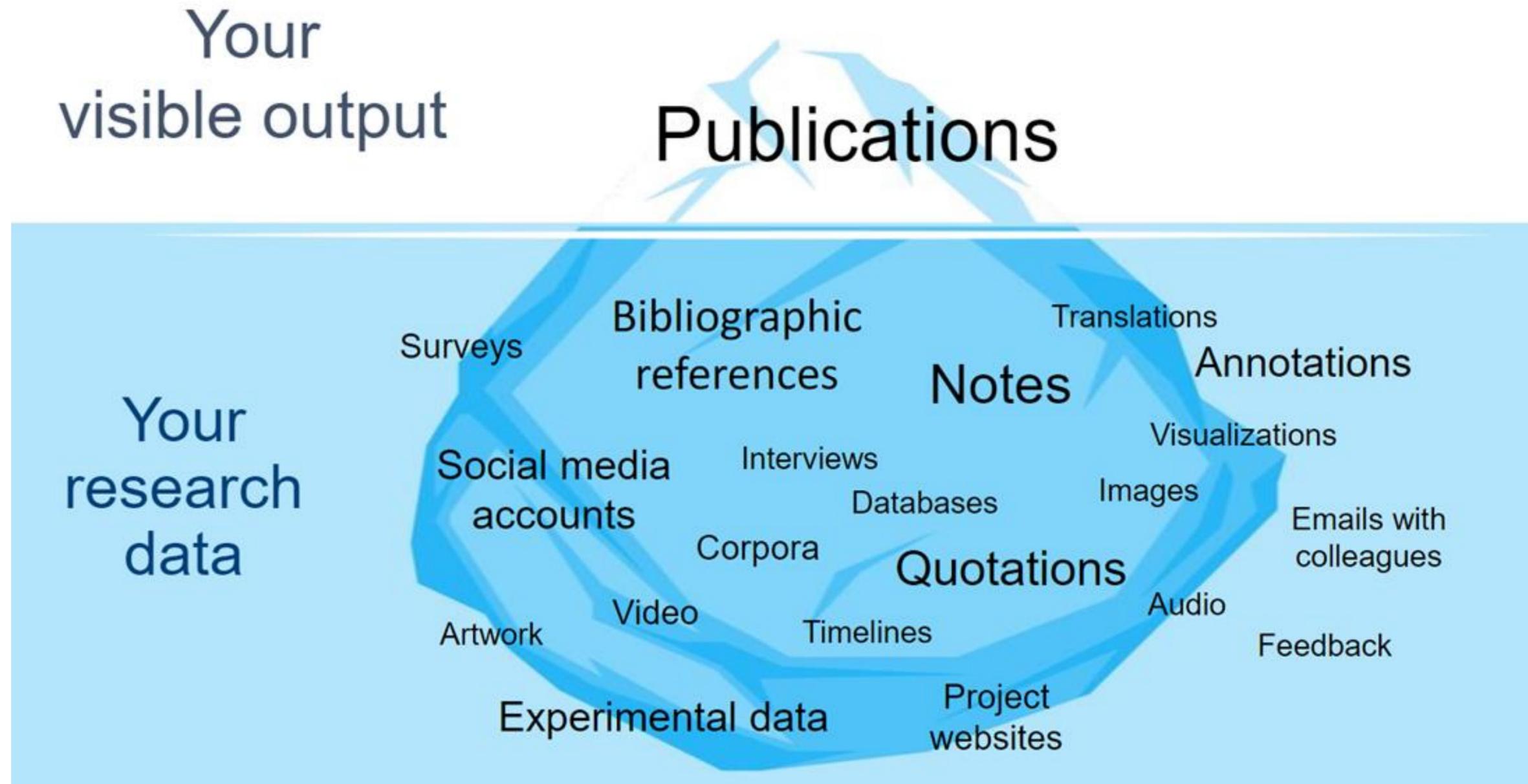
Area biomedica



INIZIAMO A PARLARE DI DATI...



INIZIAMO A PARLARE DI DATI...





QUALCHE DEFINIZIONE UFFICIALE





Informazioni, in particolare fatti o numeri, raccolte per essere esaminate e considerate come base per ragionamenti, discussioni o calcoli.





I record fattuali (numeri, testi, immagini e suoni) utilizzati come fonti primarie per la ricerca scientifica e che sono comunemente accettati dalla comunità scientifica come necessari per convalidare i risultati delle ricerche.





L'intersezione tra scienza dell'informazione e tecnologia ha cambiato la natura e il significato del dato: tutto ciò che si può rappresentare sotto forma di bits è dato.





DIVERSI TIPI DI DATI DELLA RICERCA

Esiste un'enorme varietà di tipi di dati ed è possibile classificarli in modi diversi

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



QUALI TIPI DI DATI POTRESTE INCONTRARE DURANTE LA VOSTRA RICERCA?

Dipende dal vostro ambito di
ricerca!

Oggi vediamo degli esempi collegati a...

- Codice
- Dati tabulari
- Prototipi digitali (modelli 3D)
- Database

IL CODICE

Il codice sorgente di un software è un peculiare tipo di dato, in quanto eseguibile e legato ad altri dati di input e output che vengono elaborati

È concepibile come dato semplicemente in quanto **asset e oggetto principale di output e studio di una ricerca**

Formati: generalmente **file di testo** la cui estensione dipende dal linguaggio di programmazione utilizzato (py, js, php...), raramente anche in file di testo generico, con estensione txt

Buone pratiche:

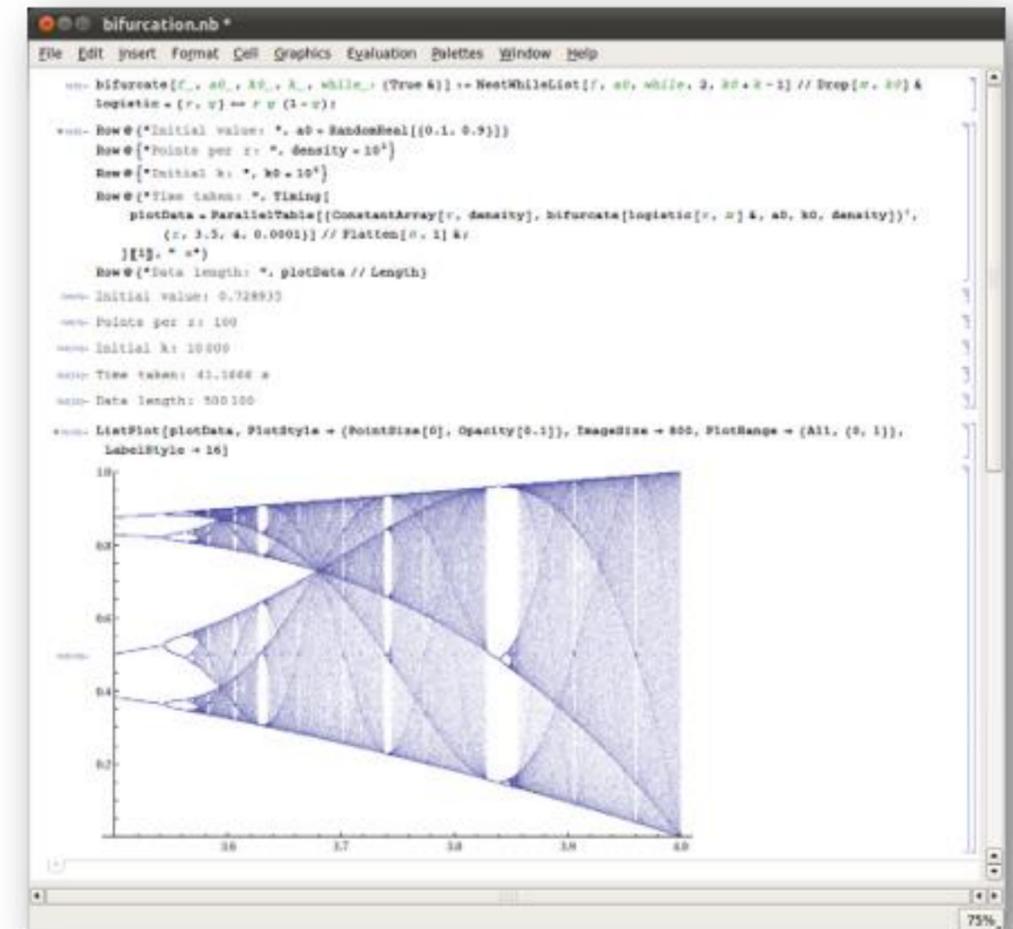
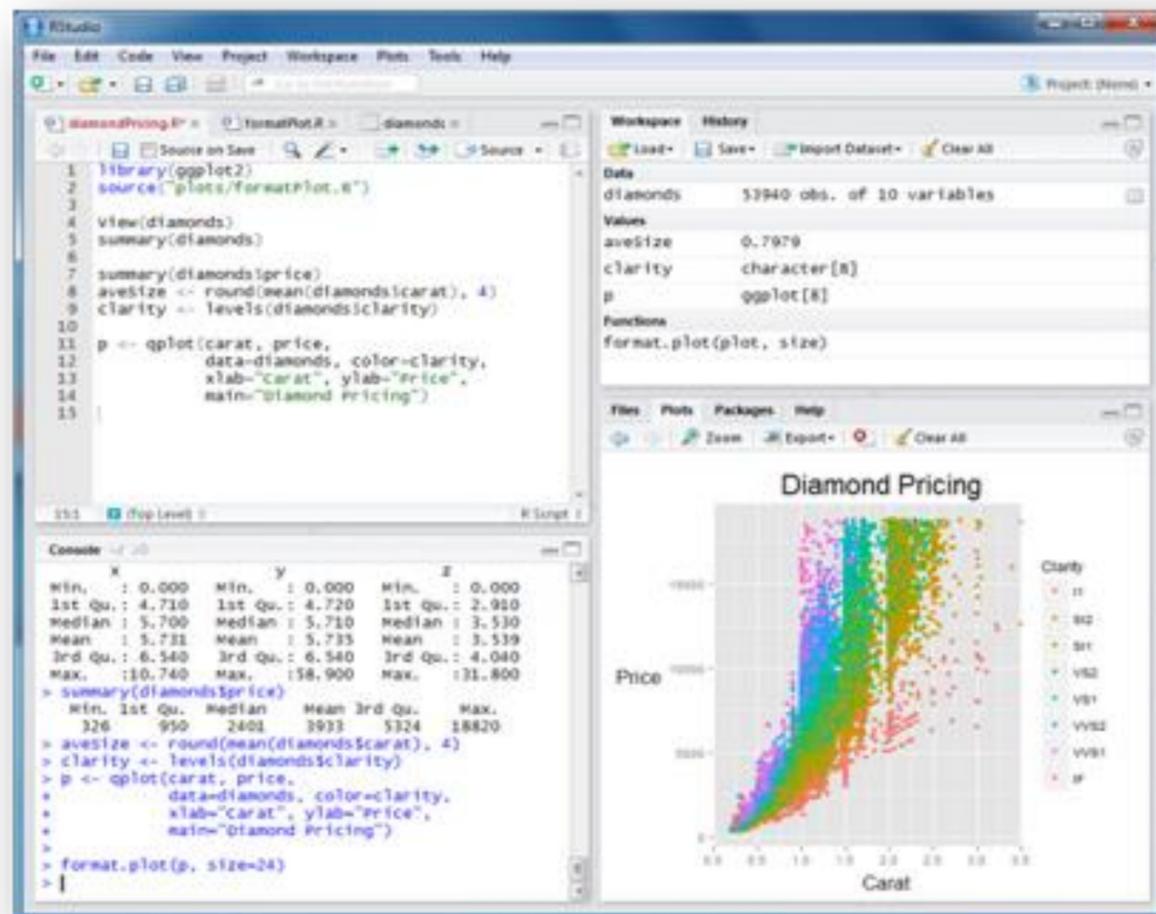
- Appropriata **metadattazione e file citazionali**
- Deposito in **repository specifici per il codice** come *Software Heritage* (<https://www.softwareheritage.org/>)
- Documentazione:
 - **File esterni** a quello di codice come documentazione (significato variabili, struttura del software ecc.)
 - **All'interno** del file di codice, commentare il codice stesso

```
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 = "${acc.output}\n\n${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
    }
  }
}
```

IL CODICE

Tools per lavorare col software in modo Open e FAIR:

- **Container e ausili per container**, come Docker e ReproZip
- Ambienti di sviluppo ed **editor di codice online** che permettono più flessibilità e collaborazione rispetto agli IDE locali. Ad esempio GitHub Codespaces, Gitpod, DataCamp Workspace
- Servizi cloud-based con specifiche di ambiente che li rendono compatibili o complementari al software per l'esecuzione di container, come i **Notebook cloud**



DATI TABULARI

Cosa sono: dati strutturati qualitativi o quantitativi, organizzati in **tabelle**

Formati più diffusi: **XLSX, CSV, TSV, ...**

Alcuni punti di attenzione nella gestione dei dati tabulari:

- **Volume** dei dati
- Utilizzo preferibilmente di **formati** aperti e/o non-proprietari
- Sistema di **codifica & standardizzazione** dei dati
- **Documentazione** sul significato dei campi

	A	B	C	D	E	F	G	H	I	J	K		
1	輪 ?? ?	銖 ???? ?	輪 ?? ?	銖餉?甄	輪 ?? ?	??	??	??	??	??	??		
2	? ?	閻 csv	http://www	rawData	? ?	閻	盞 搬 瘥 迺	?踰?輪	? http://data.g?	祥	UTF-8;		
3	? ??	刻?掛http://www	rawData	? ??	刻?掛? ?	閻	?踰?輪	? http://data.g?	祥	UTF-8;	?音 ??? 02-23566461"		
4	? ??	刻?掛http://www	other	? ??	刻?掛? ?	閻	?踰?輪	? http://data.g?	祥	UTF-8;	?音 ??? 02-23566461"		
5	編輯?媛	xml	http://www	rawData	憶折 ?	? ?	閻	?踰?輪	? http://data.g?	祥	UTF-8;	?音 ??? 02-2	
6	? 媛	xml	http://www	rawData	?刻 ?	??	閻	?踰?輪	? http://data.g?	祥	UTF-8;	?音 ??? 02-2	
7	? ?	閻 csv	http://www	rawData	? ?	閻	揮 ?? ?	1橋? ?	?踰?輪	? http://data.g?	祥	UTF-8;	?音 ??? 02-2
8	3+2	蜚潮	http://down	rawData	?祈??	?銖	銖剛 ?	菜銖	???,??	http://data.g?	祥	UTF-8;	銖音 ??? 0800-700-365
9	3+2	蜚潮	http://down	rawData	?祈??	?銖	銖剛 ?	菜銖	???,??	http://data.g?	祥	UTF-8;	銖音 ??? 0800-700-365
10	蜚???	http://down	rawData	銖剛 ?	菜3	蜚澆??	銖剛 ?	菜銖	???,??	http://data.g?	祥	UTF-8	銖音 ??? 08
11	銖剛 ?	菜txt	http://www	rawData	銖剛 ?	菜?菟???	銖	???,??	http://data.g?	祥	UTF-8	銖音 ??? 08	
12	?箇 ?	?csv	http://117.5	rawData	?祈??	?銖	?箇 ?	瘥 僑	?踰?輪	? http://data.g?	祥	UTF-8;	
13	?箇 ?	?http://117.5	rawData	?祈??	?銖	?箇 ?	瘥 僑	?踰?輪	? http://data.g?	祥	UTF-8;	? ???,04-2	
14	?箇 ?	?csv	http://117.5	rawData	?祈??	?銖	?箇 ?	瘥 僑	?踰?輪	? http://data.g?	祥	UTF-8;	
15	?箇 ?	?csv	http://117.5	rawData	?祈??	?銖	?迂?	瘥 迺	?踰?輪	? http://data.g?	祥	UTF-8;	
16	?箇 ?	?csv	http://117.5	rawData	?祈??	?銖	?箇 ?	瘥?橋? ?	?踰?輪	? http://data.g?	祥	UTF-8;	霏 ?閻?,0
17	銖剛 瘥	xml	http://www	rawData	? ?	撒 ?	憶 漱??, ?	靛 ??	?踰?輪	? http://data.g?	祥	UTF-8;	蠟 ???,02
18	??	??	??	??	??	??	??	??	??	??	??	??	??

```
File Edit Selection View Go Run ... [Extension Development Host] rainbow... X
movies.csv X
C:\wsl\share > movies.csv > data
1 original_title,year,duration,genre,director,avg_vote
10 ATLANTIC,1957,154,Drama,Romance,James Cameron,7.8
17 Blade Runner,1982,117,"Action, Sci-Fi, Thriller",Ridley Scott,8.1
18 The Matrix,1999,136,"Action, Sci-Fi","Lana Wachowski, Lilly Wachowski",8.7
19 Equilibrium,2002,107,"Action, Drama, Sci-Fi",Kurt Wimmer,7.4
20 Groundhog Day,1993,101,"Comedy, Fantasy, Romance",Harold Ramis,8.0
21 Indiana Jones and the Last Crusade,1989,127,"Action, Adventure",Steven Spielberg,8.2
22 Alien,1979,117,"Horror, Sci-Fi",Ridley Scott,8.4
23 Back to the Future,1985,116,"Adventure, Comedy, Sci-Fi",Robert Zemeckis,8.5
24 Judge Dredd,1995,96,"Action, Crime, Sci-Fi",Danny Cannon,5.6
25 Corpse Bride,2005,77,"Animation, Drama, Family",Tim Burton, Mike Johnson,7.3
26 Men in Black,1997,98,"Action, Adventure, Comedy",Barry Sonnenfeld,7.3
27 Star Wars: Episode VI,1983,131,"Action, Adventure, Fantasy",Richard Marquand,8.3
28 Star Wars: Episode III,2005,140,"Action, Adventure, Fantasy",George Lucas,7.5
29 Saving Private Ryan,1998,169,"Drama, War",Steven Spielberg,8.6
30 Desperado,1995,104,"Action, Crime, Thriller",Robert Rodriguez,7.2
31 Alita: Battle Angel,2019,122,"Action, Adventure, Sci-Fi",Robert Rodriguez,7.3
32 The Prestige,2006,130,"Drama, Mystery, Sci-Fi",Christopher Nolan,8.5
33 Dumb and Dumber,1994,107,Comedy,"Peter Farrelly, Bobby Farrelly",7.3
34 The Dark Knight,2008,152,"Action, Crime, Drama",Christopher Nolan,9.0
35 Casino Royale,2006,144,"Action, Adventure, Thriller",Martin Campbell,8.0
36 From Dusk Till Dawn,1996,108,"Action, Crime, Horror",Robert Rodriguez,7.2
37 Match Point,2005,124,"Drama, Romance, Thriller",Woody Allen,7.6
```

PROTOTIPI DIGITALI E MODELLI 3D

Prototipi di varie discipline: impianti, sistemi ecc.

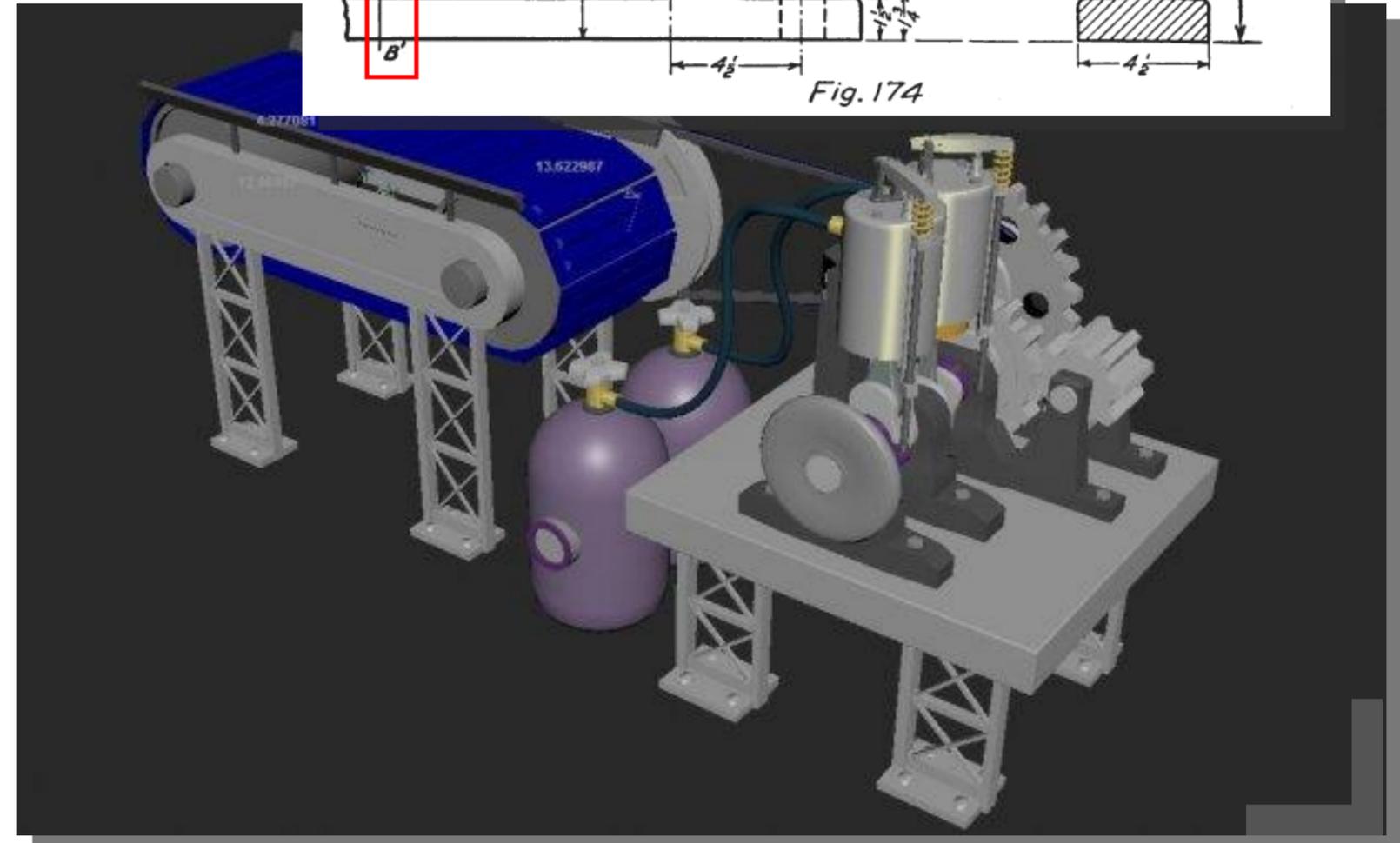
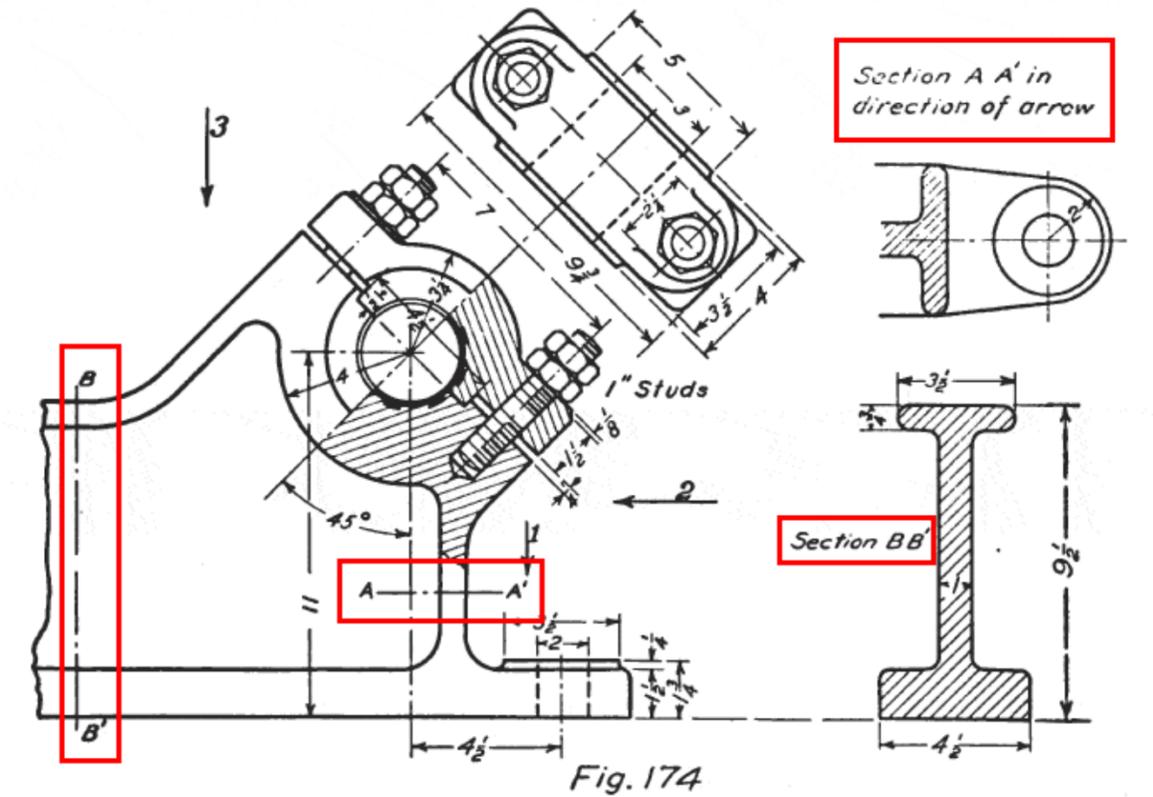
Alcuni punti di attenzione:

- **Documentazione** del modello se distribuito in formato proprietario
- **Minimizzare** le complessità strutturali non necessarie (gerarchie, risoluzione, dati nascosti, textures e superfici...)
- Considerare gli **obiettivi** di progetto (nella scelta dei formati, scelta degli standard...)

Alcuni formati maggiormente consigliabili: Adobe Illustrator (.ai), dati CAD (.dxf, .dwg o .svg), formati binari dei pacchetti GIS e CAD.

Formati aperti:

- DAE - A 3D model/scene format standardized by Khronos.
- glTF - A 3D model/scene format standardized by Khronos.
- VRML/X3D – realtime 3D data formats standardized by ISO/IEC



IL DATABASE

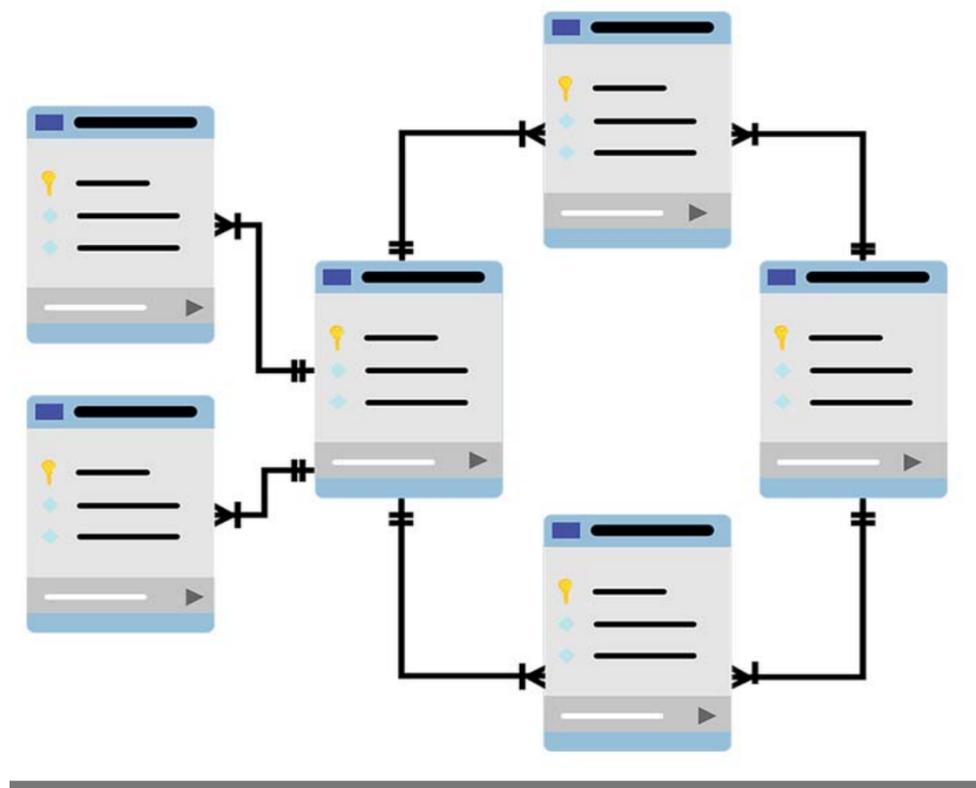
Anche se il concetto di base di dati sembra in opposizione a quello di semplice dato, possiamo considerarlo comunque come un **asset di ricerca da valorizzare e curare nel modo migliore** come qualsiasi altro "dato" in senso stretto.

I database più comunemente utilizzati sono:

- Database gerarchici
- Database di rete
- Database orientati agli oggetti
- Database relazionali
- Database NoSQL

Alcune buone pratiche nella progettazione dei database sono:

- Prediligere la **semplicità**: utilizzare standardizzazioni (ad esempio, le convenzioni di denominazione); considerare le modifiche future; prediligere la semplicità tecnica.
- **Normalizzare** i dati: evitare la ridondanza, verificare le dipendenze, prevenire le anomalie, ecc.
- Considerare le **condizioni di funzionamento**: progettare per il lungo termine, creare e fornire documentazione; diagramma dei dati, ecc.
- **Modellare** i dati in modo efficace: ridurre al minimo la manutenzione non necessaria; utilizzare la visualizzazione e testare il database.



ALTRI OUTPUT DELLA RICERCA: GLI OUTPUT NON DIGITALI

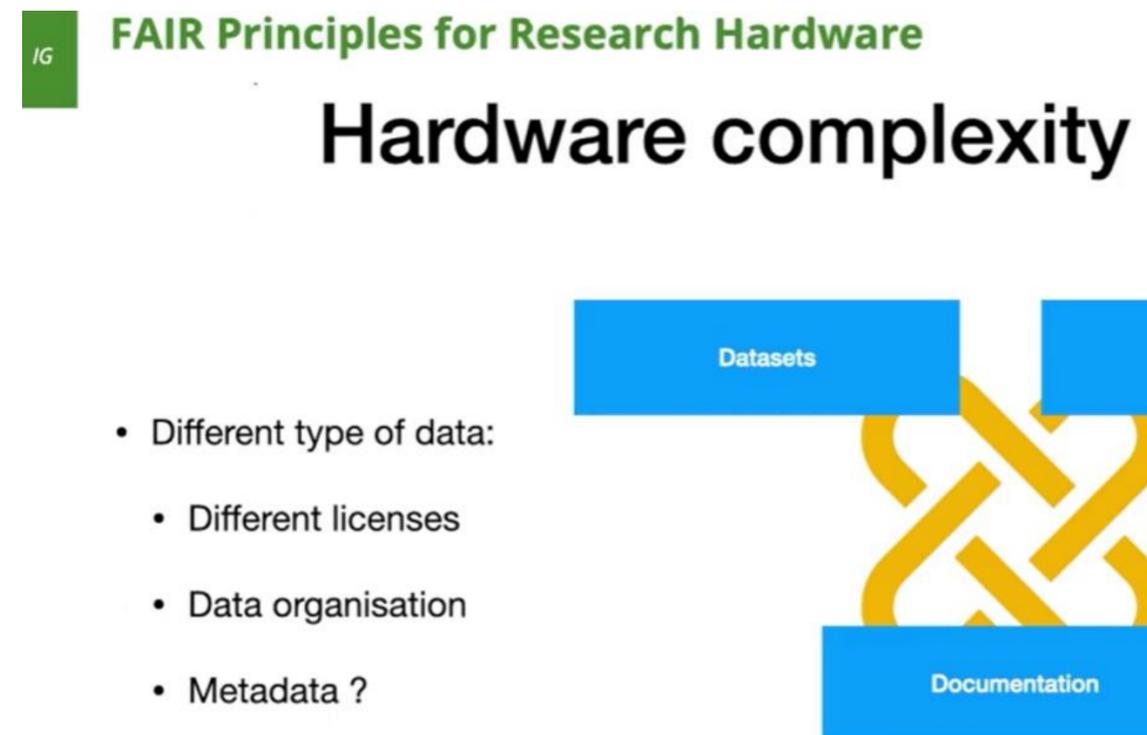
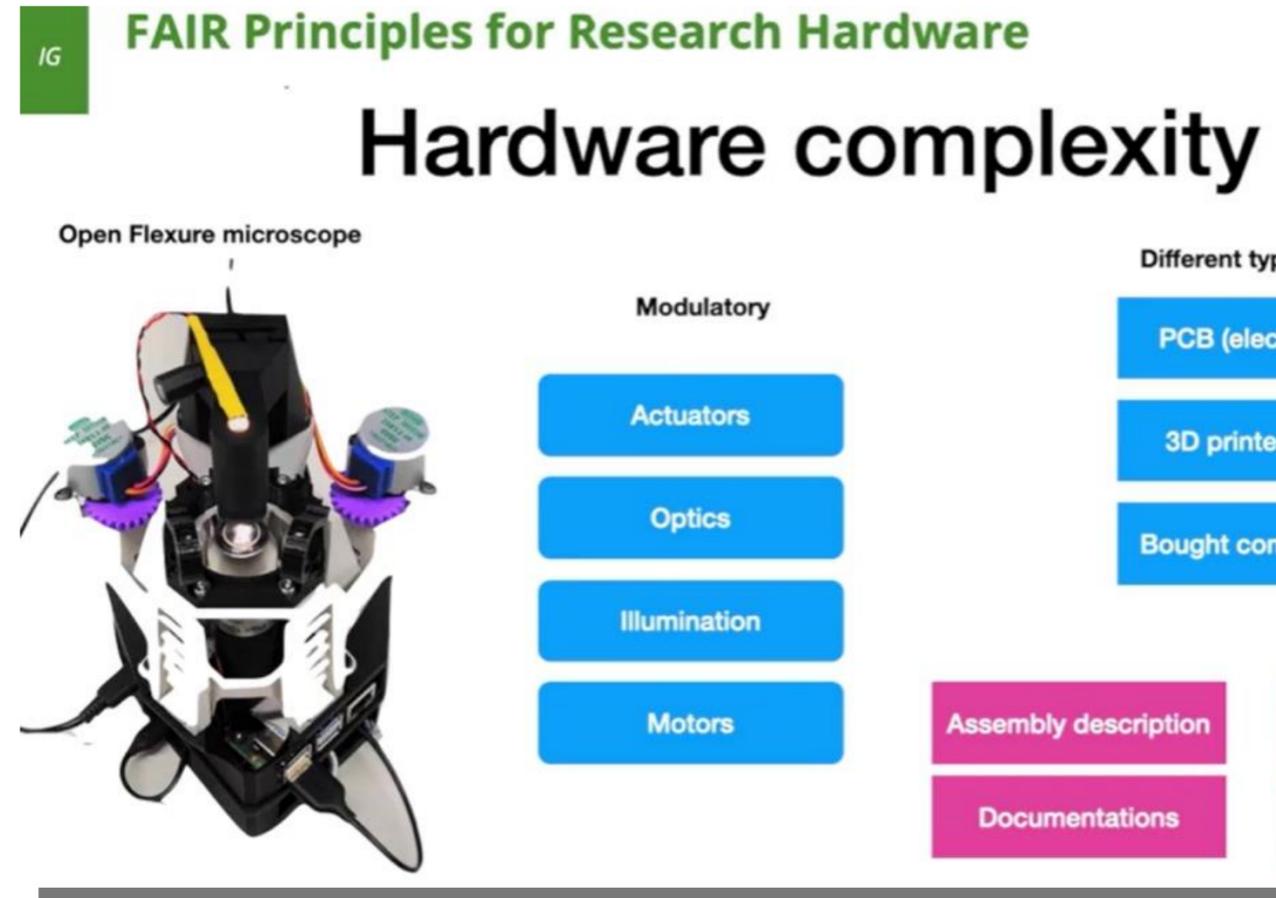
Un esempio di output non digitale della ricerca è l'hardware:

- per la ricerca
- sviluppati in ambito accademico ma destinati all'industria

Oggetti di ricerca non digitali: che cosa fare

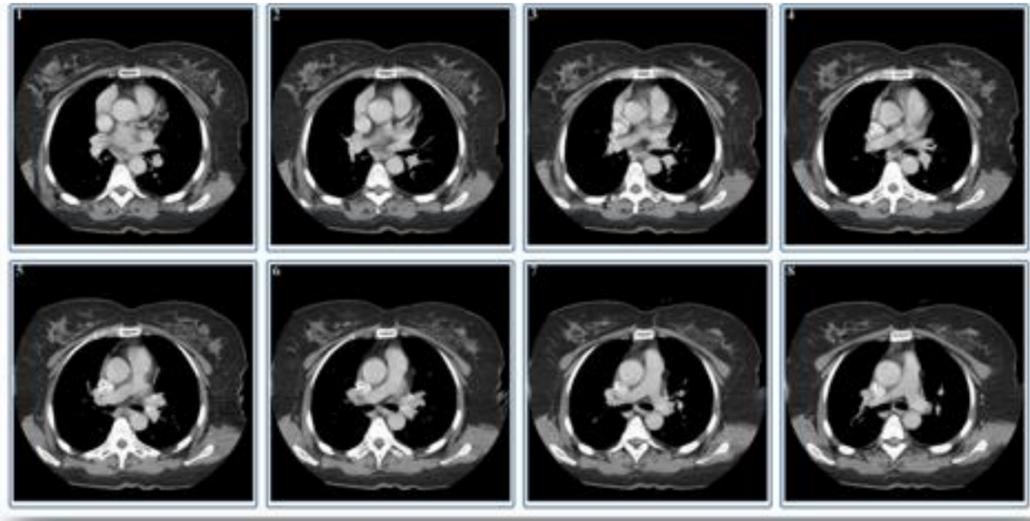
- Digitalizzare almeno i **metadati e la documentazione**
- Eventualmente **digitalizzare anche il workflow di sviluppo** del prototipo (workflow management systems, protocolli – es. Protocols.io, documenti collaborativi per appunti e documentazione)
- Nel caso si possa digitalizzare: **tenere a mente** chi è il proprietario del prototipo e ha diritto sulla digitalizzazione

06 ottobre 2023



QUALI TIPI DI DATI...

... dipende dal dominio di ricerca.



Per un ricercatore di area biomedica il dato può essere l'**immagine di una TAC**



Per un ricercatore di area sociale il dato può essere un'**intervista**

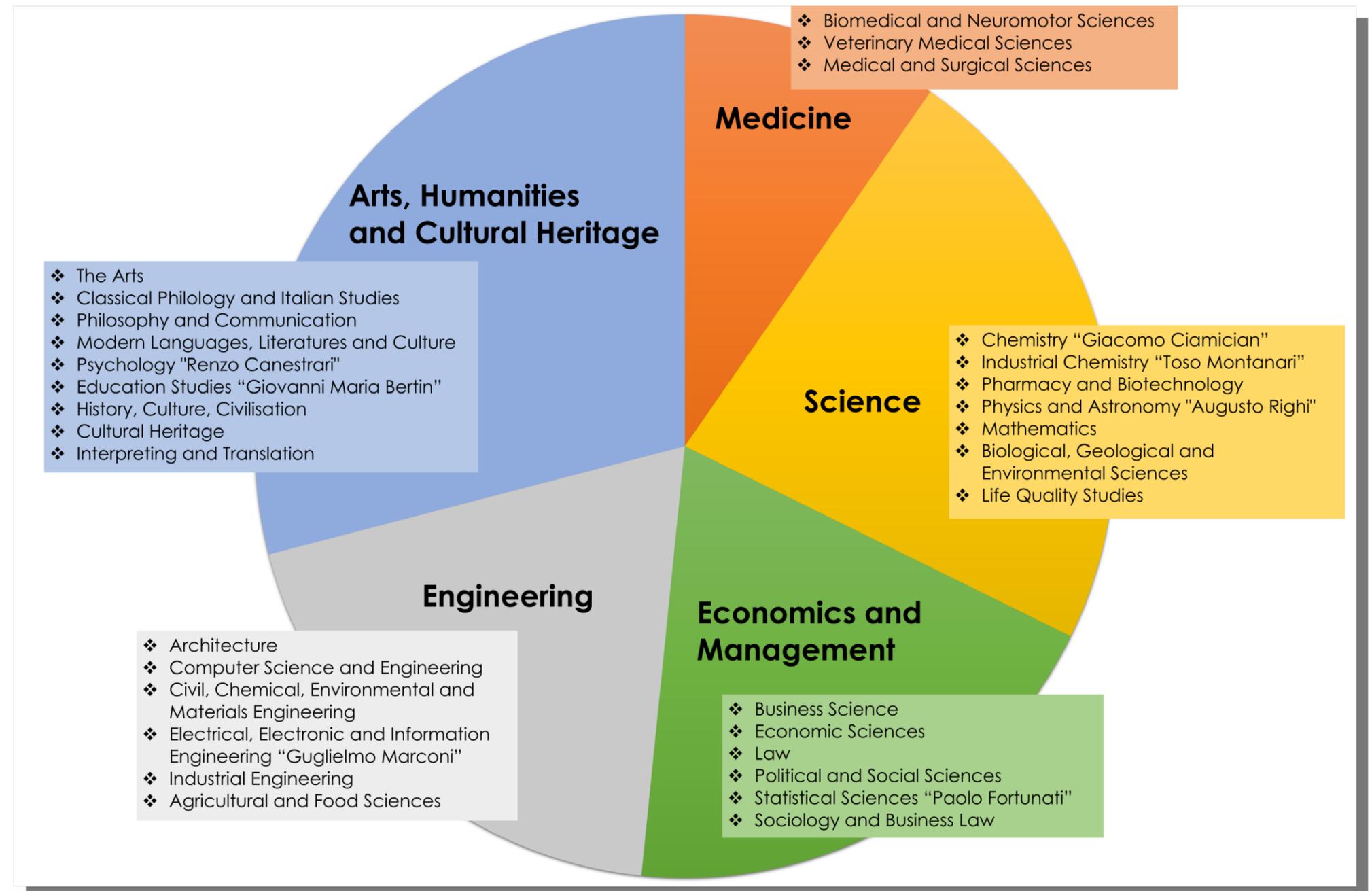


Per un ricercatore di area umanistica il dato può essere una **fonte testuale antica**

DIVERSI TIPI DI DATI...

... si incontrano in un contesto sempre più multidisciplinare.

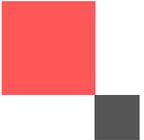
Come quello della nostra Università, che include 31 dipartimenti raggruppati in 5 macroaree.



Anche a livello di enti finanziatori, è incentivata la collaborazione tra ricercatori anche di diversi ambiti.



GESTIRE IL DATO DELLA RICERCA





La gestione dei dati di ricerca (Research Data Management, RDM) è la gestione e l'organizzazione attenta dei dati di ricerca 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.



I VANTAGGI DI UNA CORRETTA GESTIONE DEL DATO

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



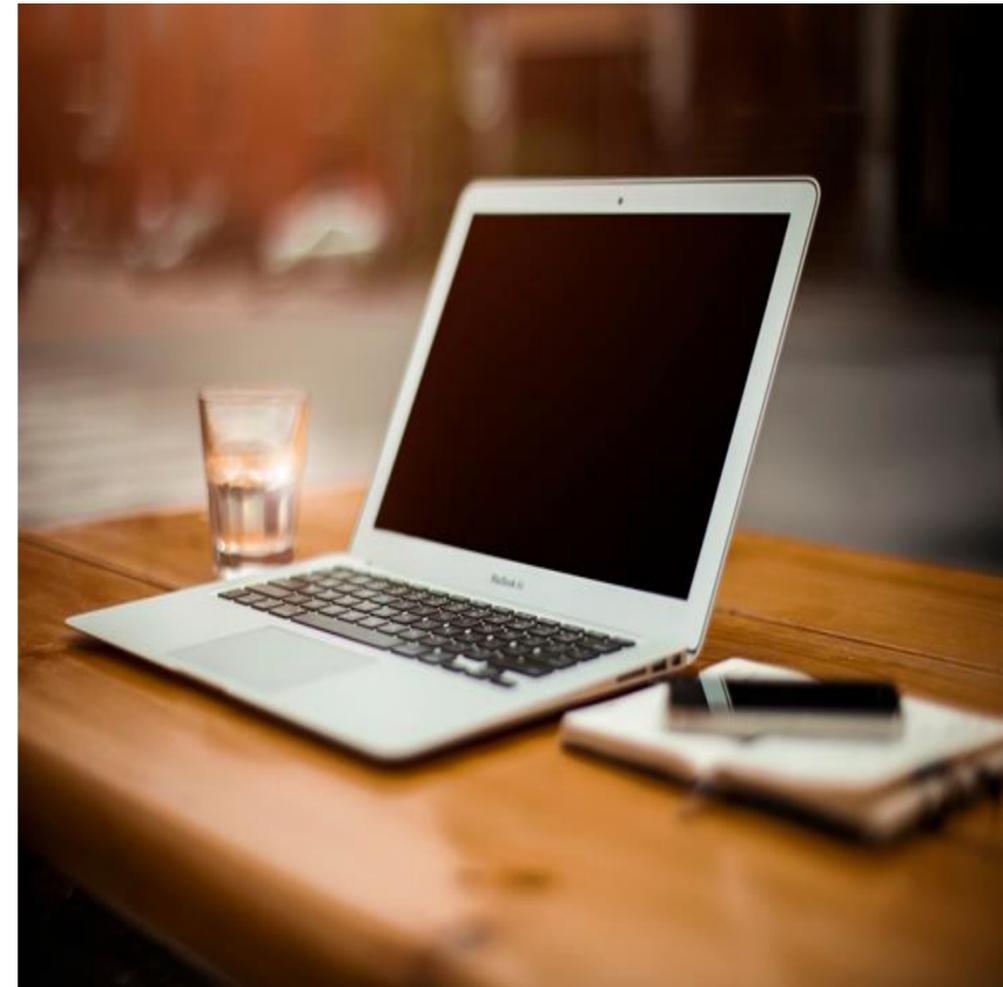
DATA MANAGEMENT HORROR STORIES

"Ho comprato un portatile nuovo di zecca solo per le mie ricerche (...).
Lo usavo per caricare i dati della ricerca (interviste e diari sul campo) e
tutto ciò che riguardava la ricerca (...).
Ho dimenticato il portatile da qualche parte (...).

I dati erano sensibili e non erano memorizzati altrove (...).

Il collega ricercatore ha comunque osato chiedermi se avessi fatto un
Data Management Plan e una valutazione dei rischi. Mi avrebbe
aiutato ad anticipare i problemi."

<https://research.tuni.fi/researchdata/news/data-horror-week-end-of-stories/>



DATA MANAGEMENT HORROR STORIES



"Abbiamo raccolto dati di ricerca ricchi e diversificati(...). Poi il progetto è terminato e ci è venuto in mente che potevamo depositare i dati per un ulteriore utilizzo (...).

Hanno iniziato a farci domande difficili come: avete chiesto ai partecipanti alla ricerca il permesso di archiviare i dati e avete menzionato l'archiviazione nella nota sulla privacy? (...) Ci era sfuggito tutto questo perché avevamo fretta di iniziare il lavoro sul campo. Di conseguenza, abbiamo perso l'opportunità di aprire i dati ad altri e persino di utilizzarli nei nostri progetti futuri."

<https://research.tuni.fi/researchdata/news/data-horror-week-end-of-stories/>

DATA MANAGEMENT HORROR STORIES



"Abbiamo diversi ricercatori in questo stesso progetto e stiamo tutti lavorando sugli stessi dati. Qualcuno sfortunato (non io) ha toccato il file dei dati grezzi e vi ha apportato alcune modifiche. Non ci sono più i dati originali. Allo stesso modo, non è chiaro chi abbia fatto cosa e quando con i dati.

Come potremo ora ottenere una descrizione valida dell'elaborazione dei dati in una rivista scientifica e vogliamo che questo studio sia trasparente? Gente, tenete i dati grezzi separati e le regole del gioco chiare."

<https://research.tuni.fi/researchdata/news/data-horror-week-end-of-stories/>

DATA MANAGEMENT HORROR STORIES

"Ho sviluppato quella che pensavo fosse una struttura logica di cartelle, c'erano sottocartelle di sottocartelle e le loro sottocartelle, e le cartelle avevano nomi con abbreviazioni fantasiose (...). Poi ho dovuto prendermi una pausa dalla ricerca (...).

È stato un po' angosciante tornare lì e riprendere in mano i miei dati. Beh, cartella dopo cartella e file dopo file tutto ha cominciato a funzionare."

<https://research.tuni.fi/researchdata/news/data-horror-week-end-of-stories/>



QUALI VANTAGGI PUOI AVERE CON UNA CORRETTA GESTIONE DEL DATO DI RICERCA?



Organizzare i dati rende il tuo lavoro più efficiente.

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



Se li gestisci, potresti non perderli.

L'archiviazione corretta dei dati e il backup regolare prevengono le perdite di dati.



Alcuni dati potrebbero essere unici e non riproducibili.

Questo li rende preziosissimi per la comunità scientifica.



Aumenta l'integrità della ricerca.

Un dato correttamente gestito facilita la validazione e il controllo.



Stimola la collaborazione con altri ricercatori.

Troveranno più facile comprendere e riutilizzare i vostri dati.



Per i ricercatori

- Maggiore visibilità e citazioni
- Opportunità di collaborazione
- Riconoscimento per la carriera
- Aiuta a prevenire errori e migliora la qualità delle analisi dei dati
- Diminuisce i rischi di "non-compliance" (etica, privacy, policy degli enti finanziatori)



Per la ricerca

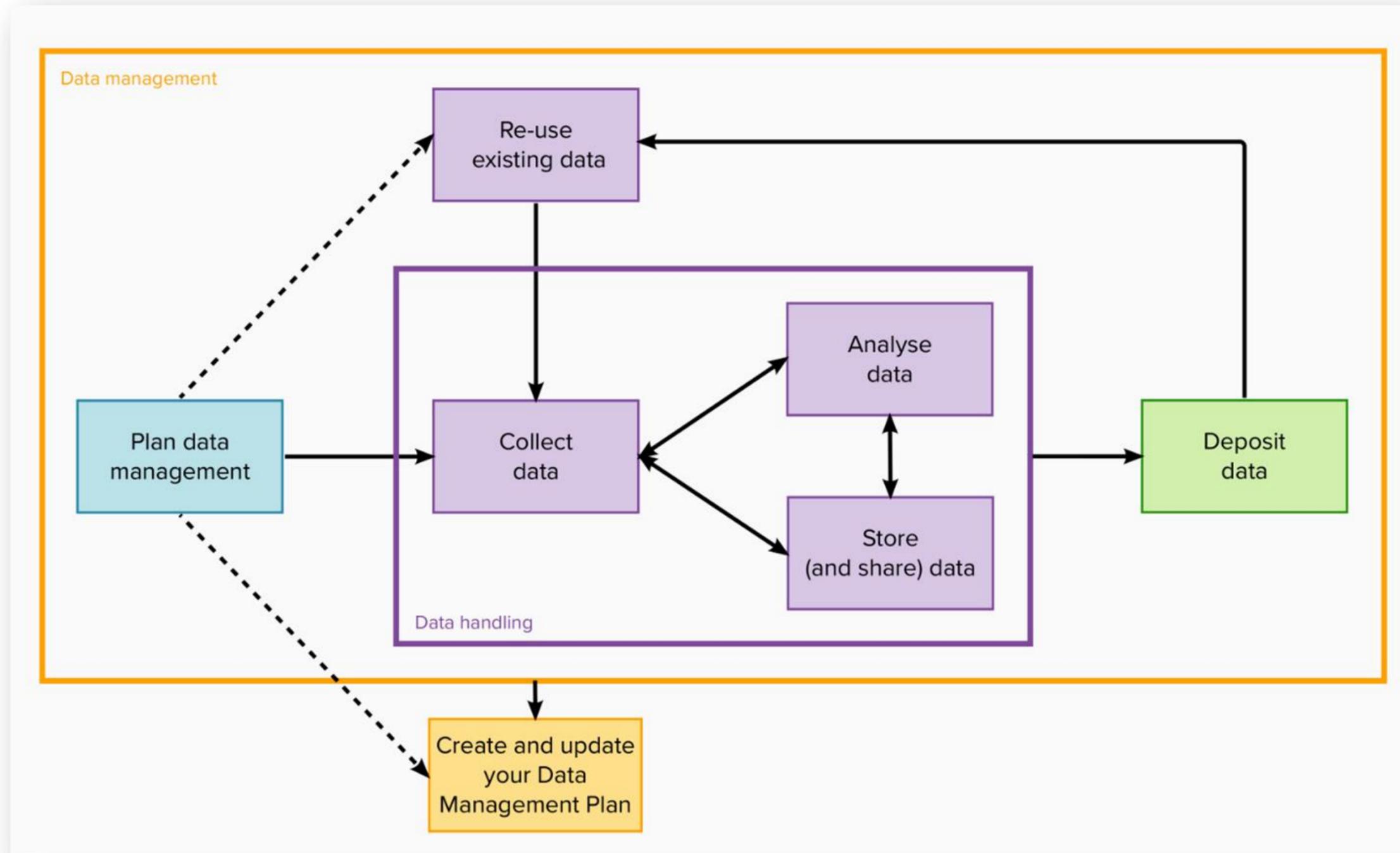
- Facilita la rintracciabilità e il riuso dei dati
- Consente nuove ricerche e nuovi approfondimenti sui dati
- Protezione dei dati di valore
- Sostiene l'integrità e la riproducibilità della ricerca



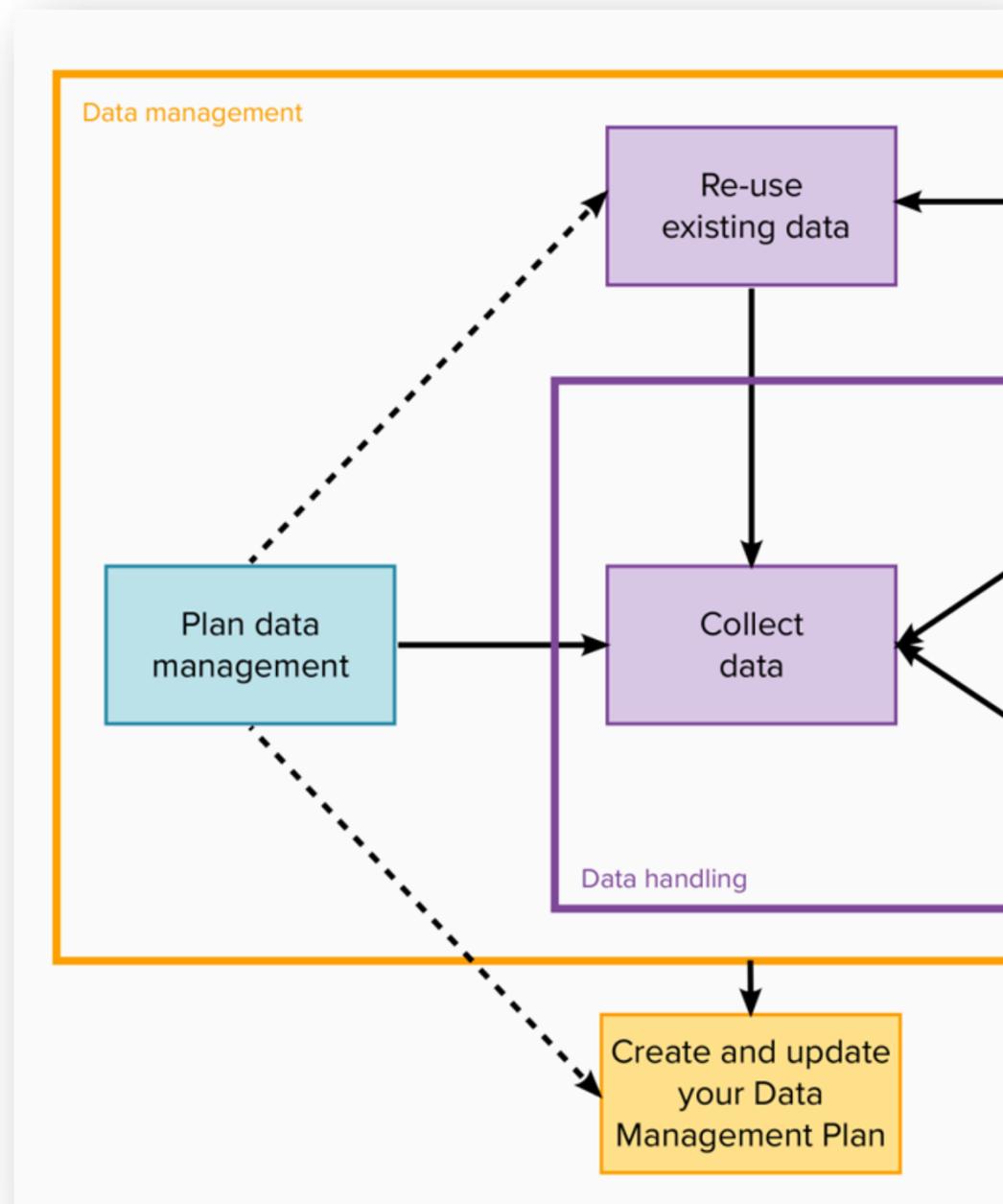
Per la società

- Uso efficiente delle risorse pubbliche
- Una migliore qualità della ricerca può dare beneficio al processo decisionale
- Opportunità di business
- Opportunità di citizen science
- Maggiore trasparenza e fiducia nella scienza

LE FASI DELLA GESTIONE DEL DATO DI RICERCA: IL CICLO DI VITA DEL DATO



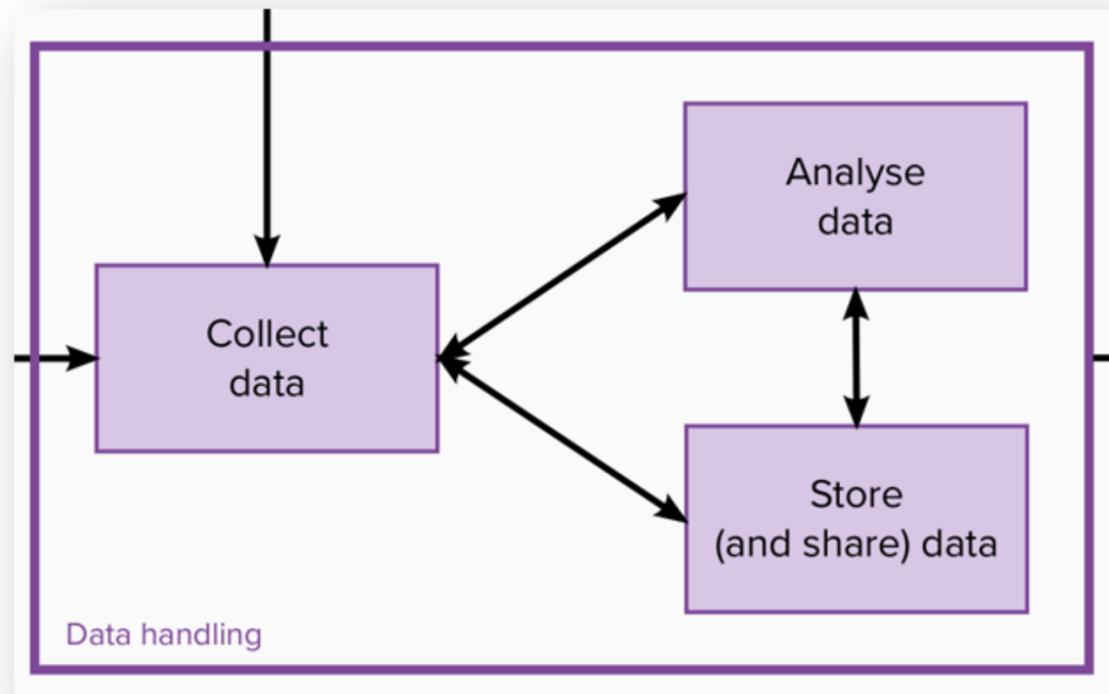
LE FASI DELLA GESTIONE DEL DATO DI RICERCA: IL CICLO DI VITA DEL DATO



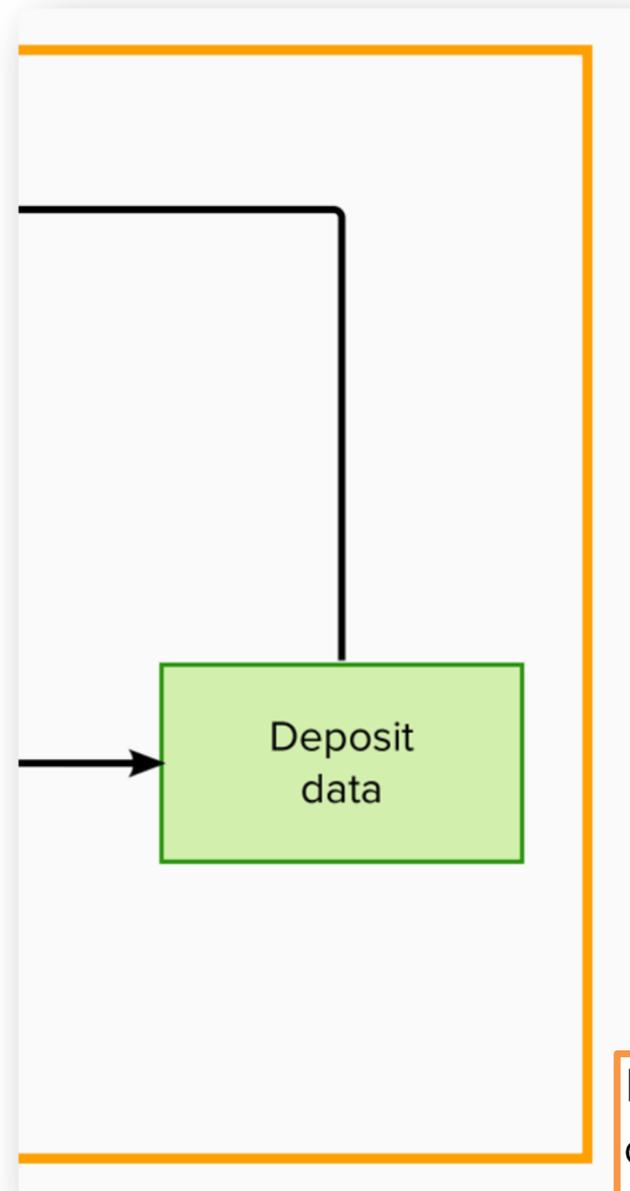
- La fase di pianificazione è cruciale per operare delle scelte ragionate durante tutto lo svolgimento del progetto.
- I dati possono essere sia generati *ex-novo* durante lo svolgimento del progetto che riutilizzati da fonti già esistenti.
- La stesura di un Data Management Plan è una buona pratica per tenere traccia dall'inizio del Progetto delle scelte fatte in termini di gestione del dato.

LE FASI DELLA GESTIONE DEL DATO DI RICERCA: IL CICLO DI VITA DEL DATO

- Le fasi attive della ricerca sono le più concitate, in cui è più facile perdere di vista la corretta gestione del dato per arrivare velocemente ad un risultato.
- Portare avanti processi di controllo qualità dei dati e tenere traccia delle diverse versioni è fondamentale per evitare errori e lavorare sempre con dati consistenti.
- In questa fase si possono scegliere i formati più adatti ai propri dati.
- Durante le fasi attive della ricerca i dati devono essere conservati in spazi di archiviazione appropriati, provvedendo ai backup necessari per non rischiare di perdere i propri dati.
- Organizzare file e cartelle in modo gerarchico e assegnando nomi coerenti aiuta a non perdersi nella navigazione!



LE FASI DELLA GESTIONE DEL DATO DI RICERCA: IL CICLO DI VITA DEL DATO



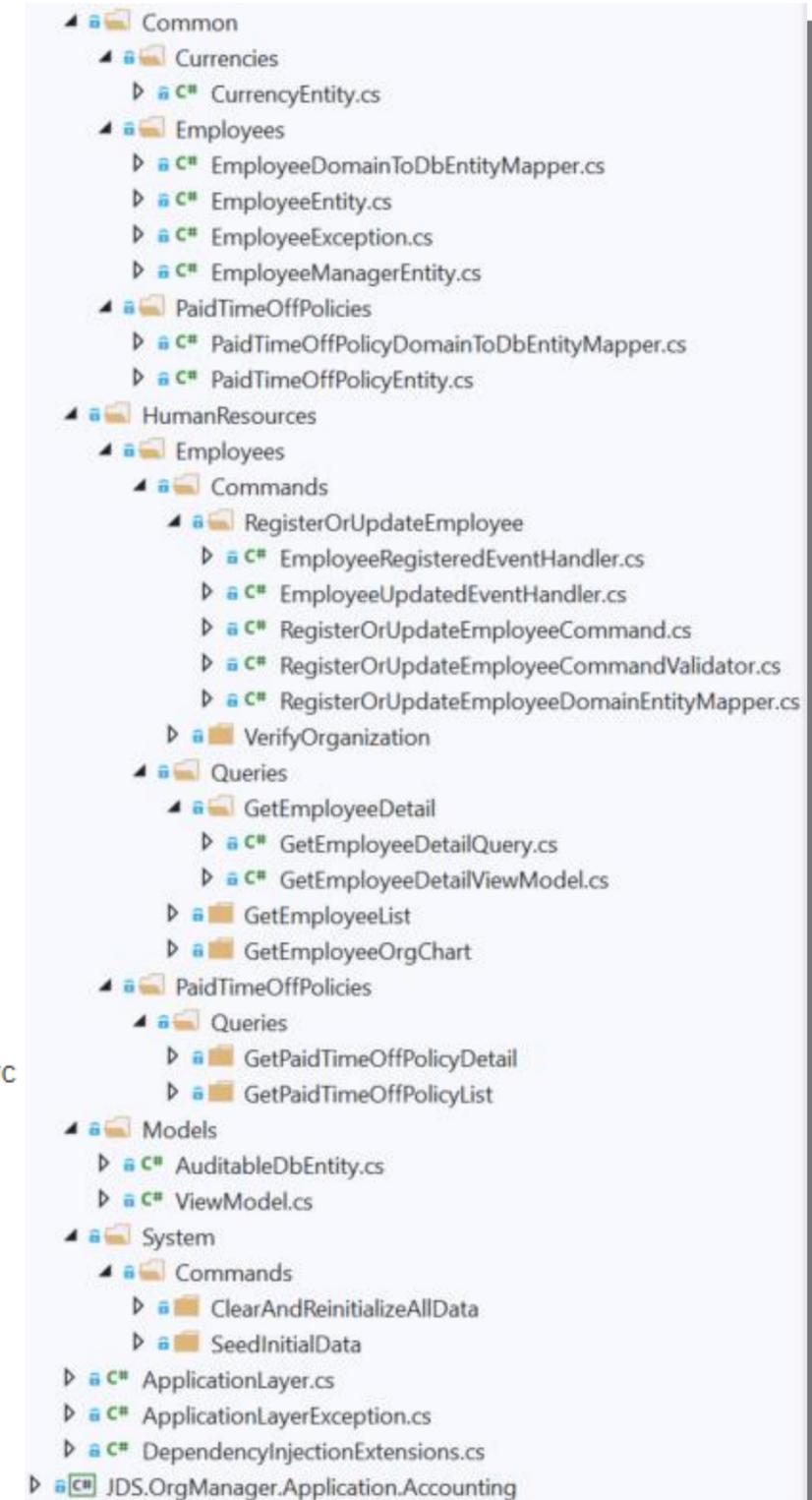
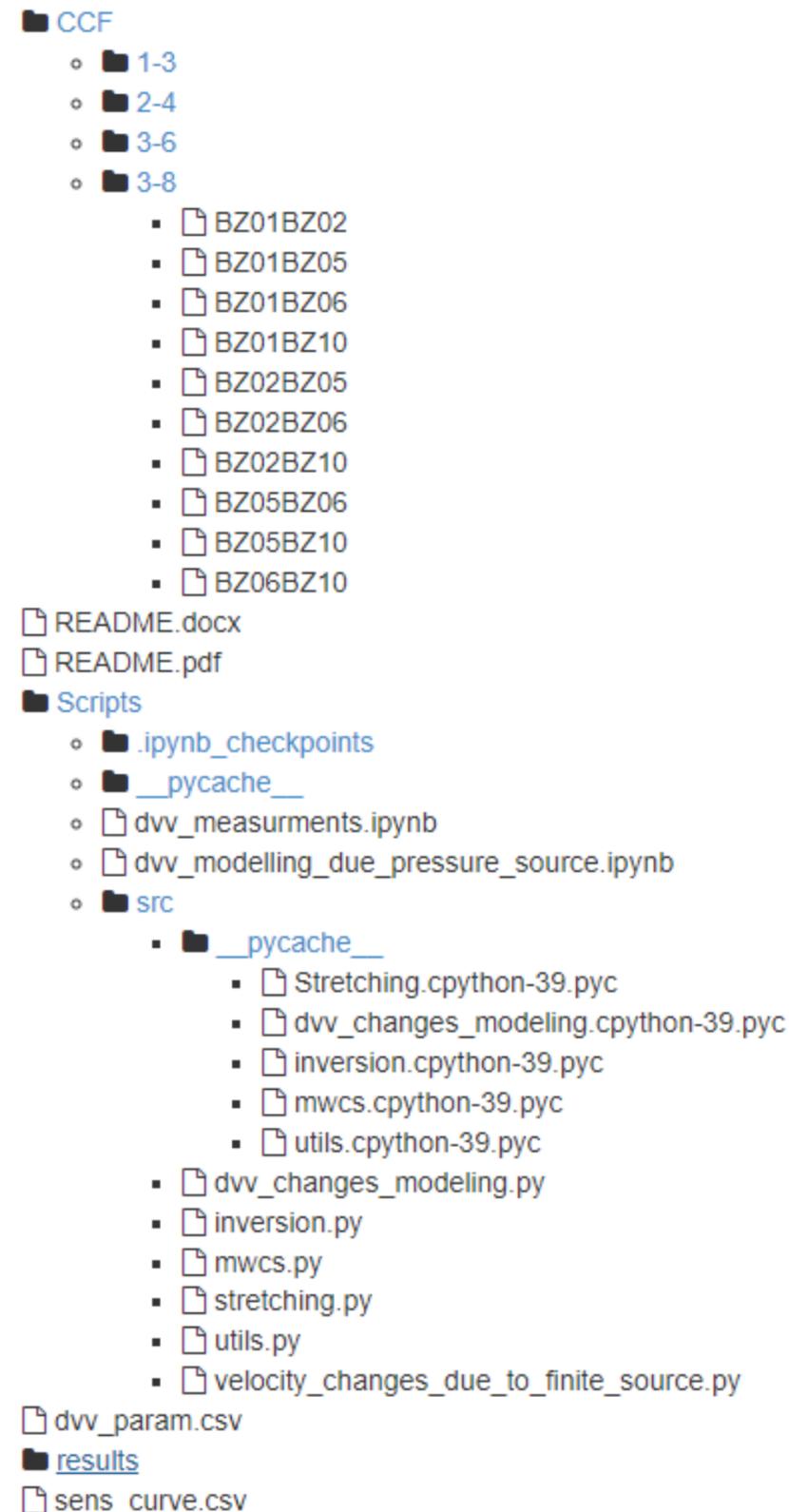
- Il deposito dei dati, organizzati in datasets, è fondamentale per la preservazione a lungo termine e per la condivisione dei dati della ricerca.
- Depositare i dati nei repositories adeguati ne garantisce la visibilità da parte di altri ricercatori, favorisce la riproducibilità e aumenta il valore della ricerca.

NB. Con “deposito dei dati” non si intende il loro caricamento su siti web personali, come siti web editori, o Servizi di cloud storage (Dropbox, Google drive, ecc.) o piattaforme di ricerca (Academia.edu, ResearchGate)!

I DATASET

DATI: informazioni raccolte per essere esaminate e utilizzate come base per ragionamenti, discussioni o calcoli e, più in generale, utilizzate per convalidare un'ipotesi.

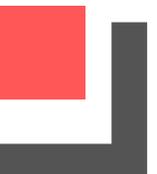
DATASET: raccolta di dati assemblati con uno **scopo comune e organizzati** per riflettere i risultati di un'attività di ricerca.





QUALCHE PUNTO CHIAVE

- Esistono diverse definizioni di dato e variano in base alla disciplina di riferimento.
- Dato è tutto ciò che è alla base dei ragionamenti a supporto di una tesi di ricerca.
- Una collezione di dati accomunati dallo stesso obiettivo è chiamata dataset.
- Il dato ha un valore intrinseco come asset della ricerca, per conservarlo deve essere gestito correttamente.
- La gestione corretta è un processo che attraversa tutte le fasi del ciclo di vita del dato della ricerca.





RENDERE DISPONIBILI I DATI AUMENTA IL VALORE DELLA RICERCA

Science Robotics

Current Issue Archive About [Submit manuscript](#)

What's the policy on publication of data sets or source code?

As a condition of publication, authors must agree to make available all data necessary to understand and assess the conclusions of the manuscript to any reader of *Science Robotics*. Data must be included in the body of the paper or in the supplementary materials, where they can be viewed free of charge by all visitors to the site. When appropriate, we encourage authors to make source code available via one of the open source project hosting web sites (e.g., SourceForge and GitHub) to facilitate recreation of experimental results and wider adoption of the technology.

materialstoday

Connecting the materials community

[Home](#) [NEWS](#) [JOURNAL ARTICLES](#) [WEBINARS](#) [COMMENT](#) [FEATURES](#) [PODCASTS](#) [PRODUCTS](#) [EVENTS](#)

Data availability statement

Authors are required to include a 'Data availability' section in their manuscript. The intention of this section is to provide readers with information on where they can obtain the research data required to reproduce the work reported in the manuscript. Research data refers to the results of observations or experimentation that validate research findings. In this context, data includes "raw" data (eg, as recorded from apparatus) and processed data (eg, as derived from raw data, and/or output from code).



ADVANCED SCIENCE

Open Access

[HOME](#) [ABOUT](#) [CONTRIBUTE](#) [BROWSE](#) [SPECIAL FEATURES](#)

Publication Ethics Guidelines

[top ↑](#)

1. Responsibilities of Authors

Authors are person(s) who generated the content of a manuscript, and whose names are given as (co-)authors on that manuscript.

- The precise criteria for authorship may vary by field, but authors should at minimum be individuals who have made a substantial intellectual contribution to the work, which should include contributions to the conception, design, analysis, and/or manuscript preparation.
- All authors are accountable for the content of the manuscript.
- An administrative role (e.g., providing laboratory space or funding) does not in itself confer authorship, and guest/gift authorship is not appropriate.
- Contributions that fall short of the criteria for authorship may be listed in the acknowledgement section. Authors should verify that individuals agree to be acknowledged.

Authors have the following responsibilities:

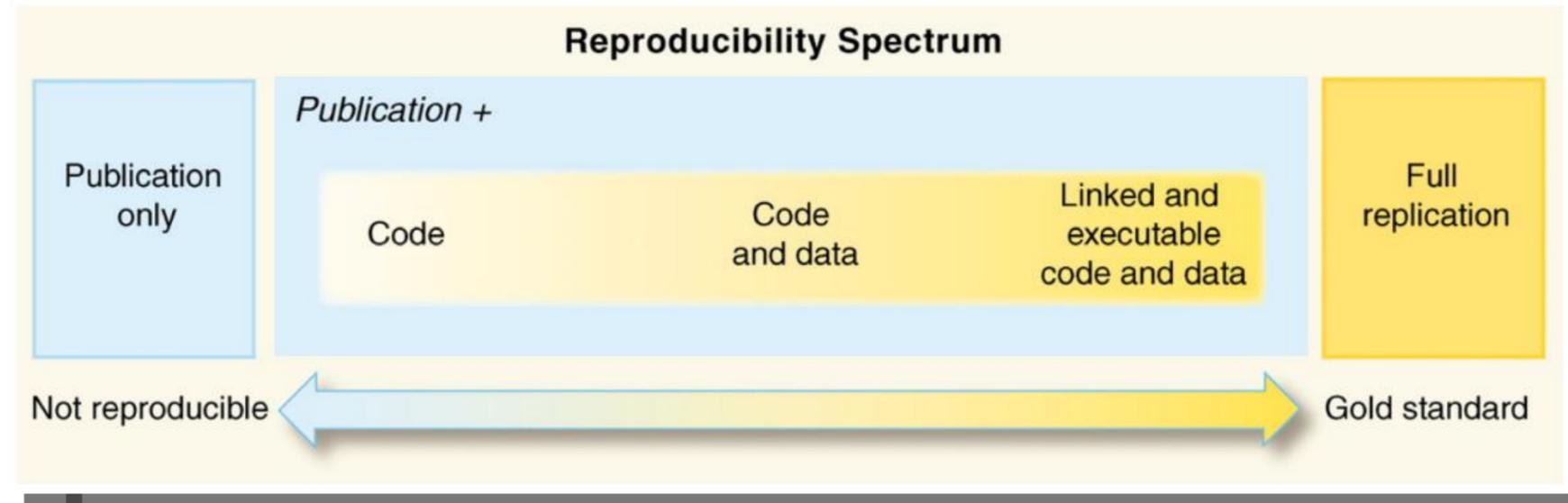
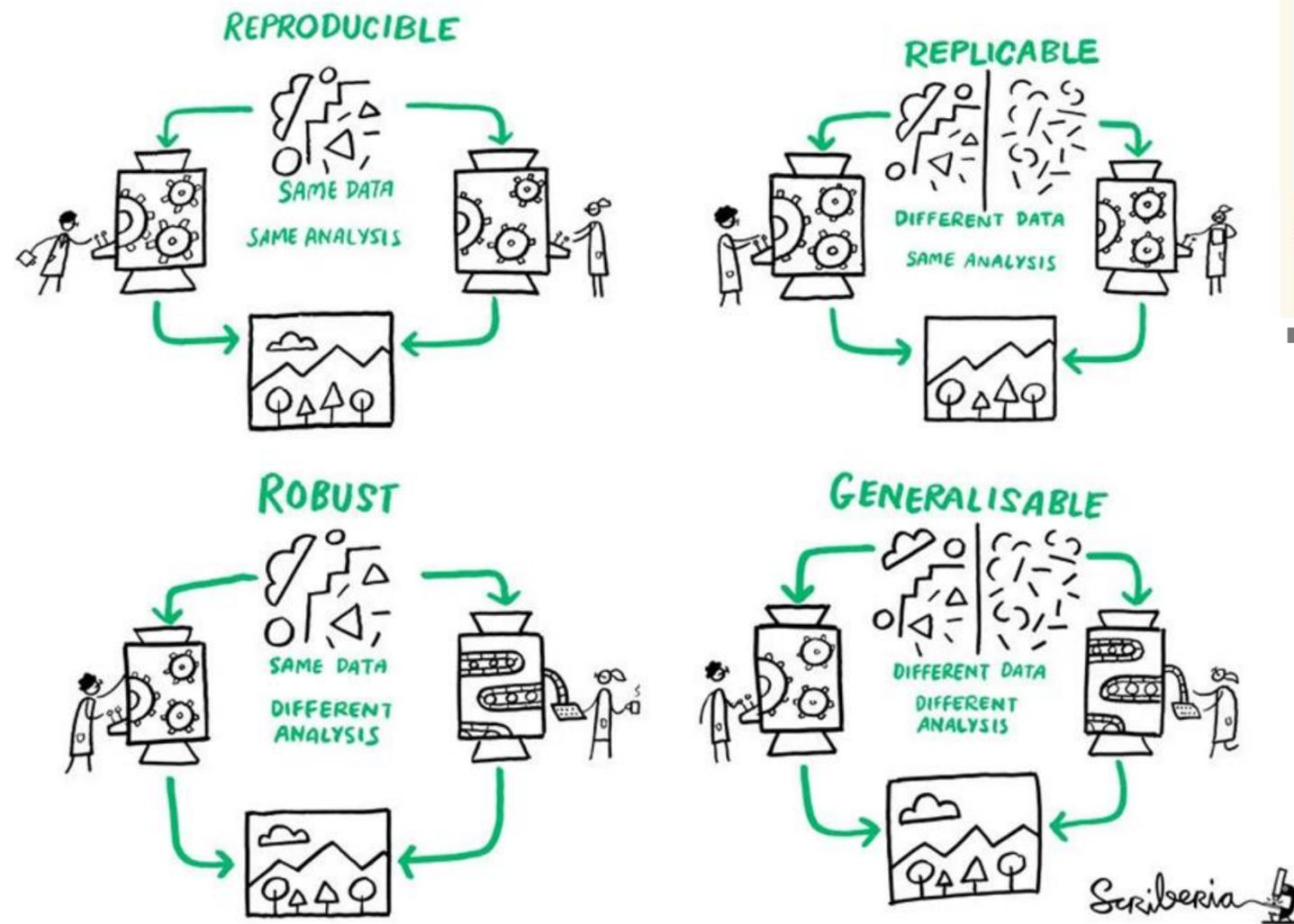
1.1. To gather and interpret data in an honest way. Data and image manipulation, as well as any other misrepresentation of the results, is unacceptable. Original unprocessed images should be utilized only. Where adjustments are necessary for scientific reasons they must be disclosed in a figure legend.

1.2. To present a concise and accurate report of their research and an objective discussion of its significance.

1.3. To provide sufficient data and methodological details that the work can be critically evaluated, reproduced, and extended by other researchers. Authors are encouraged to provide access to unique materials and applied software, with an appropriate material transfer agreement (MTA) to protect their own interests, to other researchers on request.

1.4. To make primary datasets available according to community data-sharing best practice, for example, through deposition in publicly available databases or community repositories.

LA RICERCA RIPRODUCIBILE: AFFIDABILE E FRUIBILE



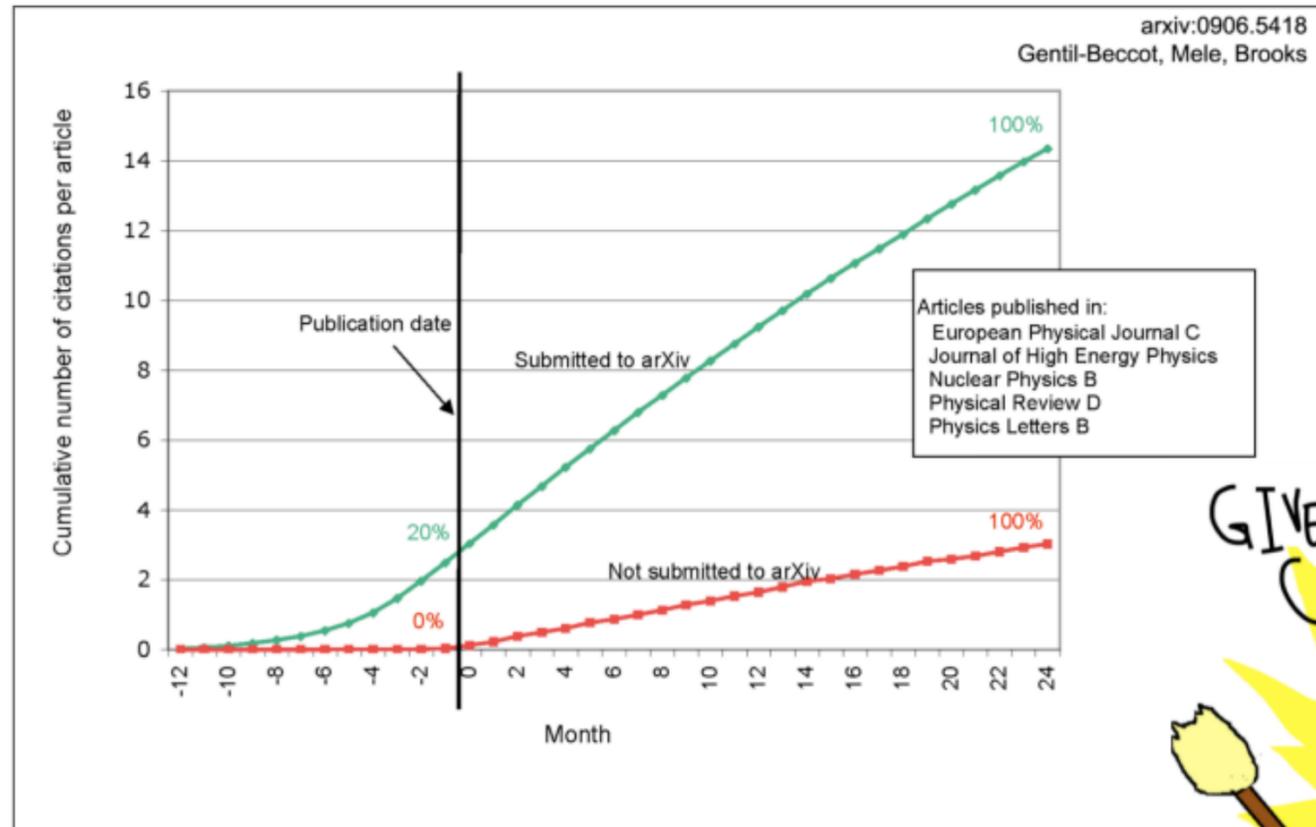
Ainsworth, Rachael. (2018, October 17). Reproducibility and Open Science. Data Science for Experimental Design (DSED), The Alan Turing Institute, London, UK. Zenodo. <https://doi.org/10.5281/zenodo.1464853>

Quando la ricerca è riproducibile è considerata più affidabile ed è effettivamente più utile per l'avanzamento scientifico

Ausili alla riproducibilità:

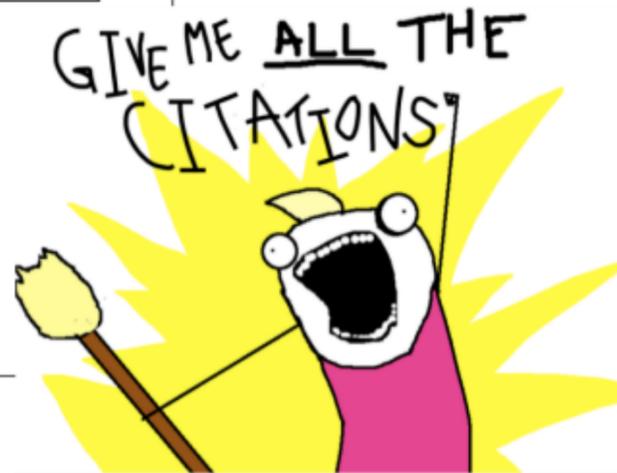
- Docker containers
- Piattaforme cloud (notebooks)
- Deposito in repository specifici, come Software Heritage

I VANTAGGI PER I RICERCATORI



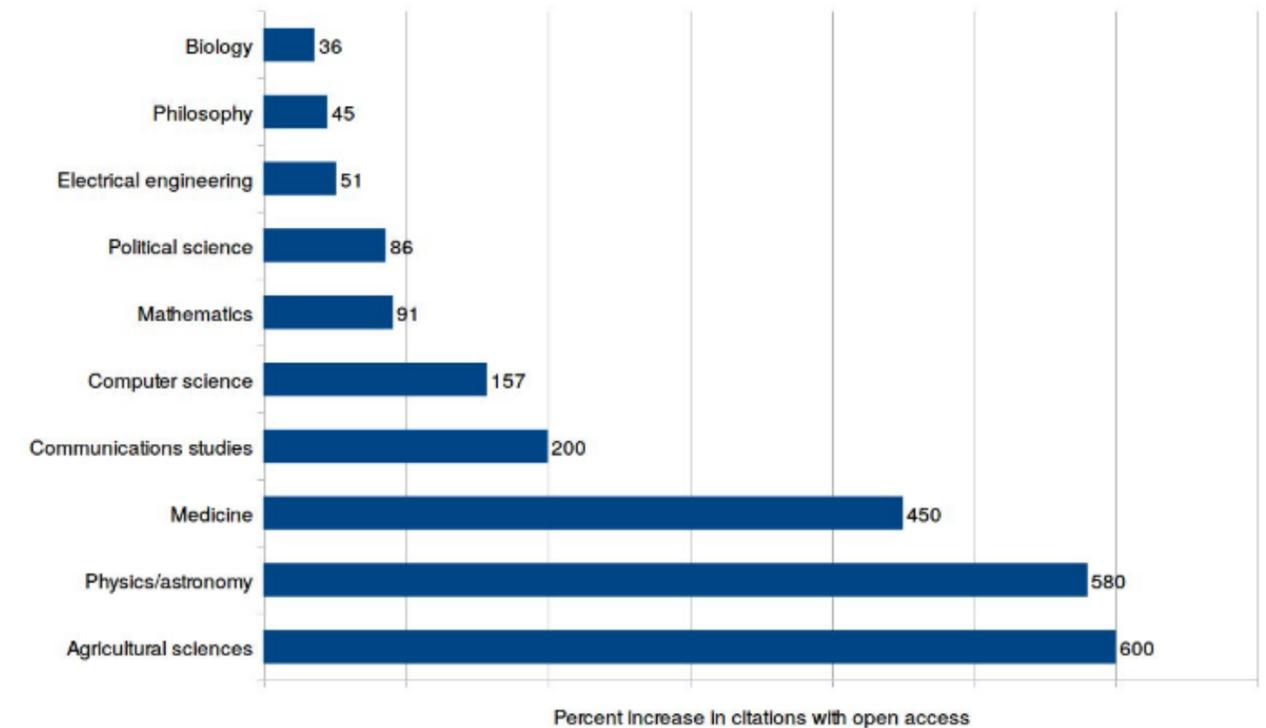
Gentil-Beccot, Mele, Brooks (2009), <https://arxiv.org/abs/0906.5418>

R Ainsworth - @rachaelevelyn



Why should you share preprints?

- Enable **rapid** publication
- Enable **open access**
- **Citation advantage**



Source: Data from [Alma Swan, 2010](#). Figure produced by E.C. McKiernan (CC BY).

Maggior visibilità e potenziale di collaborazione, networking ed eventualmente finanziamenti

✉ kevin.moerman@nuigalway.ie

🐦 @KMMoerman

I VANTAGGI PER I RICERCATORI: L'OPEN SCIENCE IN PRATICA

"Vague but exciting"

CERN

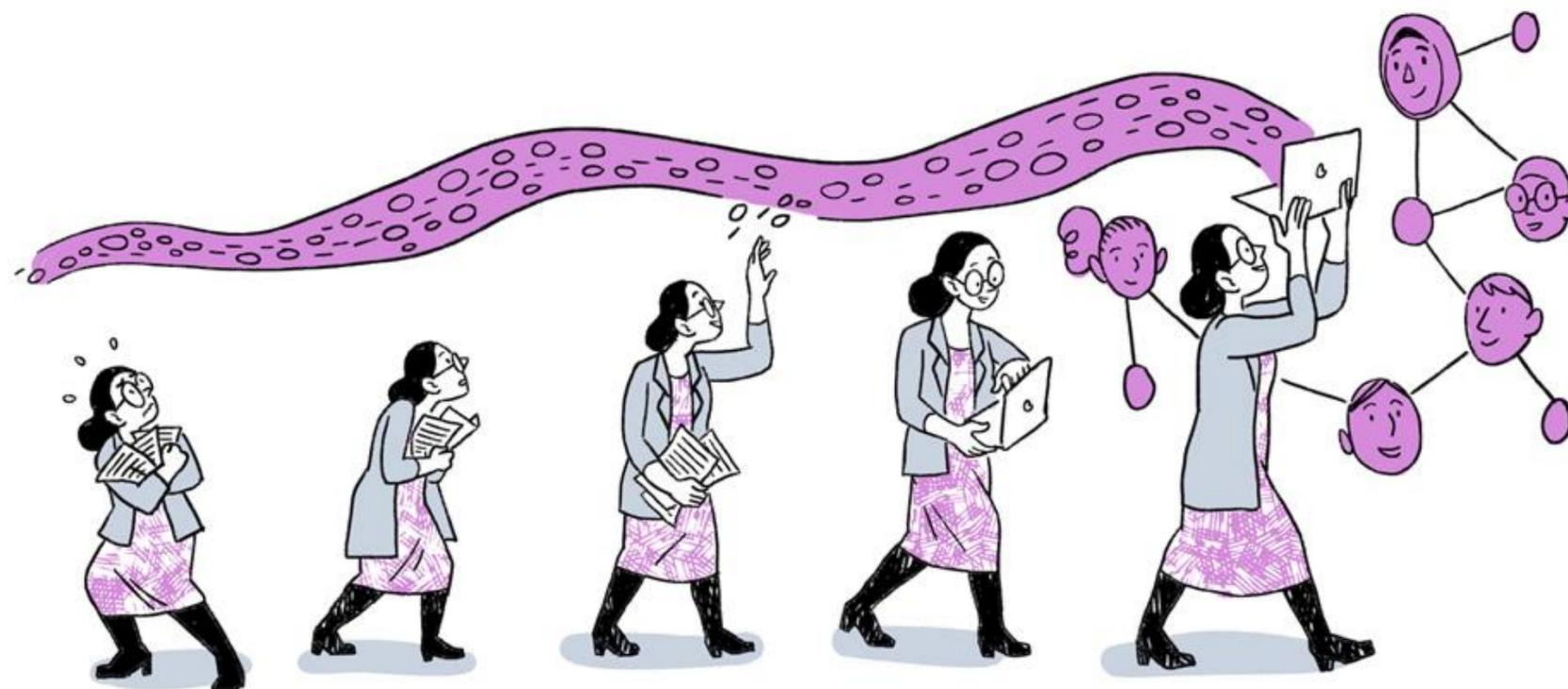
WWW.cern.ch

...THE HTTP PROTOCOL, WHICH CHANGED OUR LIVES – IT USED TO BE AN INTERNAL TOOL, CERN DECIDED TO OPEN IT UP

The image is a composite. At the top, a chalkboard has the handwritten phrase "Vague but exciting" in white chalk. Below this, on the left, is a scan of a document titled "Information Management: A Proposal" by Tim Berners-Lee, dated March 1989. The document includes an abstract and a complex diagram showing relationships between various systems like "Computer conferencing", "Hypermath", and "CERNDOC". In the center, there is a portrait of Tim Berners-Lee. To the right of the portrait, a black box contains the text: "...THE HTTP PROTOCOL, WHICH CHANGED OUR LIVES – IT USED TO BE AN INTERNAL TOOL, CERN DECIDED TO OPEN IT UP". The background of the entire image is dark, with the words "CERN" and "WWW.cern.ch" written in white on a surface that looks like a chalkboard.

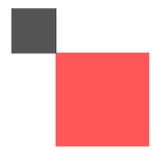
Esempio «primitivo» di Open Science che mostra i benefici della condivisione dei risultati di una ricerca:

- benefici per altri ricercatori e l'avanzamento scientifico
- benefici per la comunità



EVOLVING TOWARDS AN
ERA OF
OPEN RESEARCH

Scriberia 



COS'È L'OPEN SCIENCE

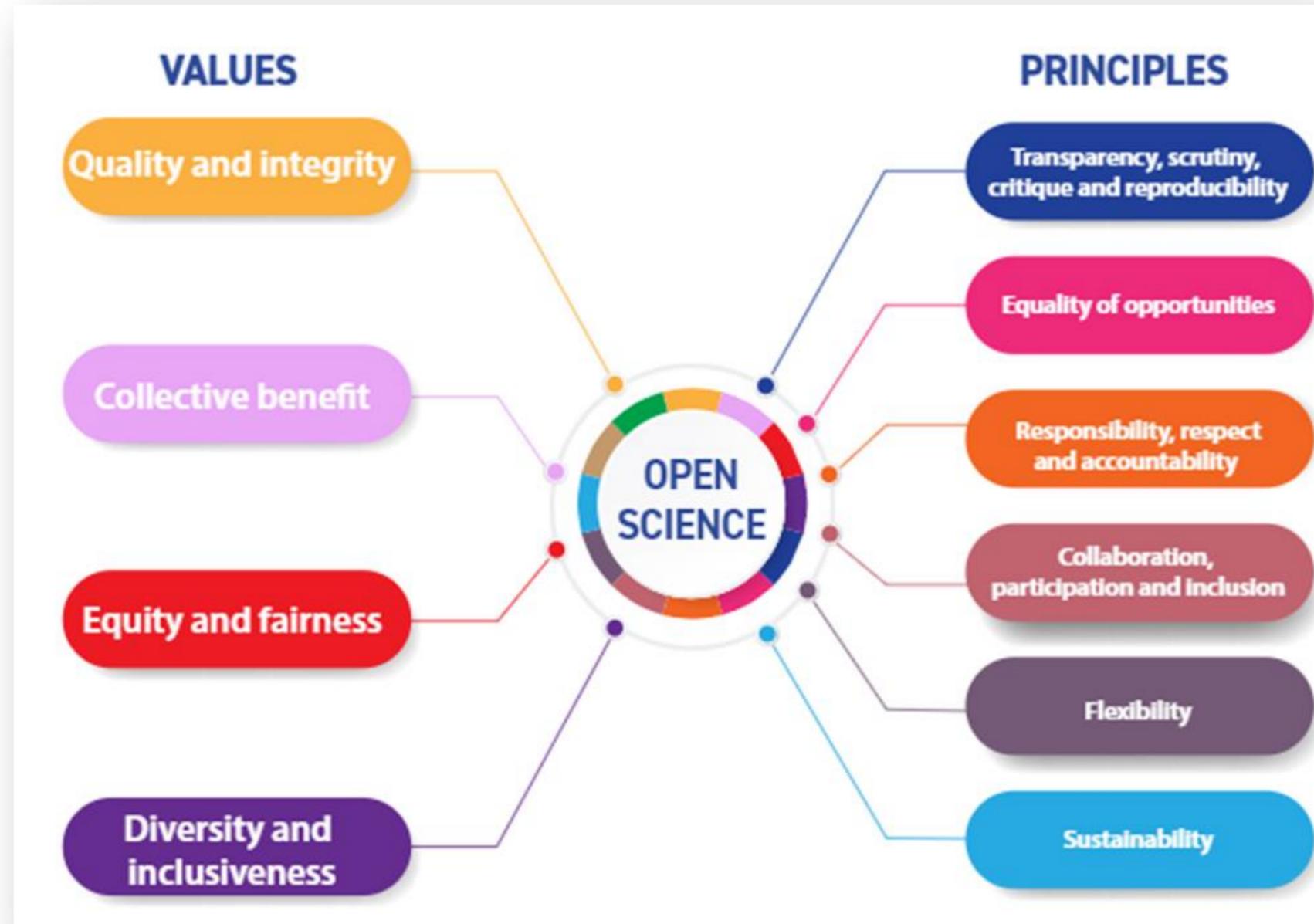




Open Science è definita come un costrutto inclusivo che combina vari movimenti e pratiche che mirano a rendere la conoscenza scientifica multilingue apertamente disponibile, accessibile e riutilizzabile per tutti, ad aumentare le collaborazioni scientifiche e la condivisione delle informazioni a beneficio della scienza e della società e ad aprire i processi di creazione, valutazione e comunicazione della conoscenza scientifica agli attori della società al di là della comunità scientifica tradizionale.

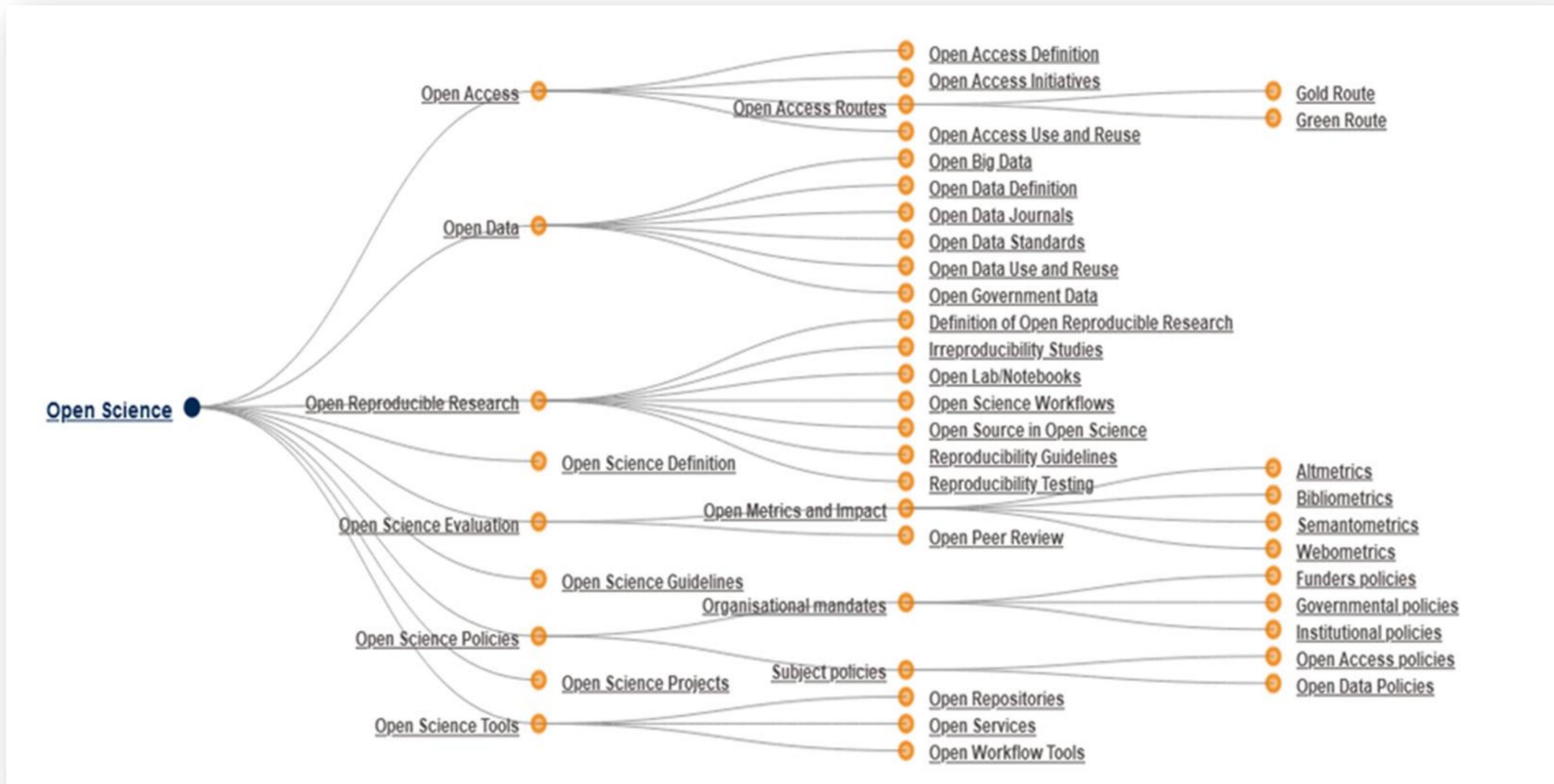


È uno standard di lavoro che si fonda sulla cooperazione e la diffusione della conoscenza.



È finalizzato a migliorare la qualità, l'efficienza e la capacità della ricerca di dare risposte concrete alle sfide poste.

Comprende l'accesso aperto a pubblicazioni e dati, ma non solo!





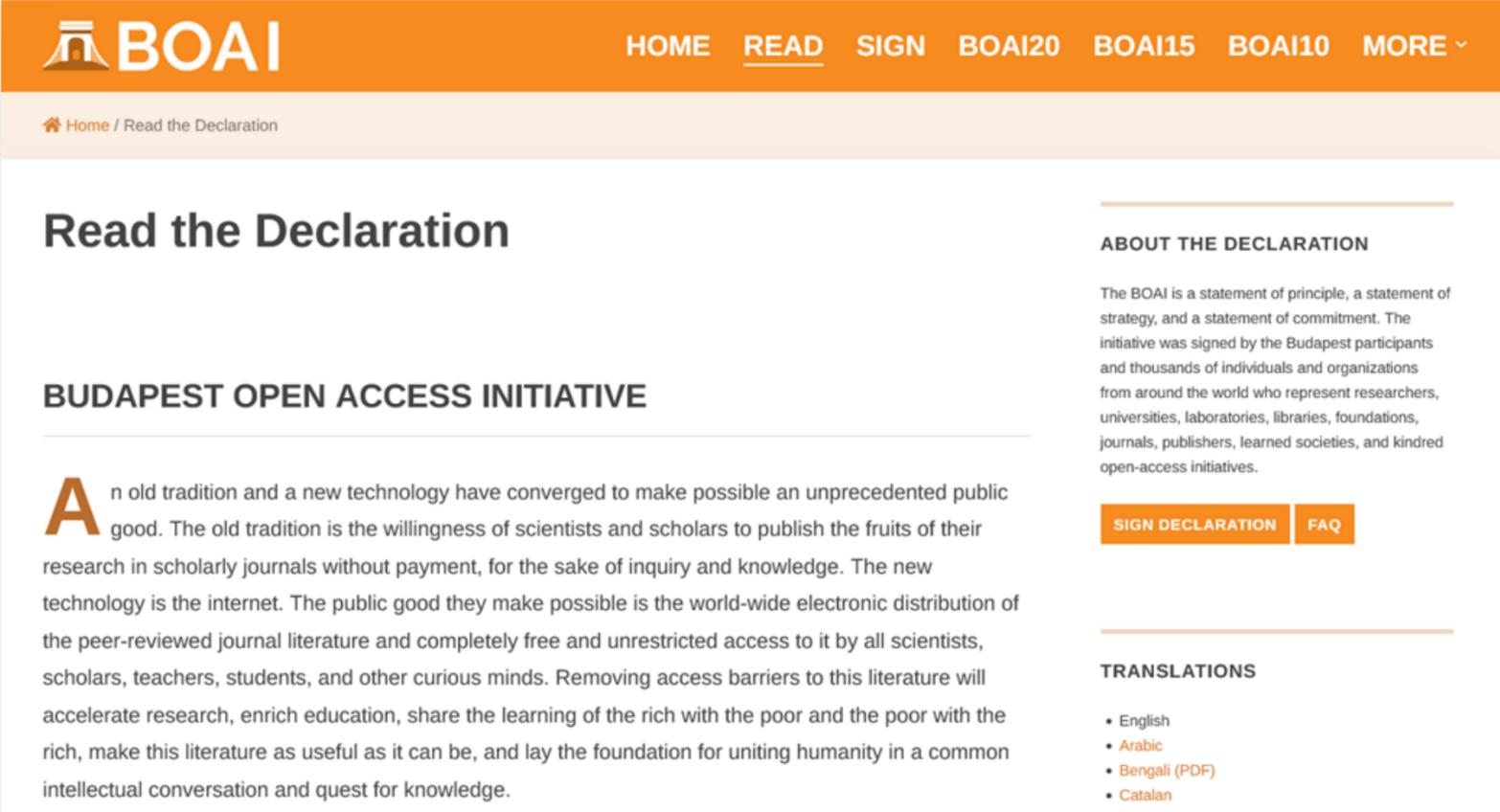
OPEN SCIENCE: COME CI SIAMO ARRIVATI?



OPEN SCIENCE: come ci siamo arrivati?

Tutto è iniziato con l'accesso aperto, definito per la prima volta nel 2002 dalla Budapest Open Access Initiative (2002):

La letteratura che dovrebbe essere **liberamente accessibile online** è quella che gli studiosi mettono **a disposizione del mondo senza aspettarsi alcun pagamento** [...]. Per "accesso aperto" a questa letteratura intendiamo la sua libera disponibilità sulla rete Internet pubblica, consentendo a qualsiasi utente di **leggere, scaricare, copiare, distribuire, stampare, ricercare o collegarsi ai testi integrali di questi articoli**, (...), di passarli come dati a un software o di utilizzarli per qualsiasi altro scopo lecito, **senza barriere finanziarie, legali o tecniche** diverse da quelle inseparabili dall'accesso a Internet stesso.



The screenshot shows the BOAI website with an orange header. The navigation menu includes HOME, READ (underlined), SIGN, BOAI20, BOAI15, BOAI10, and MORE. The main content area is titled 'Read the Declaration' and features a section for the 'BUDAPEST OPEN ACCESS INITIATIVE' with a large 'A' starting the text. A right sidebar contains 'ABOUT THE DECLARATION' and 'TRANSLATIONS' with a list of languages including English, Arabic, Bengali (PDF), Catalan, and Chinese (Simplified). Two buttons, 'SIGN DECLARATION' and 'FAQ', are visible in the sidebar.

PERCHÈ OPEN ACCESS?

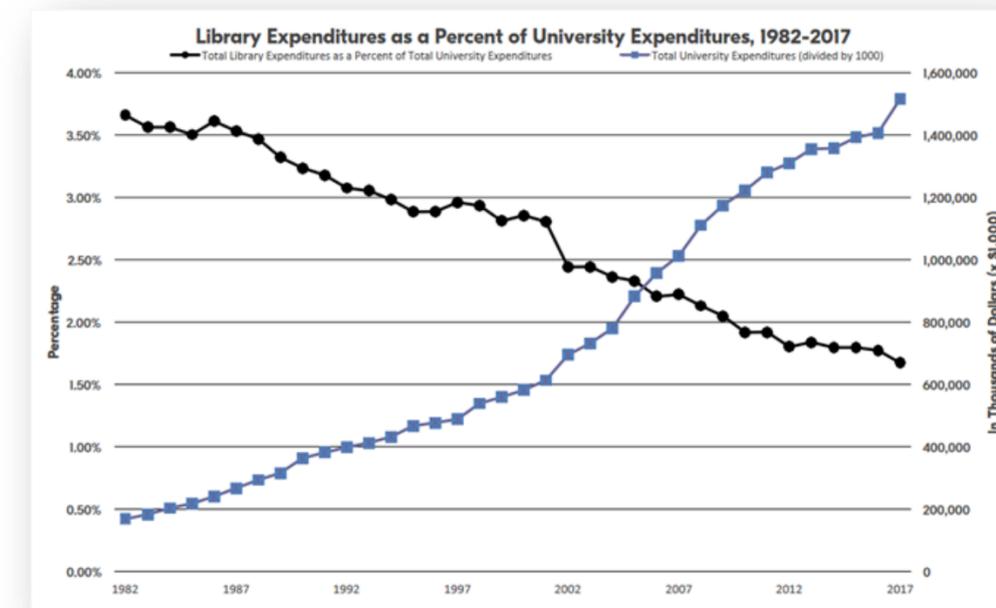
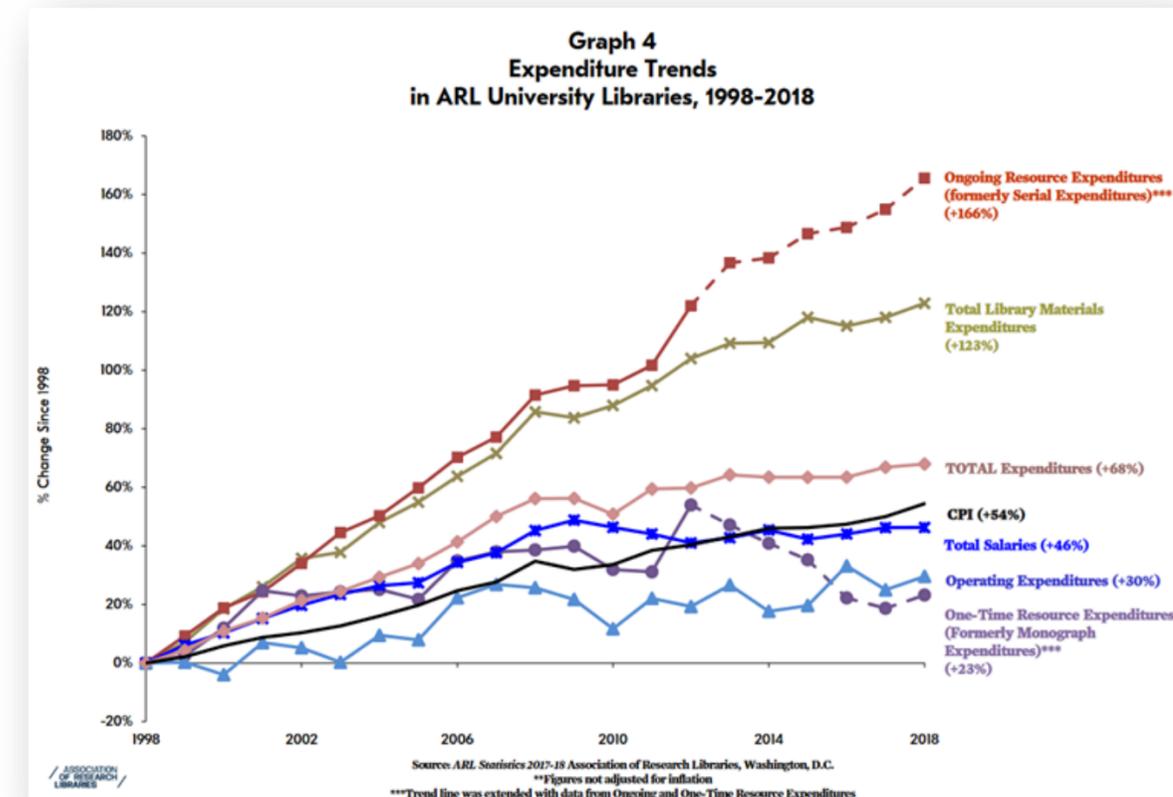
Le istituzioni si sono rese conto di pagare la ricerca 4 volte:

- Stipendi dei ricercatori
- Finanziamento della ricerca
- Abbonamenti a riviste
- Diritti di riutilizzo

Gli autori e i revisori sono pagati dalle università.

Il risultato del lavoro viene rivenduto alle biblioteche dell'Università.

Forte aumento delle spese per i periodici mentre il bilancio della biblioteca continua a essere tagliato



DA OPEN ACCESS A OPEN DATA

The screenshot shows the website for Open Access Max-Planck-Gesellschaft. The header includes navigation links for Contact, Sitemap, MPG, and Deutsch. The main content area is titled "BERLIN DECLARATION" and features a quote: "The Internet has fundamentally changed the practical and economic realities of distributing scientific knowledge and cultural heritage. For the first time ever, the Internet now offers the chance to constitute a global and interactive representation of human knowledge, including cultural heritage and the guarantee of worldwide access. We, the undersigned, feel obliged to address the challenges of the Internet as an emerging functional medium for distributing knowledge. Obviously, these developments will be able to significantly modify the nature of scientific publishing as..." Below the quote, there is a section titled "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities" with a preface. To the right, there are sections for "SIGNING INSTRUCTIONS" and "PDF VERSIONS BERLIN DECLARATION".

BERLIN DECLARATION
The Internet has fundamentally changed the practical and economic realities of distributing scientific knowledge and cultural heritage. For the first time ever, the Internet now offers the chance to constitute a global and interactive representation of human knowledge, including cultural heritage and the guarantee of worldwide access. We, the undersigned, feel obliged to address the challenges of the Internet as an emerging functional medium for distributing knowledge. Obviously, these developments will be able to significantly modify the nature of scientific publishing as...

Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities
The Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities of 22 October 2003 was written in English. It is one of the milestones of the Open Access movement. The wording of the English version shall prevail.

Preface
The Internet has fundamentally changed the practical and economic realities of distributing scientific knowledge and cultural heritage. For the first time ever, the Internet now offers the chance to constitute a global and interactive representation of human knowledge, including cultural heritage and the guarantee of worldwide access.

SIGNING INSTRUCTIONS
Governments, universities, research institutions, funding agencies, foundations, libraries, museums, archives, learned societies and professional associations who share the vision expressed in the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities are therefore invited to join the signatories that have already signed the Declaration.

Contact
Prof. Dr. Martin Stratmann
Präsident der Max Planck Gesellschaft
Hofgartenstraße 8
D-80539 München
Deutschland
Email: President or Open Access Contact

PDF VERSIONS BERLIN DECLARATION

- Arabic
- Chinese (Simp)
- Chinese (traditional)
- English
- French
- German
- Greek
- Italian
- Polish
- Portuguese
- Russian
- Spanish

Solo un anno dopo l'Iniziativa di Budapest per l'accesso aperto, la Dichiarazione di Berlino sull'accesso aperto alla conoscenza nelle scienze e nelle discipline umanistiche (2003) affermava:

Affermare l'accesso aperto come procedura utile richiede idealmente **l'impegno attivo di ogni singolo produttore di conoscenza scientifica e detentore di patrimonio culturale.**

I contributi ad accesso aperto comprendono **i risultati originali della ricerca scientifica, i dati grezzi e i metadati, i materiali di partenza, le rappresentazioni digitali di materiali pittorici e grafici e il materiale multimediale scientifico.**

DA OPEN DATA A OPEN SCIENCE...

By *scholarly communication* we mean the processes of producing, reviewing, organising, disseminating and preserving scholarly knowledge¹. Scholarly communication does not only concern researchers, but also society at large, especially students, educators, policy makers, public administrators, funders, librarians, journalists, practitioners, publishers, public and private organisations, and interested citizens.

Nel 2016 (con molte tappe intermedie) sono stati pubblicati i Principi di Vienna:

L'attenzione si sposta sulla comunicazione scientifica in generale.

Esprime la convinzione che "la scienza aperta ha il potenziale per cambiare in meglio il funzionamento del sistema di comunicazione accademica".

...IL PUNTO IN CUI CI TROVIAMO OGGI

UNESCO Recommendation on Open Science (2021)

<https://unesdoc.unesco.org/ark:/48223/pf0000379949/PDF/379949eng.pdf.multi>



UNESCO Recommendation on Open Science



June 2021
"Horizon Europe will set a new standard for dissemination of knowledge and new skills across European societies. With clear and immediate open access requirements for beneficiaries, the Open Europe Research publishing platform and a strengthened European Open Science Cloud, we are well underway in making truly open science a reality."

Mariya Gabriel Commissioner for Innovation, Research, Culture, Education and Youth

OPEN SCIENCE

EARLY KNOWLEDGE AND DATA SHARING, AND OPEN COLLABORATION

The challenge is for Europe to embrace open science as the modus operandi for all researchers. Open science consists in the sharing of knowledge, data and tools as early as possible in the Research and Innovation (R&I) process, in open collaboration

EC, Directorate-General for R&I, Horizon Europe, open science : early knowledge and data sharing, and open collaboration, 2021,

<https://data.europa.eu/doi/10.2777/18252>



EC, Research, Innovation and Science Policy Experts (2018)

<https://data.europa.eu/doi/10.2777/348700>

PERCHÈ È VANTAGGIOSO PARLARE DI OPEN SCIENCE?



Open Science è un argomento attuale in diversi contesti di finanziamento della ricerca.



La ricerca e i ricercatori si muovono in un contesto internazionale:

- in Europa le strategie legate all'Open Science sono varie e numerose;
- i vostri colleghi in tutta Europa sono formati anche per rispondere a queste richieste.



I metodi tradizionali di valutazione della ricerca sono in evoluzione:

- pratiche di Open Science entreranno a far parte dei nuovi criteri.

OPEN SCIENCE È UN ARGOMENTO ATTUALE IN DIVERSI CONTESTI DI FINANZIAMENTO DELLA RICERCA

Bando PRIN 2022

Articolo 13

Open access

1. Ciascun responsabile di unità garantisce l'accesso gratuito e *on-line* (almeno in modalità *green access*) ai risultati ottenuti e ai contenuti delle ricerche oggetto di pubblicazioni scientifiche “*peer-reviewed*” nell'ambito del progetto, secondo quanto previsto dall'art. 4, commi 2 e 2 bis, del decreto legge 8 agosto 2013, n. 91, convertito con modificazioni dalla legge 7 ottobre 2013, n. 112.
2. Restano impregiudicati tutti gli eventuali obblighi di riservatezza o di tutela dei **dati** personali.
3. Come eccezione, i responsabili di unità sono esentati dall'assicurare l'accesso aperto a parti specifiche dei propri dati di ricerca, se l'accesso aperto a tali dati dovesse compromettere il raggiungimento del principale obiettivo della ricerca stessa.



CANCER
RESEARCH
UK

UK Research
and Innovation

wellcome trust

NHS
National Institute for
Health Research

THE ROYAL
SOCIETY

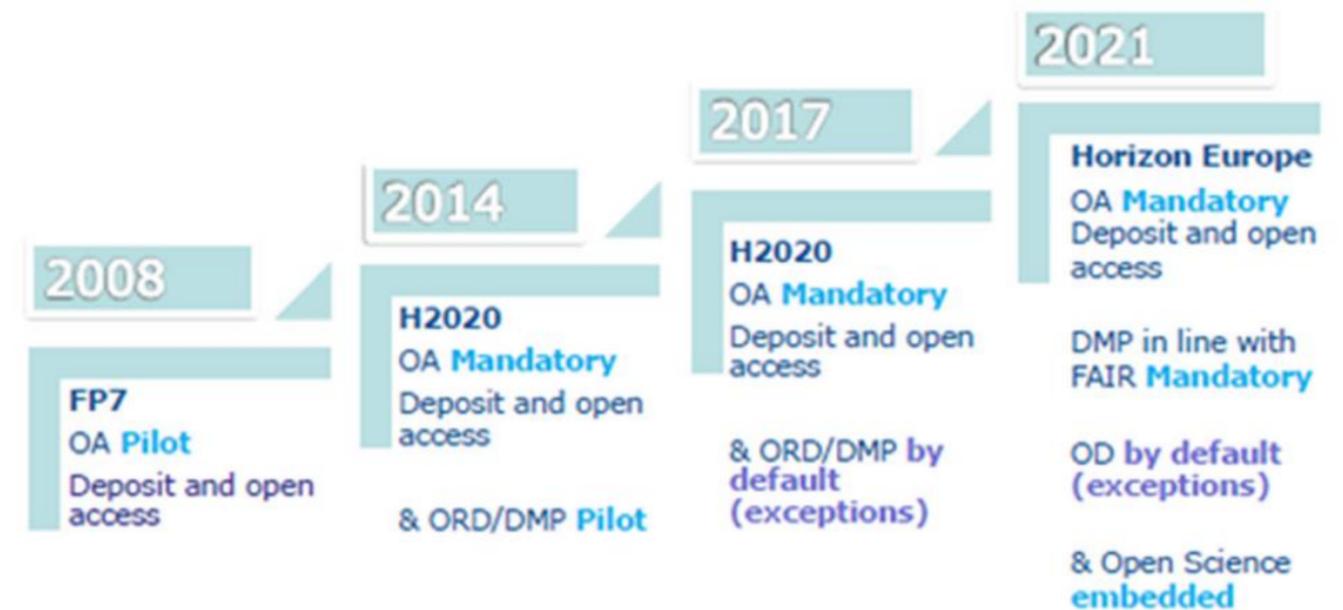
BILL & MELINDA
GATES foundation

NWO

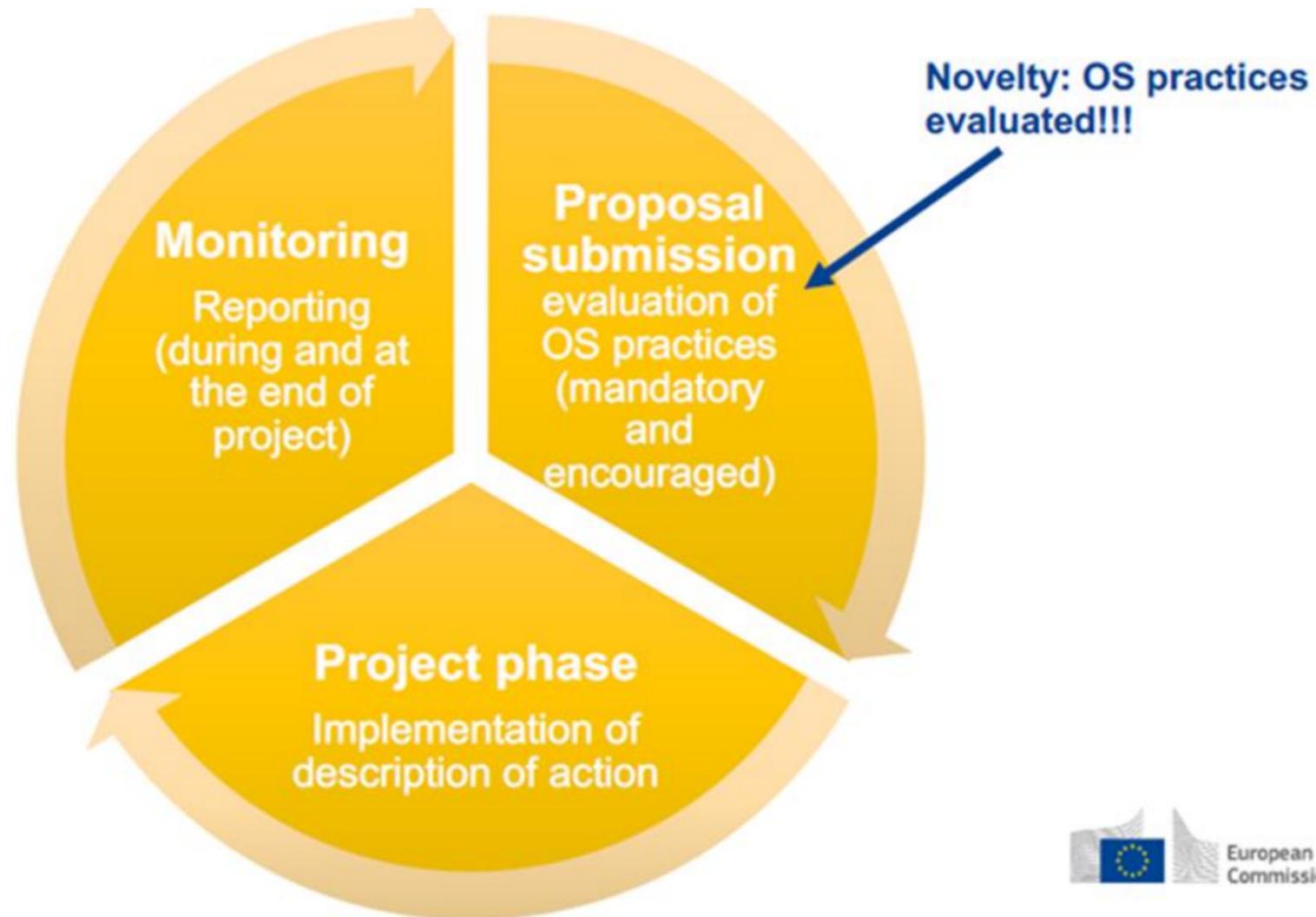
LA SCIENZA APERTA NEI FINANZIAMENTI EUROPEI

COSA? [...] la condivisione di conoscenze, dati e strumenti il più presto possibile nel processo di ricerca e innovazione (R&I), in collaborazione aperta con tutti gli attori della conoscenza, comprese le università, l'industria, le autorità pubbliche, gli utenti finali, i cittadini e la società in generale.

PERCHÈ? ha il potenziale per aumentare la qualità, l'efficienza e l'impatto della R&I, portare a una maggiore capacità di risposta alle sfide sociali e aumentare la fiducia della società nel sistema scientifico. Un obiettivo strategico dal 2015,



LA SCIENZA APERTA NEI FINANZIAMENTI EUROPEI



Pratiche obbligatorie:

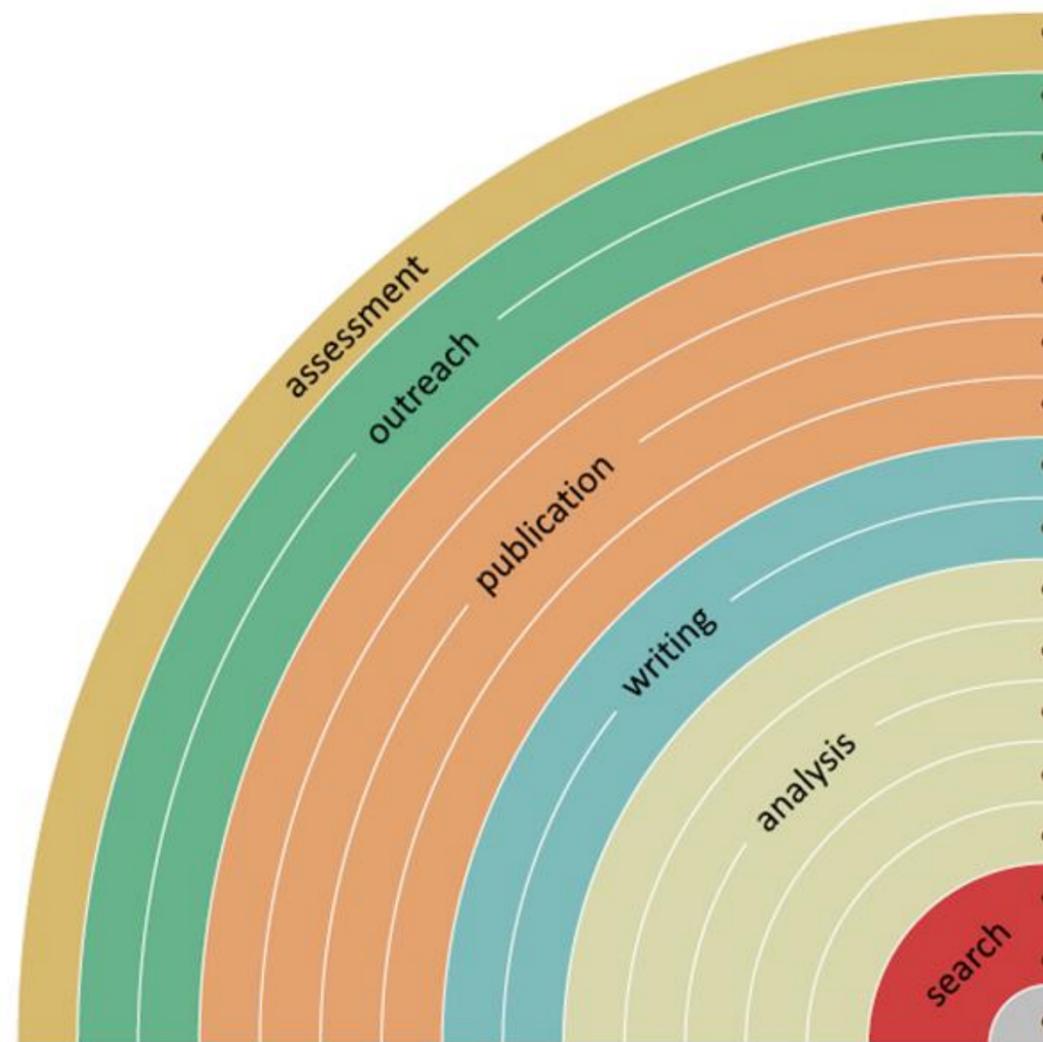
- accesso aperto alle pubblicazioni
- gestione dei dati e degli output della ricerca in linea con i principi FAIR
- DMP, accesso aperto ai dati "il più possibile aperto, il più possibile chiuso".
- accesso digitale o fisico ai risultati della ricerca in caso di emergenza pubblica, accesso aperto immediato a tutti gli output

Pratiche Raccomandate:

- strumenti e pratiche per migliorare la collaborazione, la trasparenza e l'integrità della ricerca
- gestione del flusso di ricerca, collaborazione, riproducibilità
- co-creazione con gli stakeholder e gli utenti finali

LE PRATICHE DI OPEN SCIENCE SONO VARIE E NUMEROSE

You can make your workflow more open by ...



- adding alternative evaluation, e.g. with altmetrics
- communicating through social media, e.g. Twitter
- sharing posters & presentations, e.g. at FigShare
- using open licenses, e.g. CC0 or CC-BY
- publishing open access, 'green' or 'gold'
- using open peer review, e.g. at journals or PubPeer
- sharing preprints, e.g. at OSF, arXiv or bioRxiv
- using actionable formats, e.g. with Jupyter or CoCalc
- open XML-drafting, e.g. at Overleaf or Authorea
- sharing protocols & workfl., e.g. at Protocols.io
- sharing notebooks, e.g. at OpenNotebookScience
- sharing code, e.g. at GitHub with GNU/MIT license
- sharing data, e.g. at Dryad, Zenodo or Dataverse
- pre-registering, e.g. at OSF or AsPredicted
- commenting openly, e.g. with Hypothes.is
- using shared reference libraries, e.g. with Zotero
- sharing (grant) proposals, e.g. at RIO



IL CURRICULUM NARRATIVO: DOVE DARE CONTO DI QUANTO SI LAVORA "OPEN"

List of up to 5 publications, widely-used datasets, software, goods, services, or any other achievements relevant to the call content.

Type of achievement	Short description
[Publication]	Key elements of the achievement, including a short qualitative assessment of its impact and (where available) its digital object identifier (DOI) or other type of persistent identifier (PID). Publications, in particular journal articles, are expected to be open access. Datasets are expected to be FAIR and 'as open as possible, as closed as necessary'.
[Dataset]	
[Software]	
[Good]	
[Service]	

Narrative CV format

The narrative CV is based on the premise that there is no ideal type of researcher. Different research projects require varying talents. The narrative format allows applicants to highlight their own academic profile without being limited by prescribed questions and criteria. The narrative CV consists of two parts:

- The academic profile (narrative)
- Key outputs

The key outputs section focuses on the candidate's past performance:

- A broad definition of 'output' is used. This includes not only scholarly articles but also other outputs, such as datasets, patents and software.
- The number of outputs has been maximized to 10 to stress that NWO values quality over quantity.
- Special attention is paid to Open Science by asking applicants to indicate which of their outputs are openly available.

The narrative CV has been introduced as a pilot in the Veni (part of [NWO Talent Programme](#)) and is gradually being adopted in other NWO funding instruments.

Fase di proposal dei progetti Horizon Europe, considerati per valutazione:

- esperienza in termini di Open Science
- non solo pubblicazioni
- pubblicazioni in Open Access/ dataset FAIR
- no Impact Factor per valutare la qualità delle pubblicazioni

Esempio Olandese: il Dutch Research Council, ente finanziatore nazionale olandese, ha adottato il CV Narrativo per valutare le applicazioni dei ricercatori ai bandi di finanziamento (<https://www.nwo.nl/en/dora>)

LA RIFORMA DELLA VALUTAZIONE DELLA RICERCA

Coalition for Advancing Research Assessment

Our vision is that the assessment of research, researchers and research organisations recognises the diverse outputs, practices and activities that maximise the quality and impact of research. This requires basing assessment primarily on qualitative judgement, for which peer review is central, supported by responsible use of quantitative indicators.

I believe in a research culture that recognises a diversity of contributions to science and society; that celebrates high quality and impactful research; and that values sharing, collaboration, integrity and engagement with society, transmitting knowledge from generation to generation.

Mariya Gabriel

Commissioner for Innovation, Research, Culture, Education and Youth



Italian National Agency for
the Evaluation of Universities
and Research Institutes
(ANVUR)

[website](#) →

ANVUR firmataria dell'Agreement on reforming research assessment

📅 06/10/2022

ANVUR ha sottoscritto in data odierna l' **"Agreement on reforming research assessment"**, che costituisce il documento programmatico iniziale per una discussione sulla ridefinizione delle strategie e metodologie per la valutazione delle attività di ricerca, da condurre attraverso un percorso comune e condiviso a livello internazionale. In questo senso la partecipazione dell'Agenzia alla Coalition internazionale che si sta definendo è un ulteriore elemento che testimonia la forte volontà dell'Agenzia di partecipare attivamente alle iniziative di respiro internazionale su un tema centrale della sua attività. La decisione, assunta dal Consiglio Direttivo dell'ANVUR nella seduta del 3 ottobre u.s., rappresenta un punto di partenza per le attività che l'Agenzia dovrà sviluppare per la definizione di regole comuni a livello internazionale a partire da questo documento programmatico. La firma giunge al termine di un percorso di partecipazione attiva e di confronto intrapreso attraverso la partecipazione del prof. Menico Rizzi, componente del Consiglio Direttivo con delega alla ricerca sulla valutazione e rappresentante dell'Agenzia in seno alla Coalition internazionale.

IL PROGRAMMA NAZIONALE PER LA RICERCA

SOMMARIO

PERCHÉ LA SCIENZA APERTA	
IL CONTESTO	
L'apertura come paradigma scientifico	
Gli strumenti per l'attuazione della scienza aperta.....	
L'impatto della scienza aperta.....	
STRUTTURA E OBIETTIVI DEL PIANO	
ASSI DI INTERVENTO	
1. LE PUBBLICAZIONI SCIENTIFICHE.....	
Il panorama	
Il piano di intervento	
2. I DATI DELLA RICERCA SCIENTIFICA	
Il panorama	
Il piano di intervento	
3. LA VALUTAZIONE DELLA RICERCA	
Il panorama	
Il piano di intervento	
4. SCIENZA APERTA, COMUNITA' SCIENTIFICA E PARTECIPAZIONE EUROPEA	
Il panorama	11
Il piano di intervento	12
5. APERTURA DEI DATI DELLA RICERCA SU SARS-COV-2 E COVID-19.....	12
Il panorama	12
Il piano di intervento	13
GLOSSARIO	14
DEFINIZIONI	16
DOCUMENTI DI RIFERIMENTO	16
RIFERIMENTI NORMATIVI.....	19



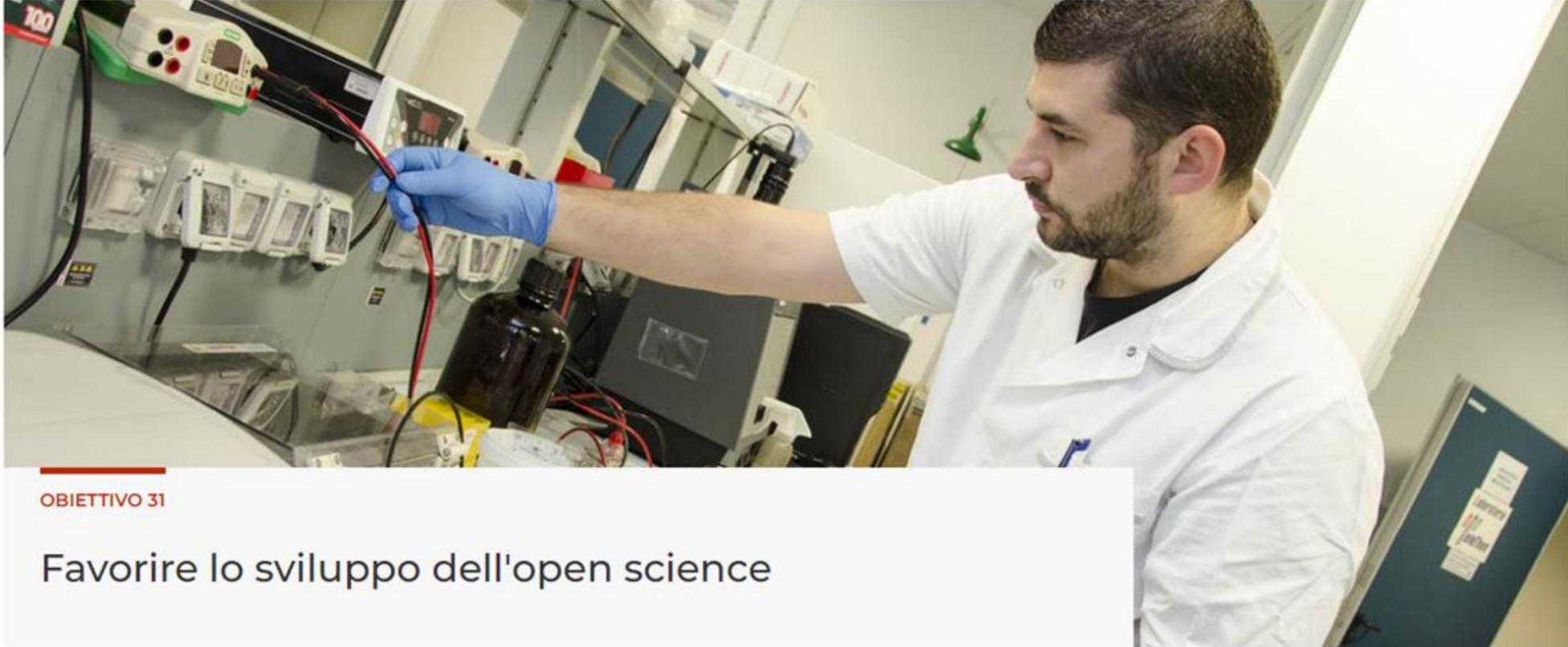
PIANO NAZIONALE PER LA SCIENZA APERTA

OPEN SCIENCE NEL PIANO STRATEGICO DI UNIBO

 **PIANO STRATEGICO 2022-2027**

Home Mission e vision **Principi** Obiettivi It En

Home > Principi > Incrementare il senso della nostra responsabilità sociale in tutte le attività



OBIETTIVO 31

Favorire lo sviluppo dell'open science

P3 PRINCIPIO 3  RICERCA



QUALCHE PUNTO CHIAVE

- Open Science è uno standard di lavoro che si fonda sulla cooperazione e la diffusione della conoscenza.
- Contiene in sé principi di trasparenza, integrità della ricerca, collaborazione piena tra attori diversi, libertà di accesso a dati e pubblicazioni, nuove metriche di valutazione.
- E' finalizzato a migliorare la qualità, l'efficienza e la capacità della ricerca di dare risposte concrete alle sfide poste



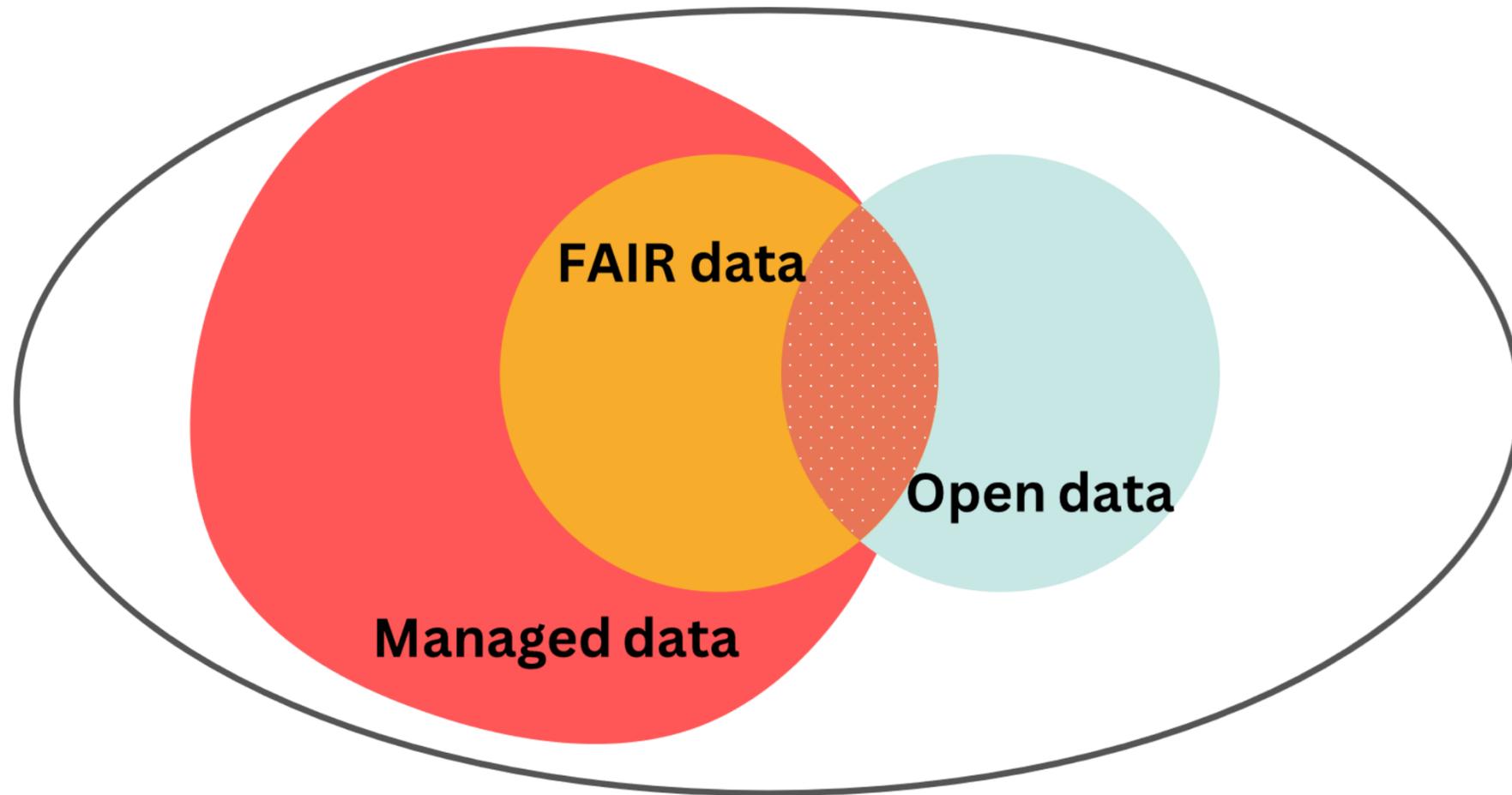




I PRINCIPI FAIR



FAIR: COS'È?



- I principi FAIR descrivono come i risultati della ricerca dovrebbero essere organizzati per essere più facilmente accessibili, compresi, scambiati e riutilizzati.
- Non tutti i dati gestiti correttamente sono FAIR.
- Non tutti i dati FAIR sono dati Open.

SCIENTIFIC DATA

Amended Addendum

OPEN

SUBJECT CATEGORIES

- » Research data
- » Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*[#]

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data by a diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers. We have come together to design and jointly endorse a concise and measurable set of principles to act as the FAIR Data Principles. The intent is that these may act as a guideline for those who wish to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on data management, the FAIR Principles put specific emphasis on enhancing the ability of machines to find and use the data, in addition to supporting its reuse by individuals. This Comment is a formal publication of the FAIR Principles, and includes the rationale behind them, and some early implementations in the community.

Received: 10 December 2015

Accepted: 12 February 2016

Published: 15 March 2016

Box 2 | The FAIR Guiding Principles

To be Findable:

- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
 - A1.1 the protocol is open, free, and universally implementable
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

To be Interoperable:

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

To be Reusable:

- R1. meta(data) are richly described with a plurality of accurate and relevant attributes
 - R1.1. (meta)data are released with a clear and accessible data usage license
 - R1.2. (meta)data are associated with detailed provenance
 - R1.3. (meta)data meet domain-relevant community standards

The FAIR4RS Principles are:

<p>F: Software, and its associated metadata, is easy for both humans and machines to find.</p> <p>F1. Software is assigned a globally unique and persistent identifier.</p> <ul style="list-style-type: none"> F1.1. Components of the software representing levels of granularity are assigned distinct identifiers. F1.2. Different versions of the software are assigned distinct identifiers. <p>F2. Software is described with rich metadata.</p> <p>F3. Metadata clearly and explicitly include the identifier of the software they describe.</p> <p>F4. Metadata are FAIR, searchable and indexable.</p>
<p>A: Software, and its metadata, is retrievable via standardized protocols.</p> <p>A1. Software is retrievable by its identifier using a standardized communications protocol.</p> <ul style="list-style-type: none"> A1.1. The protocol is open, free, and universally implementable. A1.2. The protocol allows for an authentication and authorization procedure, where necessary. <p>A2. Metadata are accessible, even when the software is no longer available.</p>
<p>I: Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.</p> <p>I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards.</p> <p>I2. Software includes qualified references to other objects.</p>
<p>R: Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).</p> <p>R1. Software is described with a plurality of accurate and relevant attributes.</p> <ul style="list-style-type: none"> R1.1. Software is given a clear and accessible license. R1.2. Software is associated with detailed provenance. <p>R2. Software includes qualified references to other software.</p> <p>R3. Software meets domain-relevant community standards.</p>

Table 1: The FAIR Principles for Research Software

☰ README.md

FAIR 4 Research Software (FAIR4RS) WG [↗](#)

This working group has completed its outcomes. Thanks for your contributions!!! The [RDA Software Source Code Interest Group](#) is the maintenance home for the principles. Concerns or queries about the principles can be raised at RDA plenary events organised by the SSC IG, where there may be opportunities for adopters to report back on progress.

Cite the FAIR4RS principles [↗](#)

- Barker, M., Chue Hong, N.P., Katz, D.S. et al. Introducing the FAIR Principles for research software. *Sci Data* 9, 622 (2022). <https://doi.org/10.1038/s41597-022-01710-x>
- May 24th, 2022. The RDA Council have endorsed the FAIR4RS Principles as an official output! Citation and download: Chue Hong, N. P., Katz, D. S., Barker, M., Lamprecht, A-L, Martinez, C., Psomopoulos, F. E., Harrow, J., Castro, L. J., Gruenpeter, M., Martinez, P. A., Honeyman, T., et al. (2022). FAIR Principles for Research Software version 1.0. (FAIR4RS Principles v1.0). Research Data Alliance. DOI: <https://doi.org/10.15497/RDA00068>

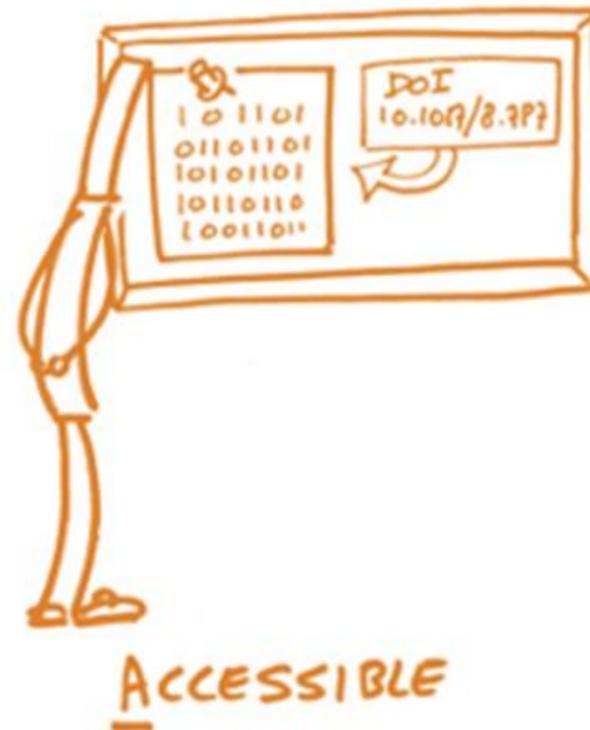
★ Star us on GitHub — it will help you find us again.

contributions
welcome
issues 0 open
pull requests 0 open
+350 members
1 code of conduct

FAIR DATA PRINCIPLES



Il primo passo per (ri)utilizzare i dati è trovarli.



Una volta trovati, l'utente deve sapere come accedere (autenticazione, autorizzazione...)



I dati devono essere integrati con altri dati e devono interoperare con applicazioni o flussi di lavoro.



L'obiettivo finale di FAIR è ottimizzare il riutilizzo dei dati.

FAIR FOR HUMANS AND MACHINES



1) Both humans and machines are intended as digesters of data.

This will lead to the creation of an ecosystem that is fast to respond to change and automatically adapts to new findings or changes: the *Internet of FAIR Data and Services*. This is the reason for focusing on standards for data, identification mechanisms, data availability, etc.

F indable

A ccessible

I nteroperable

R eusable

Per gli esseri
umani e per le
macchine

F ully

A (rtificial)

I (ntelligence)

R eady

PUNTI DI ATTENZIONE PER APPLICARE I PRINCIPI FAIR

- I principi FAIR si applicano sia ai **dati** che ai **metadati**.
- I principi non riguardano necessariamente i dati **aperti** («as open as possible, as closed as necessary»).
- I principi FAIR non sono regole o standard, ma un adattamento delle **routine di lavoro**.

COSA INTENDIAMO QUANDO PARLIAMO DI METADATI?

Informazioni strutturate sui dati

- Li descrivono, spiegano, localizzano
- Ne facilitano l'uso o la gestione
- Possono essere aggiunti manualmente o automaticamente
- Esistono standard specifici per disciplina



COSA INTENDIAMO QUANDO PARLIAMO DI METADATI PER I DATI DIGITALI?

```
{
  "@context": "https://raw.githubusercontent.com/codemeta/codemeta/master/codemeta.jsonld",
  "@type": "SoftwareSourceCode",
  "identifier": "codemetar",
  "description": "Codemeta defines a 'JSON-LD' format for describing software metadata.\n",
  "name": "codemetar: Generate CodeMeta Metadata for R Packages",
  "codeRepository": "https://github.com/codemeta/codemetar",
  "issueTracker": "https://github.com/codemeta/codemetar/issues",
  "license": "https://spdx.org/licenses/MIT",
  "version": "0.1.0",
  "programmingLanguage": {
    "@type": "ComputerLanguage",
    "name": "R",
    "version": "3.4.0",
    "url": "https://r-project.org"
  },
  "runtimePlatform": "R version 3.4.0 (2017-04-21)",
  "author": [
    {
      "@type": "Person",
      "givenName": "Carl",
      "familyName": "Boettiger",
      "email": "cboettig@gmail.com",
      "@id": "http://orcid.org/0000-0002-1642-628X"
    }
  ],
}
```

Metadati
descrittivi

Metadati
strutturali

```
  "copyrightHolder": [
    {
      "@type": "Person",
      "givenName": "Carl",
      "familyName": "Boettiger",
      "email": "cboettig@gmail.com",
      "@id": "http://orcid.org/0000-0002-1642-628X"
    }
  ],
}
```

Metadati
amministrativi

```
"softwareRequirements": [
  {
    "@type": "SoftwareApplication",
    "name": "jsonlite",
    "version": "1.3",
    "provider": {
      "@id": "https://cran.r-project.org",
      "@type": "Organization",
      "name": "Central R Archive Network (CRAN)",
      "url": "https://cran.r-project.org"
    }
  },
  {
    "@type": "SoftwareApplication",
    "name": "utils"
  },
  {
    "@type": "SoftwareApplication",
    "name": "methods"
  },
  {
    "@type": "SoftwareApplication",
    "name": "stats"
  },
  {
    "@type": "SoftwareApplication",
    "name": "R",
    "version": "3.0.0"
  }
],
"contIntegration": "https://travis-ci.org/codemeta/codemetar",
}
```

FILE CITAZIONALI: DOCUMENTAZIONE OLTRE AI METADATI

Che cos'è un file CITATION.cff?

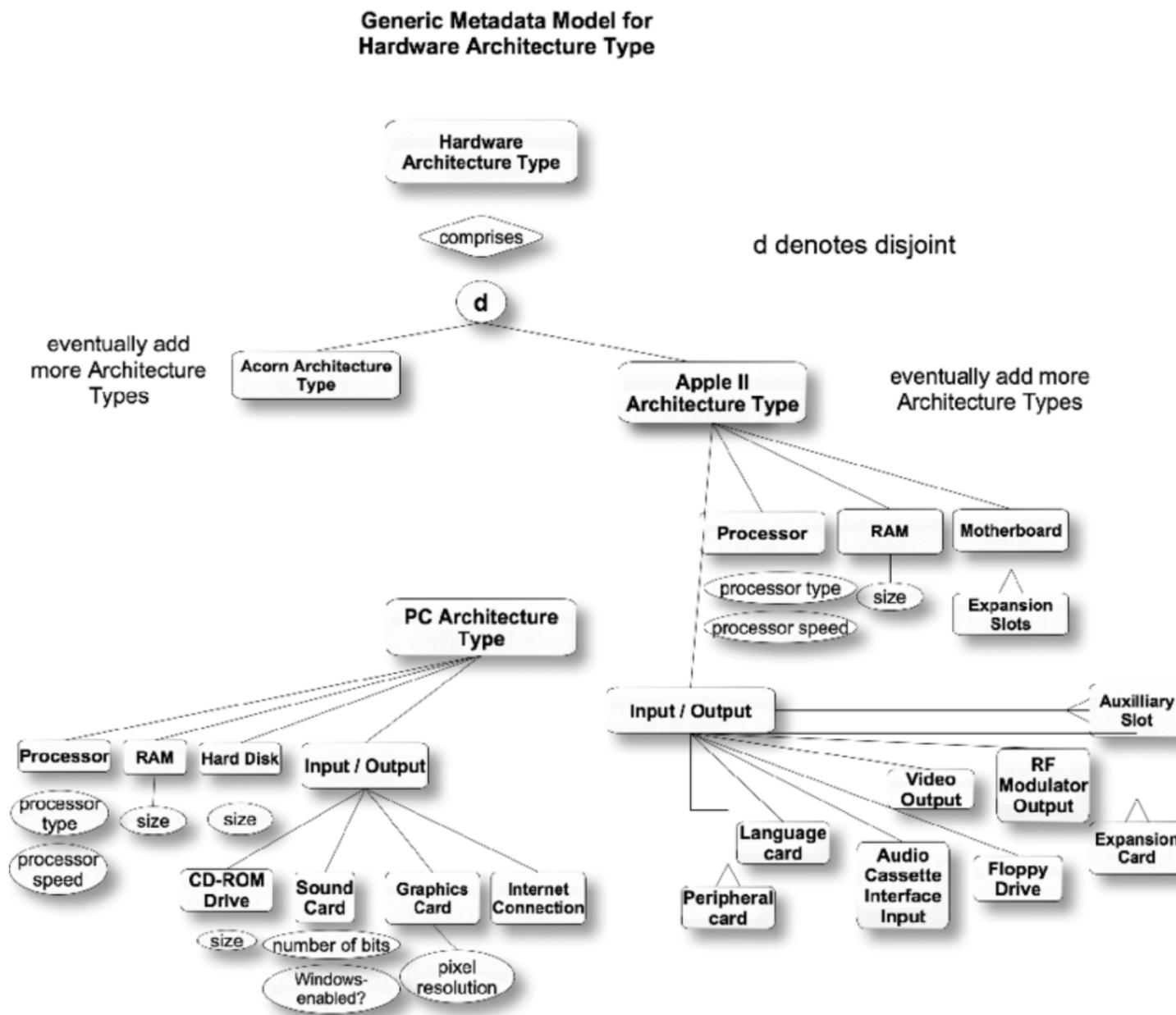
I file CITATION.cff sono file di testo semplice con informazioni di citazione per il software (e i dataset) leggibili dall'essere umano e dai sistemi informatici. Gli sviluppatori possono includerli nei loro repository per indicare agli altri come citare correttamente il loro software.

Questo è un esempio di un semplice file CITATION.cff:

```
cff-version: 1.2.0
message: "If you use this software, please cite it as below."
authors:
  - family-names: Druskat
    given-names: Stephan
    orcid: https://orcid.org/1234-5678-9101-1121
title: "My Research Software"
version: 2.0.4
doi: 10.5281/zenodo.1234
date-released: 2021-08-11
```

<https://citation-file-format.github.io/>

COSA INTENDIAMO QUANDO PARLIAMO DI METADATI PER GLI OUTPUT FISICI?



64 AIDA64 Business

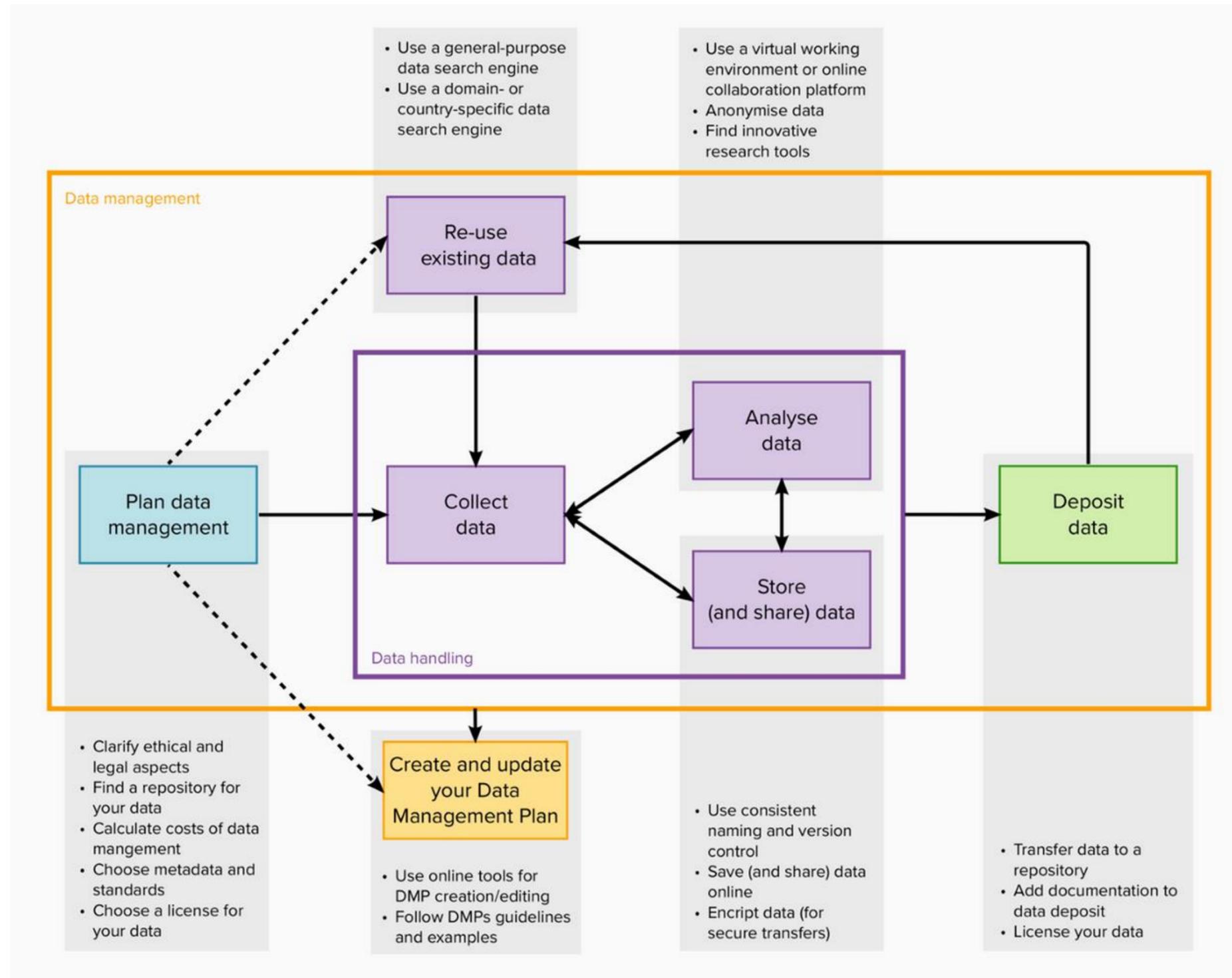
File View Report Remote Favorites Tools Help

< > ^ ↻ 👤 📈 📄 Report

Menu	Field	Value
64 AIDA64	Motherboard Properties	
> Computer	Motherboard ID	63-0100-000001-00101111-041312-Chipset\$B85-HD3_BI...
▼ Motherboard	Motherboard Name	Gigabyte GA-B85-HD3
CPU		
CPUID		
Motherboard		
Memory		
SPD		
Chipset		
BIOS		
ACPI		
> Operating System		
> Server		
> Display		
> Multimedia		
> Storage		
> Network		
> DirectX		
> Devices		
> Software		
> Security		
> Config		
> Database		
> Benchmark		
	Front Side Bus Properties	
	Bus Type	BCLK
	Real Clock	97 MHz
	Effective Clock	97 MHz
	Memory Bus Properties	
	Bus Type	Dual DDR3 SDRAM
	Bus Width	128-bit
	DRAM:FSB Ratio	24:3
	Real Clock	779 MHz (DDR)
	Effective Clock	1558 MHz
	Bandwidth	24926 MB/s
	Chipset Bus Properties	
	Bus Type	Intel Direct Media Interface v2.0
	Motherboard Physical Info	
	CPU Sockets/Slots	1 LGA1150
	Expansion Slots	2 PCI, 2 PCI-E x1, 2 PCI-E x16

Fornire metadati e documentazione digitali quando l'output di ricerca è fisico

I PRINCIPI FAIR SONO TRASVERSALI: DA APPLICARE DURANTE L'INTERO CICLO DI VITA DEL DATO





F
indable



- Ai dataset devono essere assegnati identificatori persistenti (ad esempio, DOI, Handle, URN).
- I dataset devono essere accompagnati da metadati e parole chiave significative (contenenti il PID).



A

ccessible



- I metadati devono essere apertamente accessibili attraverso l'uso di protocolli standard.
- I dati possono essere apertamente accessibili (default) o accessibili attraverso un sistema di autenticazione e autorizzazione, se la natura dei dati ne impedisce l'apertura.
- I dati e i metadati devono essere conservati in archivi o **depositi** che li rendano persistenti nel tempo e rintracciabili in rete.

"AS OPEN AS POSSIBLE AS CLOSED AS NECESSARY"

- Lavorate con dati **personali**? Seguire il GDPR e proteggere la privacy
- Lavorate con dati **sensibili**? Seguire il GDPR, potrebbero essere coinvolti i comitati etici.
- Volete riutilizzare i dati in base ai diritti di proprietà intellettuale o al **copyright**?
- Volete **brevettare**? Prevedere un embargo sui dati
- Volete pubblicare (ad esempio, una monografia) e temete che le idee possano essere rubate? Si può imporre un **embargo** e/o pubblicare i dati sottostanti insieme alla pubblicazione.

- 
- Consenso informato
 - Anonimizzazione
 - Accesso controllato
 - Comitati di accesso

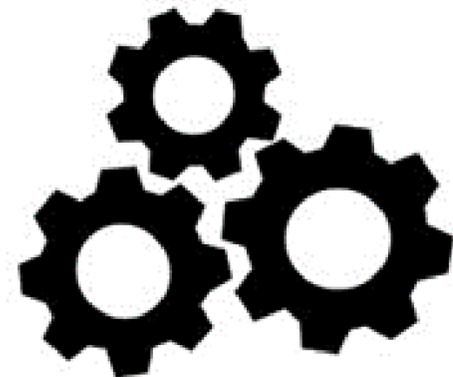


Controllare le licenze! Il riutilizzo può essere limitato, poiché l'accessibilità dei nuovi risultati è limitata.

- 
- Deposita i dati con embargo
 - Accesso controllato



I nteroperable

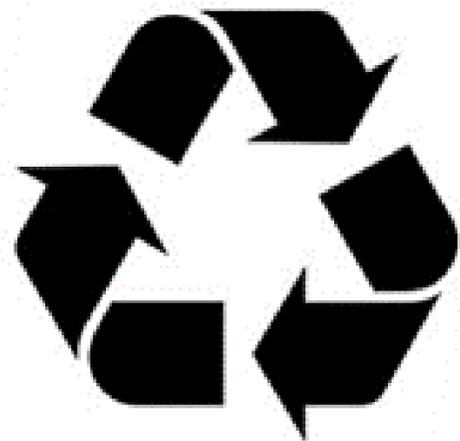


- I dati devono essere combinabili e utilizzabili con altri dati o strumenti.
- Il formato dei dati deve essere aperto e interoperabile da vari strumenti, compresi altri database.
- I metadati devono utilizzare un linguaggio standardizzato e condiviso a livello internazionale dai diversi servizi di indicizzazione.
- Includere riferimenti qualificati ad altri (meta)dati.

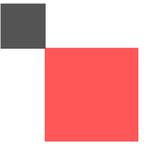


R

Reusable



- I dati devono essere descritti e documentati nel miglior modo possibile, per garantirne la qualità e per poter essere replicati e combinati in contesti diversi.
- L'elaborazione dei dati deve essere conforme agli standard riconosciuti dalle comunità scientifiche di riferimento.
- Il riutilizzo dei (meta)dati deve essere dichiarato sotto una o più licenze aperte chiare e accessibili (ad esempio, Creative Commons, GPL per il software e altre).



COME VALUTARE SE UN DATASET È FAIR?

TESTIAMO UN DATASET INSIEME:



F-UJI is a web service to programmatically assess FAIRness of research data objects at the dataset level based on the FAIRsFAIR Data Object Assessment Metrics [↗](#)

[Click here to assess a dataset](#)

<https://www.fairsfair.eu/f-uji-automated-fair-data-assessment-tool>

<https://www.f-uji.net/>

<https://zenodo.org/record/8378595>



QUALCHE PUNTO CHIAVE

- I dati della ricerca devono essere gestiti in modo trasparente: seguire i principi FAIR e renderli reperibili, accessibili, interoperabili e riutilizzabili.
- I dati FAIR non sono necessariamente dati gestiti correttamente né tantomeno Open Data → "As open as possible, as closed as necessary".
- Un repository per i dati non fa tutto il lavoro per rendere i vostri dati il più FAIR possibile. Tuttavia, offre un'ottima struttura per ottenere le basi giuste.



CONSIGLI DI APPROFONDIMENTO

Open Science:

Open Science: A Practical Guide for Early-Career Researchers (1.0) <https://doi.org/10.5281/zenodo.7716153>

Open Science Toolkit (UNESCO) <https://www.unesco.org/en/open-science/toolkit>

Portale Italiano Open-Science <https://open-science.it>

Webinar Open Science Cafè <https://www.icdi.it/it/attivita/competence-centre/open-science-cafe>

Gestione FAIR dei Dati di Ricerca:

Open AIRE guidelines <https://guidelines.openaire.eu/en/latest/>

The Turing Way handbook <https://the-turing-way.netlify.app/index.html>

The hands-on guide to research data management for KU Leuven researchers, students, and research support staff in the humanities and social sciences. <https://doi.org/10.5281/zenodo.4596323>

LEGO® Metadata for Reproducibility game pack <http://eprints.gla.ac.uk/196477/>

Reprojuice, the reproducibility game <https://seriousgaming.tudelft.nl/game/reprojuice>

CONSIGLI DI APPROFONDIMENTO

Gestione FAIR dei Dati di Ricerca – strumenti e infrastrutture per discipline di area tecnologica:

- Software FAIR:
 - 5 recommendations for FAIR software: <https://fair-software.eu/>
- Licenze per il software:
 - Choose a License: <https://choosealicense.com/>
 - Interactive License Selector for code and data: <https://ufal.github.io/public-license-selector/>
 - (Software) Licenses in Plain English: <https://www.tldrlegal.com/>
- Metadati e citazioni per il software:
 - CodeMeta generator: <https://codemeta.github.io/codemeta-generator/>
 - Citation File Format con generatore automatico di file citazionali di metadati: <https://citation-file-format.github.io/>, <https://citation-file-format.github.io/cff-initializer-javascript/#/>
- Altro:
 - Awesome Research Software Registries on GitHub: <https://github.com/NLeSC/awesome-research-software-registries>
 - Quality check del software: <https://fair-software.eu/recommendations/checklist>
 - SciCodes Consortium of scientific software registries and repositories, <https://scicodes.net/>
 - Awesome Open Science for software and computational science: <https://github.com/INRIA/awesome-open-science-software>
 - Referencing and citing content using GitHub and Zenodo: <https://docs.github.com/en/repositories/archiving-a-github-repository/referencing-and-citing-content>
 - Open repository of tech datasets hosted by IEEE: <https://ieee-dataport.org/>





GRAZIE

