

CORSI DI FORMAZIONE



**Consiglio Nazionale
delle Ricerche**
ILIESI

Gestire dati FAIR e Open

Elena Giglia

Elena Giglia

elena.giglia@unito.it

 @egiglia



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/). Photos are mine. Feel free to reuse from Flickr/eg65

In questo modulo impareremo:

1. come gestire i propri dati correttamente e rendere la ricerca più efficace

2. la differenza fra dati FAIR e Open

MESSAGGI CHIAVE

- gestire bene i dati è nell'interesse di chi fa ricerca - non c'entra Open!
- solo dati gestiti bene possono essere resi FAIR e se possibile aperti
- NON ci sarà una ricetta per DMP, ma strumenti utili (da imparare)

[DMP]

... lo so...

NON USCIRETE DI QUI OGGI CON UN DMP PRONTO
MA CON GLI STRUMENTI PER SCRIVERLO – OGNUNO DIVERSO

- 1) NON È FACILE GESTIRE I DATI
- 2) NON C'È UNA RICETTA, OGNI DATASET UNICO
- 3) CI SONO MOLTI ASPETTI DA CONSIDERARE
- 4) MOLTI STRUMENTI DA IMPARARE A USARE
- 5) SEMBRA RICHIEDERE COSÌ TANTO TEMPO
- 6) MA I BENEFICI SONO ENOOOOOOOOOORMI

<http://www.insideoutproject.net/>

« On ne sait pas
ce que nous réserve
l'avenir,
alors profitons »

<http://www.insideoutp>

« Laissez
être heur

...«dati» è ancora una parolaccia nelle scienze umane?

fact that data in the humanities are also an effect of operationalisation and interpretive processes.



Are artistic/humanities disciplines/methodologies only considered valid if we call them 'data'?



In other words, can we not accept different kinds of research methodologies as valid *on their own terms*, rather than on those imported from (or imposed by) other fields? #munin2018



In the humanities, we all use research data, although we may not be aware of it. It is like in the case of Monsieur Jourdain, the title character of Molière's *Le Bourgeois gentilhomme*, who learnt, to his great satisfaction, that unwittingly he had been speaking prose all his life. With research data in the humanities it is exactly the same: you are using it, even if you don't know it, and once you realise it, it will affect your research workflow forever.

PRIMA LI CHIAMIAMO «DATI»
MEGLIO È

of Social Anthropologists. One reason for the uneven adoption of the term could be that in the humanities, "[w]e resist the blanket term 'data' for the very good reason that we have more and precise terminology (e.g. primary sources, secondary sources, theoretical documents, bibliographies, critical editions, annotations, notes, etc.) available to us to describe and make transparent our research processes" (Edmond & Tóth-Czifra, 2018:1). The resistance to 'data' in the humanities, as an oversimplifying abstraction of complex phenomena, was voiced by many critics, most notably by Johanna Drucker (2011), who opposed the objectifying term 'data' (something given) and proposed to use 'capta' (something captured, taken) instead. This

...sta cambiando

What do we mean by “data”? A proposed classification of data types in the arts and humanities

2022

Bianca Gualandi, Luca Pareschi, Silvio Peroni

2024

Journal of Documentation

DOWNLOADS ALTMETRICS

Originality/value

Our findings confirm that “data” within the FAIR framework should include all types of inputs and outputs humanities research work with, including publications. Also, the participants of this study appear ready for a discussion around making their research data FAIR: they do not find the terminology particularly problematic, while they rely on precise and recognised methodologies, as well as on sharing and collaboration with colleagues.

allea | All European Academies

Working Group E-Humanities addresses challenges in digital humanities, focusing on transparency in linking resources to publications, recognising updates as scholarly contributions, reevaluating authorship, fostering digital skills, and adjusting evaluation methods.

The second section offers recommendations for assessing specific digital outputs like editions, databases, infographics, code, blogs, and podcasts. Each case study includes practical examples and suggested readings.

2023

Year : 2023

DOI: 10.26356/OUTPUTS-DH

[Read on Zenodo](#)



Download



Recognising Digital Scholarly Outputs in the Humanities

ALLEA Report | November 2023



• ...vi è mai capitato...

DI PERDERE DATI?

DI AVERE I DATI SUL COMPUTER
DEL PhD CHE POI SE NE È
ANDATO??

DI APRIRE DATI ALTRUI E NON
RIUSCIRE A LEGGERLI??

DI CHIEDERE DATI DOPO AVER
LETTO UN ARTICOLO E
L'AUTORE NON LI TROVA PIÙ??

DI NON RIUSCIRE AD APRIRE
UN FILE PER IL FORMATO
OBSOLETO?

...O SEMPLICEMENTE DI
ESSERE FELICI QUANDO
TROVATE DATI CONDIVISI DA
ALTRI CHE POTETE RIUSARE?

Perché occuparci dei dati

GRAZIE PER L'ATTENZIONE, PER OGGI
ABBIAMO FINITO...

Digital Science Report
The State of Open Data 2021

The longest-running longitudinal survey and analysis on open data

Foreword by Natasha SIMONS, Australian Research Data Commons (ARDC)

2021

November 2021

Open data saves lives. The global pandemic has highlighted beyond anything that came before it the importance of data sharing in solving the big challenges of our time. COVID-19 data may be the

Perché occuparci dei dati?

1. SONO IL FONDAMENTO DI UNA RICERCA SOLIDA



2. IL COVID HA DIMOSTRATO CHE SERVONO I DATI, NON SOLO GLI ARTICOLI, E SERVONO SUBITO

3. I DATI SONO FRAGILI, SI PERDONO

4. ALCUNI SONO UNICI E NON POSSONO ESSERE RIPRODOTTI (METEO, TERREMOTI...)

5. POSSONO ESSERE MANIPOLATI, GESTIRLI GARANTISCE INTEGRITÀ

6. PERMETTONO VALIDAZIONI E RIPRODUCIBILITÀ

7. I DATI CREANO PONTI FRA LE DISCIPLINE


8. POSSONO ESSERE RIUTILIZZATI (IN MODO INEDITO)

CAREER COLUMN | 01 February 2024 2024

PER EVITARE GARBAGE
IN/GARBAGE OUT CON
L'INTELLIGENZA ARTIFICIALE


In the AI science boom, beware: your results are only as good as your data

Machine-learning systems are voracious data consumers – but trustworthy results require more vetting both before and after publication.

By [Hunter Moseley](#) 



Marco Rossi • 3rd+ 2024

Marketing & Digital Temporary Manager | ex Sky - Mond...
5d • 

Spunti interessanti da [Netcomm Forum](#).

[Shalini Kurapati, PhD](#) pone l'accento sul concetto di etica del dato:

- ▶ non c'è intelligenza artificiale senza dati
- ▶ l'80% del lavoro nell'AI è relativo all'acquisizione di dati
- ▶ ma i dati devono essere di qualità
- ▶ e l'utente è sempre più attento e sensibile alla sua privacy

Perché occuparci dei dati?

PERCHÉ DOBBIAMO.
DIRETTIVA OPEN DATA

PERCHÉ DOBBIAMO. C'È
EOSC

26.6.2019 IT Gazzetta ufficiale dell'Unione europea L 172/56

DIRETTIVA (UE) 2019/1024 DEL PARLAMENTO EUROPEO E DEL CONSIGLIO

del 20 giugno 2019

relativa all'apertura dei dati e al riutilizzo dell'informazione del settore pubblico
Open data directive

EOSC Association
Advancing Open Science in Europe

I DATI DELLA RICERCA SONO INCLUSI
NELLA DIRETTIVA «OPEN DATA».
DEVONO ESSERE APERTI

PERCHÉ DOBBIAMO.
UN NUMERO CRESCENTE DI
RIVISTE CHIEDE DI PUBBLICARLI
INSIEME ALL'ARTICOLO PER
TRASPARENZA E RIPRODUCIBILITÀ

PERCHÉ DOBBIAMO. IN HORIZON
EUROPE I DATI SONO SIA NELLE
PRATICHE OBBLIGATORIE CHE IN QUELLE
RACCOMANDATE

ANNEX 5

V.1 Feb 2021



Horizon Europe (HORIZON)
Euratom Research and Training
(EURATOM)
General Model Grant A
EIC Accelerator Co
DE MGA - M&I & M&I
Open Access
© 2021 Horizon Europe

**COMMUNICATION, DISSEMINATION, OPEN SCIENCE AND VISIBILITY (—
ARTICLE 17)**

Open science: research data management

The beneficiaries must manage the digital research data generated in the action ('data') responsibly, in line with the FAIR principles and by taking all of the following actions:

In Italia

- Raccomandazione (UE) **2018/790** della Commissione Europea del 25 aprile 2018
Accesso all'informazione scientifica e sulla sua conservazione
[non vincolante]
- Direttiva (UE) **2019/1024** del Parlamento e del Consiglio Dell'UE del del 20 giugno 2019
Apertura dei dati e al riutilizzo dell'informazione del settore pubblico
[si limita a stabilire degli obiettivi]
- Regolamento (UE) **2023/138** della Commissione Europea del 21 dicembre 2022
Elenco di serie di dati di elevato valore e le relative modalità di pubblicazione e riutilizzo
[specifico e vincolante]

Il legislatore nazionale ha chiarito aspetti fondamentali (1 di 2)

La Direttiva UE 2019/1024 è stata recepita con il D.Lgs. 8 novembre 2021 n. 200 2023
che modifica il D.Lgs. 24 gennaio 2006 n. 36 definendo:

- **Titolare del dato:** *la pubblica amministrazione o l'organismo di diritto pubblico che ha originariamente formato per uso proprio o commissionato ad altro soggetto pubblico o privato il documento che rappresenta il dato, o che ne ha la disponibilità*
[art. 2, c. 1, lett. e]
- **Licenze:** *le pubbliche amministrazioni, gli organismi di diritto pubblico e le imprese pubbliche nonché le imprese private di cui all'articolo 1, comma 2-quater, adottano licenze standard, disponibili in formato digitale, per il riutilizzo dei propri documenti*
[art. 8, c. 1]
- **Licenza standard per il riutilizzo:** *il contratto, o altro strumento negoziale, ove possibile redatto in forma elettronica, compatibile con le licenze pubbliche standardizzate disponibili online, nel quale sono definite le modalità di riutilizzo in formato digitale dei documenti*
[art. 2, c. h]

Il legislatore nazionale ha chiarito aspetti fondamentali (2 di 2)

Con la modifica del D.Lgs. 24/01/2006 n.36 viene introdotto l'Art. 9-bis Dati della ricerca

- 1. I dati della ricerca sono riutilizzabili a fini commerciali o non commerciali conformemente a quanto previsto dal presente decreto legislativo, nel rispetto della disciplina sulla protezione dei dati personali, ove applicabile, degli interessi commerciali, nonché della normativa in materia di diritti di proprietà intellettuale ai sensi della legge 22/4/1941, n. 633, e dei diritti di proprietà industriale ai sensi del decreto legislativo 10/2/2005, n. 30.*
- 2. La previsione del comma 1 si applica nelle ipotesi in cui i dati siano il **risultato di attività di ricerca finanziata con fondi pubblici e quando gli stessi dati siano resi pubblici, anche attraverso l'archiviazione in una banca dati pubblica**, da ricercatori, organizzazioni che svolgono attività di ricerca e organizzazioni che finanziano la ricerca, tramite una banca dati gestita a livello istituzionale o su base tematica.*
- 3. I dati della ricerca di cui ai commi precedenti rispettano i requisiti di **reperibilità, accessibilità, interoperabilità e riutilizzabilità**.*

2023



Nessun dato?

Is withholding your data simply bad science, or should it fall under scientific misconduct?

22 comments | 5 shares

Estimated reading time: 5 minutes



A recent study sent data requests to 200 authors of economics articles where it was stated 'data available upon request'. Most of the authors refused. What does the scientific community think about those withholding their data? Are they guilty of scientific misconduct? **Nicole Janz** argues that if you don't share your data, you are

breaking professional standards in research, and are thus committing scientific misconduct.

Classifying data secrecy as misconduct may be a harsh, but it is a necessary step.



Alastair Dunning

@alastairdunning

Following

To me, data are like footnotes. I might not always read them, but I get suspicious if they are not there.

Traduci dalla lingua originale: inglese

12:49 - 27 feb 2018

<https://twitter.com/alastairdunning/status/968453078218395648>

2 Retweet 8 Mi piace



NESSUN DATO?
PIGRIZIA O FRODE?
I DATI COME NOTE A PIE'
PAGINA: POSSO NON
LEggerLE, MA DIVENTO
SOSPETTOSO SE NON CI SONO

2015

Gold Standard
Research Integrity

Questionable Research
Practices

Scientific
Misconduct



Data secrecy

Open data
Open code
Pre-registration
Version control

P-hacking
Sloppy statistics
Peer review abuse
Inappropriate research design
Not answering to replicators
Lying about authorships

Fabrication
Falsification
Plagiarism

Parliamo di dati

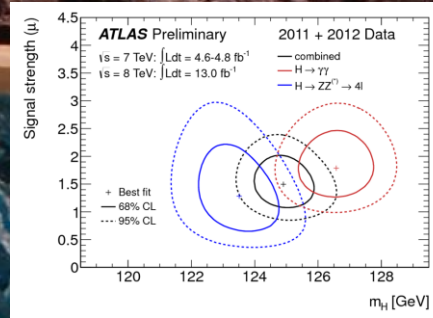


Table S1. Number of reads per prokaryotic operational taxonomic unit (OTU) and sample for

A04_TP1	A04_T1
25707	6
119	
73	2
382	
15	
1843	
5429	
29	
0	

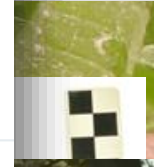
 **Wilma van Wezenbeek**
 @wvanwezenbeek

Following

#osc2018 Wolfram Horstmann wants us to talk about datadiversity, like we do with biodiversity #openscience

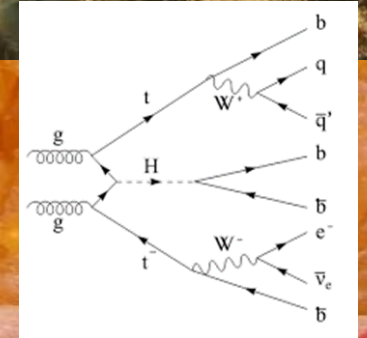
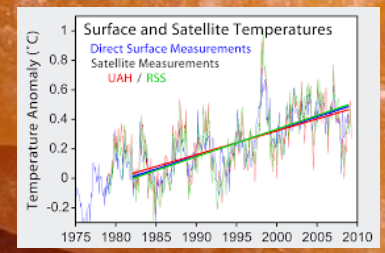
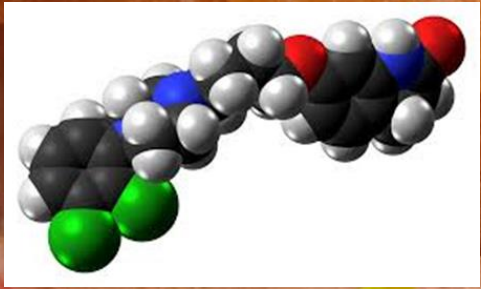
Traduci il Tweet
 12:51 - 13 mar 2018
 3 Retweet 1 Mi piace
<https://twitter.com/wvanwezenbeek/status/973527086685093893>

Batus vir, qui non abiit in oratio in pioz, et in via peccatorum non stetit: et in cathedra petulencie non sedet. Sed in lege domini voluntas eius: et in lege eius meditabitur die ac nocte. Et erit tamquam lignus, quod plantatus est secus decursus aquarum: quod fructus suum dabit in tempore suo. Et folium eius non defluet: et omnia quecumque faciet prosperabunt. Non sic impij, non sic: sed



Gaucelm Faidit

I.
 Ara nos sia guitz
 lo vers dieus Iesu Cristz,
 car de franca gen gaia
 soi per Lui partitz,
 on ai estat noiritz
 et onratz e grazitz;
 per so-l prec no-ill desplaia
 s'ieu m'en vauc marritz.
 A! gentils lemozis,
 el vostr'onrat pais
 lais de bella paria
 seignors e vezis
 e domnas ab pretz fis,
 pros, de gran cortesia,
 don plane e languis
 e sospir nueg e dia.



Data



We could then define data in the humanities broadly as all materials and assets scholars collect, generate and use during all stages of the research cycle. In this report we focus on digital assets.

DATI = TUTTO CIÒ CHE VIENE RACCOLTO, GENERATO E USATO NEL PROCESSO DI RICERCA

PENSATE A TUTTI GLI ELEMENTI DELLA VOSTRA RICERCA COME «DATI» CHE POSSONO ESSERE RIUSATI DA ALTRI. E CONSIDERATE QUANTO SAREBBE UTILE PER VOI POTER RIUSARE DATI DI ALTRI



RECOMMENDATIONS

» Think of all your research assets as research data that could be potentially reused by other scholars. Consider how useful it would be for your own work if others shared their data.

I 3 passi

E GESTIRE I DATI CORRETTAMENTE È NELL'INTERESSE
PRIMARIO DI CHI FA RICERCA,
PERCHÉ L'INTERA RICERCA SCORRE PIÙ FLUIDA

OPEN

FAIR

MANAGED

1. I DATI DEVO ESSERE «AS OPEN AS POSSIBLE»

2. MA SE I DATI NON SONO «FAIR», APRIRLI COMPORTA RISCHI
(USO SCORRETTO, CATTIVE INTERPETAZIONI, ...)

3. MA SE I DATI NON SONO CORRETTAMENTE GESTITI, RENDERLI
«FAIR» COSTA TROPPO TEMPO E DENARO. CON EOSC, DATI GESTITI E
DATI FAIR TENDONO A COINCIDERE, FAIR BY DESIGN

1. Gestire i dati

DESCRIZIONE
(metadati)

RIPRODUCIBILITÀ

ORGANIZZAZIONE
(file naming,
folders,
versioning...)

DOCUMENTAZIONE



ASPETTI LEGALI

CONSERVAZIONE
SUL LUNGO
PERIODO

BACKUP E
STORAGE

LUNGO TUTTO IL CICLO DI VITA

2. I dati devono essere FAIR BY DESIGN

To be Findable:

F1. (meta)data are assigned a globally unique and eternally persistent identifier.

F2. data are described with rich metadata.

F3. (meta)data are registered or indexed in a searchable resource.

F4. metadata specify the data identifier.

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 RE-USABLE:

R1. meta(data) have a plurality of accurate and relevant attributes.

R1.1. (meta)data are released with a clear and accessible data usage li

R1.2. (meta)data are associated with their provenance.

R1.3. (meta)data meet domain-relevant community standards.

«AS FAIR AS POSSIBLE»

«ACCESSIBLE»

NON SIGNIFICA «OPEN».

I DATI POSSONO RIMANERE
CHIUSI PER VALIDE RAGIONI,
PURCHÉ ALTRI RICERCATORI (E
LE MACCHINE) SAPPIANO
DOVE TROVARLI E CON QUALI
CONDIZIONI DI ACCESSO

Sharing Data
Why share data
 2. Why share data?



3) Se possibile, renderlo

COME LE ENERGIE RINNOVABILI: IL RIUSO CREA NUOVO VALORE

"Open data is like a renewable energy source: it can be reused without diminishing its original value, and reuse creates new value."

Digital Science Report

The State of Open Data 2017
 A collection of analyses and articles about open data, curated by Figshare
 Foreword by Jean-Claude Burgelman
Oct. 2017
 OCTOBER 2017

Better research

- Demonstrates research integrity, as there is transparency and accountability in the production of the data
- Encourages research enquiry and debate
- Promotes innovation and potential new research
- Encourages the improvement of research
- Prevents research fraud

UNA RICERCA MIGLIORE

- INTEGRITÀ
- DIBATTITO
- RIUSO INEDITO

Better impact

- Enables peer scrutiny of the research findings, validating the work carried out
- Increases the visibility of the research
- Provides credit for the creation of the data
- Can lead to new collaborations
- Produces a public record of the research

UN MAGGIORE IMPATTO

- VISIBILITÀ
- CREDITO
- COLLABORAZIONI

Better value

- Avoids duplication of effort in data creation
- Provides resources for use in teaching and learning
- Meets funder requirements
- Ensures data can be re-visited for future research
- Maximises return on research investment
- Preparing data for sharing also prepares it for reuse

PIÙ VALORE

- EVITA DUPLICAZIONI
- MASSIMO RITORNO SU INVESTIMENTI

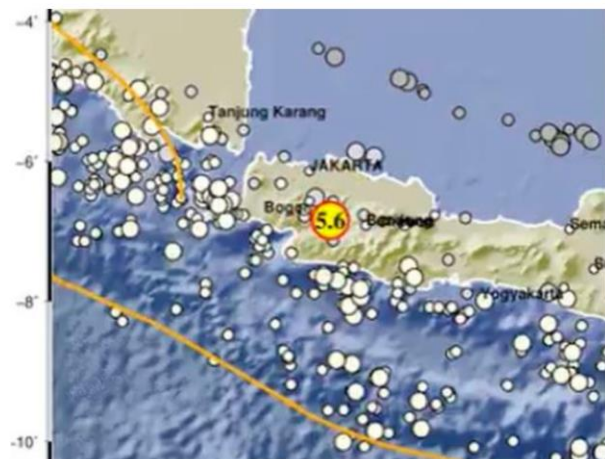
...il selfie...

How we can get those data

This was the best map that we can get (cited by the media)

Those data points are not really data points. They're just a selfie of data points.

They're not reusable.



IN «FAIR»
ACCENTO SU
«R»,
REUSABLE

RICORDATE...
SE I DATI
NON SONO ACCESSIBILI E RIUSABILI
SONO **SOLO UN SELFIE DI DATI**, QUINDI
DEL TUTTO INUTILI
[Dasapta Erwin Irawan]

[parlare di dati significa anche data stewards]



The number of people with these skills needed to effectively operate the EOSC is, we estimate, likely exceeding half a million within a decade. As we further argue below, we believe that the implementation of the EOSC needs to include instruments to help train, retain and recognise this expertise, in order to support the 1.7 million scientists and over 70 million people working in innovation⁹. The success of the EOSC depends upon it.

- SERVONO 500.00 DATA STEWARDS
- DATA STEWARDS SONO UNO DEI FATTORI DI SUCCESSO DI EOSC

Strategic Research and Innovation Agenda
(SRIA)
of the
European Open Science Cloud (EOSC)
SRIA 1.0 Version 1.0 15 February 2021

7.4. *Critical success factors*

The developments and expected impacts described above will not happen spontaneously. For these benefits to materialise a number of critical success factors (CSFs) must be in place. The following CSFs have been identified for EOSC:

- Researchers performing publicly funded research make relevant results available as openly as possible;
- Professional data stewards are available in research-performing organisations in Europe to help implement FAIR principles and support Open Science;



Time to re-think the divide between academic and support staff Apr. 2022

Research professionals should not be split into two categories, say Marta Teperék, Maria Cruz and Danny Kingsley.

DATA STEWARDS NON SONO «SUPPORTING STAFF». SONO PARTE DEL TEAM DI RICERCA

[competence profi

Education core content

This 1-year degree should build upon students' educational/job background through domain specific data knowledge and leverage with theoretical and practical competences. The education can be viewed as a Data Steward specialisation within the domain of their previous degree/jobs. The education contains **60 ECTS** and is expected to finish with a 15 ECTS project.

Preliminary Content

The 60 ECTS should be distributed among the following main areas:

- 22,5-30 ECTS: IT competences – including computational thinking, data modelling, data management, data harvesting, cleaning, and storing, infra-structure (storage & compute). An introduction to data science, machine learning, and their derived data needs.
 - 7,5-15 ECTS: Legal and ethical competences – including GDPR, FAIR, data security, and data & AI ethics.
 - 7,5-15 ECTS: Domain specific data competences – including knowledge about data, infrastructure, and practice within the students primary domain, e.g., health, life-science, finance/fintech, or the public sector.
 - 15 ECTS: Graduate project (possibly in collaboration with academia, industry, or the public sector)
- Competences such as project management, communication skills, and change management should be

Competence Profile

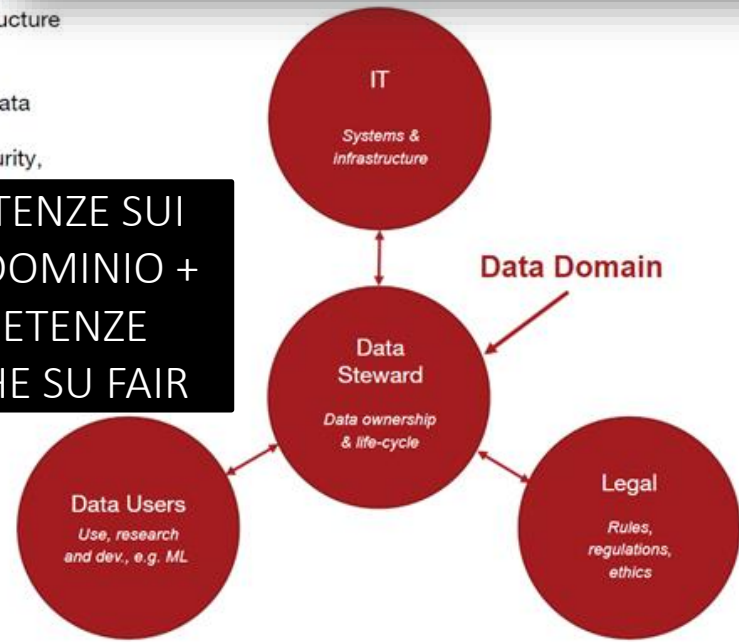
A data steward is a data specialist with strong domain-specific knowledge who understands and appreciates the relevance of data, data sources, data infrastructure and constraints within a scientific or other application domain.

The future Data Steward must assume ownership and responsibility for data, data quality, and the data life-cycle as their primary function. They should ensure collaboration and coherence between IT competences, quality assurance, security, rules & regulations, and facilitate the application and use of data internally and externally in the organisation.

Competence profile examples

- Domain-specific data understanding
- Ability to ensure that structured and unstructured data is modelled, harvested, stored, and maintained in documented, and regulated fashion with focus and findability, accessibility, interoperability, and reusability.
- Competences to facilitate HPC (High Performance Computing) during development and research through handling of large-scale data in public and private enterprises.
- Understanding of and competences within legal, ethical and security aspects of data handling, data sharing, e.g., integrity and GDPR.

COMPETENZE SUI DATI DI DOMINIO + COMPETENZE TECNICHE SU FAIR



Ma chi è il data steward?



Data stewardship is the responsible planning and executing of all actions on digital data before, during and after a research project, with the aim of optimising the usability, reusability and reproducibility of the resulting data.

It differs from data management, in the sense that data management concerns all actual, operational data-related activities in any phase of the data lifecycle, while data stewardship refers to the assignment of responsibilities in, and planning of, data management.

PIANIFICA E METTE IN ATTO
PRIMA, DURANTE E DOPO IL PROGETTO DI RICERCA,
LE AZIONI VOLTE A OTTIMIZZARE
L'USABILITÀ, RIUSABILITÀ E RIPRODUCIBILITÀ DEI DATI

ASSOCIATES, Data steward – cosa fa BACHELORS, MASTERS, DOCTORS O



The screenshot shows the Consortium GARR website. At the top, there is a navigation menu with links for 'CHI SIAMO', 'INFRASTRUTTURE', 'SERVIZI', 'PROGETTI', 'FORMAZIONE', and 'MEDIA'. Below the menu, a date 'Pubblicato il 18 Aprile 2023' is visible. The main content area features a banner for a survey titled 'Sondaggio sui Data Steward in Italia'. The banner includes the Consortium GARR logo, the Skills 4 Europe logo, and a photograph of a person's hands typing on a laptop. Below the banner, there is a short paragraph explaining the survey's purpose: 'Il sondaggio ha lo scopo di identificare la presenza dei Data Steward all'interno di Università e Istituti di Ricerca in Italia. Il Data Steward è un professionista specializzato nella gestione dei dati di ricerca e svolge un ruolo cruciale nella promozione della scienza aperta. Agisce come intermediario tra i ricercatori, le infrastrutture e le organizzazioni di'.

GARR news

Il Data Steward è un professionista specializzato nella gestione dei dati di ricerca e svolge un ruolo cruciale nella promozione della scienza aperta. Agisce come intermediario tra i ricercatori, le infrastrutture e le organizzazioni di ricerca, fornendo un supporto tecnico, disciplinare e legale nella gestione dei dati.

Il Data Steward lavora perciò a stretto contatto con i ricercatori per garantire che i dati della ricerca siano gestiti in modo responsabile e rispettoso della privacy e dei diritti di proprietà intellettuale. Può aiutare i ricercatori a scrivere i Data Management Plan (DMP), a identificare le migliori pratiche per la gestione dei dati, a garantire l'accessibilità e la condivisione dei dati e a garantire la conformità alle normative di protezione dei dati e può contribuire alla gestione del repository istituzionale.

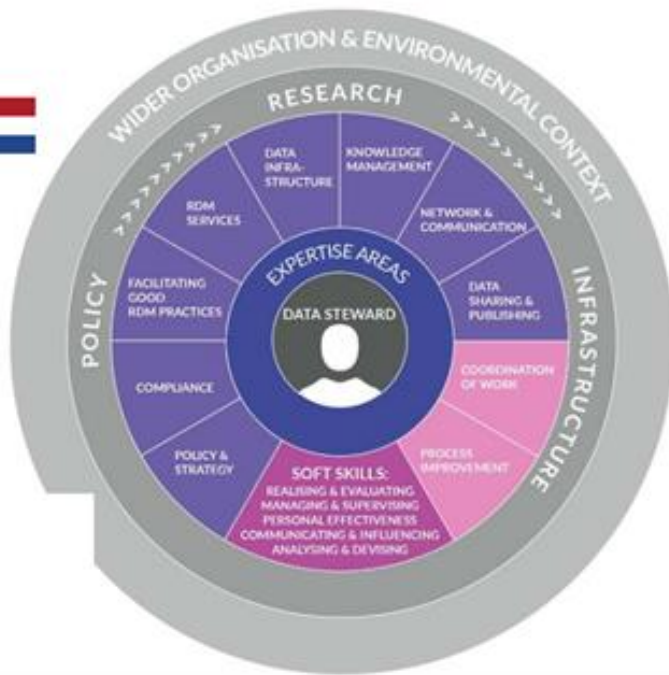
Il ruolo del data

Open Science Café

Il data steward: chi è, cosa fa e perché è fondamentale

3 DICEMBRE 2021 | SARA DI GIORGIO | NOTIZIE
ARTICOLO LETTO 4669 VOLTE

Chi è il «data steward»? (2)



Competenze

- ✓ Servizi di *research data management*, infrastrutture dati e di ricerca
- ✓ Gestione della conoscenza
- ✓ Conservazione e pubblicazione dei dati
- ✓ Conformità a leggi e normative istituzionali, nazionali ed internazionali in materia di gestione dei dati

Abilità

- ✓ Organizzazione e coordinamento del lavoro
- ✓ Comunicazione e sensibilizzazione
- ✓ Gestione e miglioramento di processi
- ✓ Didattica e formazione

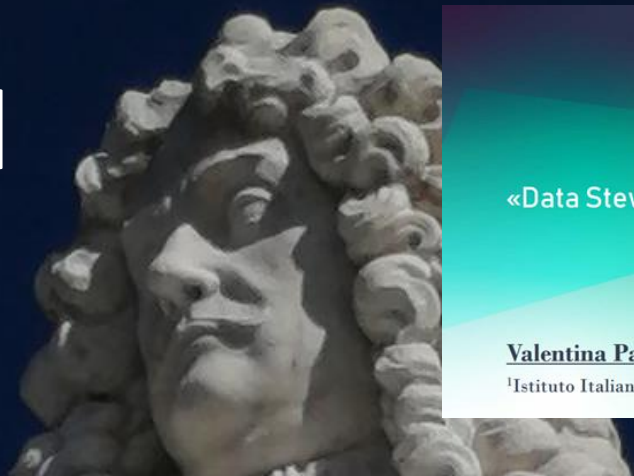
Table 3. Overview of the eight defined competence areas for all data steward roles

2019

Competence area	This concerns
Policy/strategy	Development, implementation and monitoring of research data management policy and strategy for the research institute
Compliance	Compliance to the Netherlands Code of Conduct for Academic Practice, the Netherlands Code of Conduct for Research Integrity, the General Data Protection Regulation (GDPR), and other relevant legal and ethical standards
Alignment with FAIR data principles	Alignment to the FAIR data principles and the principles of Open Science
Services	Availability of adequate support on research data management, in staff or services
Infrastructure	Availability of adequate data infrastructure for research data management
Knowledge management	Adequate level of knowledge and skills on research data management within the institute, department or project
Network	Obtaining and maintaining a network of aligned expertise areas and relevant departments and organisations inside and outside the institute, department or project
Data archiving	Adequate support and data infrastructure for FAIR and long-term archiving of data of the institute, department or project

8 AREE DI COMPETENZA
CHIAVE: FARE DA «PONTE»

Data steward



«Data Steward» per i dati FAIR

2021

it ISTITUTO ITALIANO DI TECNOLOGIA

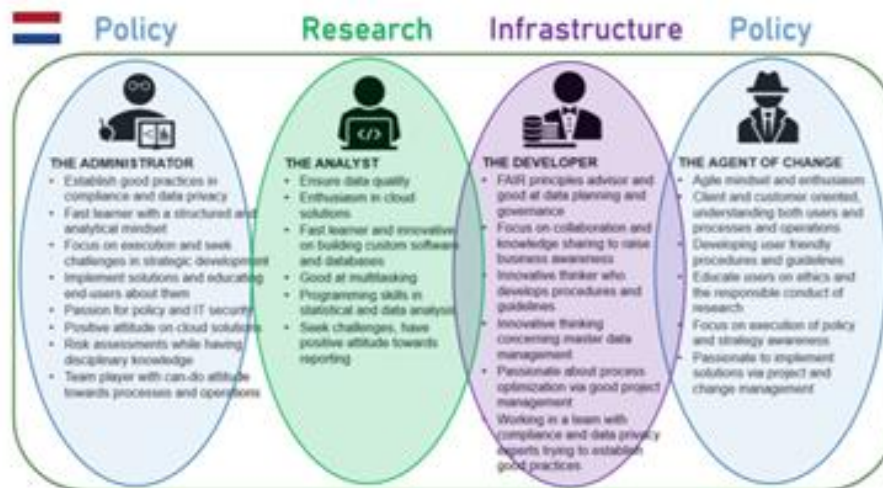
Valentina Pasquale¹, Emma Lazzeri², Elena Giglia³

¹Istituto Italiano di Tecnologia, ²GARR, ³Università di Torino



Chi è il «data steward» (2)

Profili professionali del data steward



Lorna Wildgaard et al. (2020). National Coordination of Data Steward Education in Denmark: Final report to the National Forum for Research Data Management. Zenodo. <https://doi.org/10.5281/zenodo.3609515>

I «data steward»...

... hanno (preferibilmente) un **PhD** e possiedono **nozioni su come i dati vengono gestiti in un dominio di ricerca specifico**

... possibilmente hanno esperienza pregressa in **programmazione, sviluppo software, gestione di database e infrastrutture di ricerca, sicurezza dei dati**

... hanno buone capacità **comunicative, di insegnamento e organizzative**

... possiedono nozioni su **aspetti legali della gestione dei dati (privacy, proprietà intellettuale) ed etici**

... **comprendono la psicologia dei ricercatori e parlano lo stesso linguaggio specifico**

... **desiderano intraprendere un percorso di carriera che non è né puramente scientifico né tecnico**

Servono competenze

Skills
4 eosc

ABOUT ▾

KERS

NEWS



<https://www.skills4eosc.eu/>

- Skills for the European
- Open Science
- Commons

PROGETTO SKILL4EOSC PER CURRICULA DATA STEWARD E CREAZIONE COMPETENCE CENTER

Objectives of the project are:

1. **Map career profiles related to Open Science** and define, through co-creation the "**Minimum Viable Skillset**" (MVS) for each of them; create a shared framework for the recognition of competencies acquired by university students, trainers and new professionals as a part of an academic path or a lifelong learning process.
2. **Define a methodology and a Quality Assurance process** to ensure the quality and relevance of OS learning materials and the management of their life-cycle, thus enhancing their re-usability.
3. **Offer training on OS and the usage of data in evidence-based policy for civil servants** and policymakers and empower CCs, researchers and "honest brokers" through the offering of resources to carry out training for this target.
4. **Define "OS and data-intensive science essentials"** for inclusion in generic undergraduate, postgraduate and PhD curricula as a key skill that anyone doing research is expected to acquire.
5. Design and implement a **collaboration model between national and regional CCs and international Research Infrastructures** and communities to provide specialised OS competencies targeting the needs of researchers and thematic RI professionals.
6. **Support lifelong learning** through professional networks as an enabling environment to discuss, cocreate and exchange best practices and solutions among OS professionals and researchers.
7. **Coordinate national, regional and thematic Competence Centres on OS and EOSC** in Europe and leverage their expertise to create a widespread user support network and an environment that fosters and harmonises training and skills activities.
8. Create and implement a strategy for engaging with **relevant stakeholders to co-create and promote the project outputs** (Curricula, shared certification and QA frameworks, human networks), building partnerships to embed project activities and results among the broadest network of stakeholders.
9. **Establish synergies with key actors within the Member States and in the EOSC arena**, and with human capital and training programmes at the national, regional and European levels to maximise the impact of the project activities and results and pave the way for their long-term sustainability.

La comunità



CHI SIAMO ▾ INFRASTRUTTURE ▾ SERVIZI ▾ PROGETTI ▾ FORMAZIONE ▾ MEDIA

Publicato il 18 Aprile 2023

Sondaggio sui Data Steward in Italia

SONDAGGIO

Per l'attivazione della rete professionale dei data steward

Il sondaggio ha lo scopo di identificare la presenza dei Data Steward all'interno di Università e Istituti di Ricerca in Italia.

Il Data Steward è un professionista specializzato nella gestione e promozione della scienza aperta. Agisce come intermediario tra i



HOME NAVIGA IL SITO EVENTI CATALOGO ENGLISH CERCA



HOME CHI SIAMO ATTIVITÀ DOCUMENTI

Supporting **oeosc**



THE ITALIAN EDUCATION & RESEARCH NETWORK

Data steward, riepilogo e riflessioni dal Kick-off meeting della Comunità Italiana

06 dicembre 2023

1° Meeting in presenza del Competence Centre ICDI e Kick-off della Data Steward Community Italiana



EOSC: il potenziale

...IL VALORE DEGLI OPEN DATA:
ATTIVARE IL POTENZIALE DEI DATI DELLA
RICERCA PER ACCELERARE PROGRESSO E
INNOVAZIONE

eosc EOSC Strategy – Status Current Thinking

What

EOSC is a web of FAIR data and related services for research
Research data that is easy to find, access, interoperate and reuse (FAIR)
Trusted and sustainable research outputs are available within and across scientific disciplines

Why

Unlock the full potential of research data to accelerate discoveries and innovation

How

Access and interoperability of research data and results

- Define ownership, authorship and responsibility of data and research outputs
- Ensure long-term preservation of data throughout its lifecycle
- Enable the creation of standards for all research domains
- Make data machine-actionable
- Enable new scientific discovery methods and science disciplines
- Train researchers on adopting FAIR principles as an integral part in their activity

A sustainable coordinated infrastructure

- Establish and maintain a coordinated federated reference architecture
- Implement an operational infrastructure framework that is long term sustainable
- Ensure high quality of data and services
- Ensure secure access to data and services
- Define clear standards for API and interoperability of data and services
- Apply user friendly practices
- Inspire EOSC ambassadors to assist in on-boarding of researchers

Inspired people and robust governance

- Communicate an inspiring EOSC vision and strategy
- Implement an unambiguous and clearly mandated governance structure
- Establish a framework to engage human capital in institutions, countries and scientific communities
- Enable disciplinary and cross-disciplinary transnational research to find new insights from existing and new research data and outputs



...un passo indietro...

[il fondamento]

Information Guide: Introduction to Ownership of Rights in Research Data. CREATE, University of Glasgow, 2018

2018

Barrow, S. , Margoni, T.  and McCutcheon, V.  (2018) Information Guide: Introduction to Ownership of Rights in Research Data. CREATE, University of Glasgow, 2018. Documentation. University of Glasgow.



Guides for Researchers

How do I know if my research data is protected?

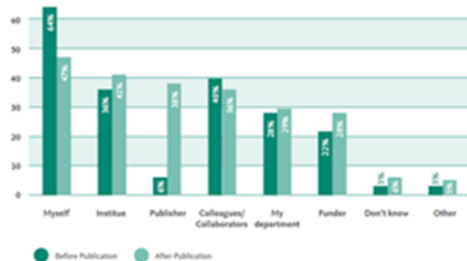
Learn more about what is research data and their protection by intellectual property rights

OpenAIRE

I DATI GREZZI
NON SONO «MIEI»
NON ESISTE COPYRIGHT
PERCHÉ NON SONO CREATIVI

This time though it happened. What it was: 64% of researchers believe they own the data they generated for their research.

Figure 3. Research data ownership before and after publication (%; n=1162)



The result comes from a **solid piece of academic research** based on equally solid (open) data. The study and the report 'Open Data - the Researcher Perspective' were done by **CWTS / Leiden** and **Elsevier**. Credit giving, check.

Of course, the study reports other equally surprising results



Wainer Lusoli

@w_lusoli

Following

repeat with me: [#researchdata](#) is NOT mine. I was paid to get it, I'll get a [#nobel](#) 4 it, but it's NOT mine [linkedin.com/pulse/repeat-m ...](https://www.linkedin.com/pulse/repeat-m...) [#opendata](#)

Traduci dalla lingua originale: inglese



Repeat with me: research data is not mine

Seldom do I see something that truly shakes me at work. You know, work is work, I am no neurosurgeon, no médecin sans frontières nor am I a social

[linkedin.com](https://www.linkedin.com)

11:18 - 12 apr 2017

14 Retweet 18 Mi piace



Lusoli, Apr.2017

R = Reusable – Aspetti legali

1. THE PROTECTION OF DATA, DATA SETS AND DATABASES

European Union (EU) law defines “databases”, but not data sets or, at least for copyright purposes, data. Databases that meet the legal definition^① can be protected by copyright if they are original. Data sets, if they correspond to the definition of database, are protected by copyright otherwise not. Data as such are normally excluded from copyright protection [2,3]. It is important to understand that copyright protects original expressions in the “literary and artistic” domain^②, an expression that has historically included works such as books, musical works, choreographies, cinematographic works, drawings, etc [4]. Ideas, procedures, methods of operation or mathematical concepts as such, news of the day and miscellaneous facts are excluded from copyright protection [4,5,6].



MIT Press Direct



2020

Data Intelligence

Volume 2, Issue 1-2

Winter-Spring 2020



< Previous Article Next Article >

Article Contents

Abstract

1. THE PROTECTION OF DATA, DATA SETS AND DATABASES

2. SUITABLE OPTIONS FOR LICENSING DATA AND DATABASE RIGHTS

January 01 2020

Licensing FAIR Data for Reuse

Ignasi Labastida, Thomas Margoni

> Author and Article Information

Data Intelligence (2020) 2 (1-2): 199–207.

https://doi.org/10.1162/dint_a_00042



Cite



PDF



Permissions



Share

Abstract

The last letter of the FAIR acronym stands for Reusability. Data and metadata should be made available with a clear and accessible usage license. But, what are the choices? How can researchers share data and allow reusability? Are all the licenses available for sharing content suitable for data? Data can be covered by different layers of copyright protection making the relationship between data and copyright particularly complex. Some research



Le basi

5 MODI PER PENSARE I DATI:

- COME SONO RACCOLTI (ESPERIMENTI, SIMULAZIONI...)
- COME SI PRESENTANO (TESTI, QUESTIONARI, VIDEO...)
- IL LORO FORMATO ELETTRONICO (.TXT, .MKV...)
- IL LORO VOLUME (BIG DATA...)
- IN CHE FASE SONO DEL CICLO (RAW DATA...)

▣ The way the data is collected.

- ▣ By experimenting, simulations, observations, derived data, reference data.

▣ The data forms.

- ▣ For example text documents, spreadsheets, lab journals, logs, questionnaires, software code, transcripts, code books, audio and video recordings, photos, samples, slides, artefacts, models, scripts, databases, metadata, etc.

▣ The formats for electronic storage of the research data.

▣ The size (volume) of the data files.

▣ The *research lifecycle* phase the data is in.

RICHIEDONO
STRUMENTI E
TRATTAMENTI
DIVERSI

Il ciclo di vita dei dati



PIANIFICARE LA
GESTIONE DEI DATI
IN OGNI PASSO DEL
CICLO DI VITA È
CRUCIALE

Il ciclo di vita dei dati

The research data lifecycle

Data Management Planning



Active data

- Frequent access
- Short-term fast storage
- Shared with collaborators
- Not finalised

“Cold” data

- Seldom access
- Long-term slow storage
- Open
- Finalised (but can be updated if needed)

11 5%

nature

Feb. 25, 2020

Subscribe

WORLD VIEW · 25 FEBRUARY 2020

Invest 5% of research funds in ensuring data are reusable



It is irresponsible to support research but not data stewardship, says Barend Mons.

Barend Mons

Bunk. First, taking care of data is an ethical duty, and should be part of good research practice. Second, if data are treated properly, researchers will have significantly more time to do research. Consider the losses incurred under the current system. Students in PhD programmes spend up to 80% of their time on 'data munging', fixing formatting and minor mistakes to make data suitable for analysis – wasting time and talent. With 400 such students, that would amount to a monetary waste equivalent to the salaries of 200 full-time employees, at minimum. So, hiring 20 professional data stewards to cut time lost to data wrangling would boost effective research capacity.

Many top universities are starting to see that the costs of not sharing data are significant and greater than the associated risks. Data stewardship offers excellent returns on investment.

Funders hold the stick: they should disburse no further funding without a properly reviewed and budgeted data-stewardship plan. The carrot is that FAIR data allow much more effective artificial intelligence (FAIR can also mean 'fully AI ready'), which will open up unprecedented research opportunities and increase reproducibility.

I tell research institutions that, on average, 5% of overall research costs should go towards data stewardship. With €300 billion (US\$325 billion) of public money spent on research in the European Union, we should expect to spend €15 billion on data stewardship. Scientists, especially more experienced ones, are often upset when I say this. They see it as 5% less funding for research.

- PRENDERSI CURA DEI DATI È ETICO
- ASSUMERE DATA STEWARDS FA RISPARMIARE TEMPO
 - FAIR=FULLY ARTIFICIAL INTELLIGENCE READY

Costi

COSTI DEL **NON** AVERE
DATI FAIR

Cost of not having FAIR
research data

Cost-Benefit analysis for FAIR research data

10,2 bn DIRETTI
16 bn INDIRETTI
26,2 bn TOTALI

Following this approach, we found that the annual cost of not having FAIR research data costs the European economy at least €10.2bn every year. In addition, we also listed a number of consequences from not having FAIR which could not be reliably estimated, such as an impact on research quality, economic turnover, or machine readability of research data. By drawing a rough parallel with the European open data economy, we concluded that these unquantified elements could account for another €16bn annually on top of what we estimated. These results relied on a combination of desk research, interviews with the subject matter experts and our most conservative assumptions.

CI SONO COSTI PER GESTIRE, RENDERE FAIR E CONSERVARE I DATI...
MA PENSIAMO

- A QUANTO COSTEREBBE NON CONSERVARLI E NON GESTIRLI
- A QUANTO TEMPO PERDETE PER «PULIRLI» PRIMA DI POTERLI USARE (79% DEL TEMPO PER PREPARARLI]

Data management costing tool and checklist

CHECKLIST DEGLI ASPETTI DA CONSIDERARE PER VALUTARE I COSTI

The costing tool

Activity	Comments and suggestions	✓	Cost
Data description <ul style="list-style-type: none"> Are data in a spreadsheet or database clearly marked with variable and value labels, code descriptions, missing value descriptions, etc? Are labels consistent? Do textual data like interview transcripts need description of context, e.g., included as a heading page? 	<ul style="list-style-type: none"> If data descriptions are implemented as part of data creation, data input or data transcription - low or no additional cost. If needed to be added afterwards - higher cost. Codebooks for datasets can often be easily exported from software packages. 		
Data cleaning <ul style="list-style-type: none"> Do quantitative data need to be cleaned, checked, or verified before sharing, e.g., check validity of codes used, check for anomalous values? 	<ul style="list-style-type: none"> If carried out as part of data entry and preparation before data analysis - low or no additional cost. If needed afterwards - higher cost. 		

[sui costi]

How to use the costing tool

Step 1: Check

Check the data management activities in the table and tick those that may apply to your proposed research.

Step 2: Estimate

For each selected activity, estimate the additional time and/or other resources needed and cost this, e.g., people's time or physical resources needed such as hardware or software. Find out which resources are available to you from your institution. Consider whether you need a dedicated data manager.

Step 3: Implement

Add these data management costs to your research application. Coordinate resourcing and costing with your institution, research office, and institutional IT services.

Step 4: Plan

Plan the data management activities in advance to avoid them competing with the need to focus on research excellence.

Formatting and organising <ul style="list-style-type: none"> Are your data files, spreadsheets, interview transcripts, records, etc. all in a uniform format or style? Are files, records and items in the collection clearly named with unique file names and well organised? 	<ul style="list-style-type: none"> If planned beforehand by developing templates and data entry forms for individual data files (transcripts, spreadsheets, databases) and by constructing clear file structures - low or no additional cost. If needed afterwards - higher cost. Free software exists for batch file renaming to harmonise file names. 		
Transcription <ul style="list-style-type: none"> Will you transcribe qualitative data (e.g., recorded interviews or focus group sessions) as part of your research; or will you need to do this specifically so data can be more easily shared and reused? Is full or partial transcription needed? Is translation needed? Will you need to develop a... 	<ul style="list-style-type: none"> If transcription is part of research practice - very low or no additional cost. If transcription not planned as part of research practice - potentially high cost. Is additional hardware /software needed? Consider cost of time needed for developing procedures, templates, and guidance for transcribers. 		

1. GESTIRE I DATI



Jisc

How and why you should manage your research data: a guide for researchers

An introduction to engaging with research data management processes.

[JISC Guide](#)



D | C | C

because good research needs good data

Perché gestire i dati

RISPARMIO DI TEMPO
MAGGIORE EFFICIENZA
CONSERVA E PROTEGGE
I DATI SONO UN «PRODOTTO»
TRASPARENZA/INTEGRITÀ



- **Save Time** – By spending a little bit of up-front time and planning and organising the data you produce you will save time and resources in the long run.
- **Increase your efficiency** – If you document your data properly whenever you or someone else comes to it they will be able to understand it quickly and without difficulty. Thus saving time and increasing efficiency.
- **Preserve and protect your data** – It is relatively easy to produce data that will be useful only the once and for a very specific purpose. Learn how to ensure that the data can be useful again and again, and how to make sure that it is never lost.
- **Data is an output in its own right** – that's right; data itself is increasingly being seen as an important output of research. If shared, it can better enable researchers. The REF (Research Excellence Framework) now takes note of it.
- **Meet grant requirements** – Many funding bodies now require that researchers archive data as well as the resulting publications as part of their project. Good data management will make this easy rather than a last minute chore.
- **Open Access** – In the UK government policy has moved to an open access framework. Producing and making available data is a vital part of this process. Journals are increasingly making room for data alongside articles, for example.
- **Transparency/research integrity** – If required you have all the documents and materials easily available making your research more transparent if questioned.

[Why data management](#)



An overview

Research Data Management*

*Standard/best practices for accurate data/code collection, processing, documentation, analysis, storage & preservation as a prerequisite for open science (FAIR ≠ Open).

- What decisions do researchers make to achieve 'FAIR' data management?



Decision making procedures in data management and data stewardship for Open Science

2022

Connie Clare, PhD



UNA BUONA GESTIONE DEI DATI È UN PREREQUISITO PER FAIR/OPEN ...E HA ENORMI VANTAGGI



The importance & benefits of RDM



Efficiency & avoids duplication of efforts (saves time, money & resources)



Transparency with internal/external collaborators



Easier data publication for long-term preservation 'prepare to share'



Reproducibility & verification → reusability



Accountability for data quality → integrity & confidence



Increased impact, greater visibility & citations



Compliance with legislation (GDPR, legal & ethical)



Compliance with policy (Institutional, Journal & Funder)

FAIR data management



Decision making procedures in data management and data stewardship for Open Science

2022

Connie Clare, PhD



DATI «FAIR BY DESIGN» – LA GESTIONE DEI DATI DEVE ESSERE FAIR IN PARTENZA



What decisions do researchers make to achieve 'FAIR' data management?

Who is responsible for data management?

How will new data be collected or produced and/or how will existing data be reused?

What data types, formats, and volumes will be collected or produced?
Structured data, formats: JSON, XML, CSV

What metadata and documentation will accompany data?

What resources will be dedicated to data management and ensuring that data will be FAIR?

How, when and which will data be shared? Are there possible restrictions to data sharing or embargo reasons?

Are there disciplinary standards and vocabularies that should be used?

What methods or software tools will be needed to access and use the data?

Will the application of a unique persistent identifier (e.g., DOI) be assigned to the dataset?



What data quality control measures will be used?

If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?

Which license will be suitable to specify data modification, redistribution and reuse?

How will data for preservation be selected? Where will data be preserved long-term?

How will data and metadata be stored and backed up during the research process?

Is an ethical review (HREC, ERB) required?

Is informed consent required?

Does data need to be anonymised or pseudonymised?

How will other legal issues, such as IPR and ownership, be managed? What legislation is applicable?

How will possible ethical issues be taken into account, and codes of conduct followed?

Data decision tree

STRUMENTO UTILISSIMO PER UN PRIMO PAPPROCCIO AL DATA MANAGEMENT [PLAN]

Caldoni, Giulia, Gualandi, Bianca, & Marino, Mario. (2022). Research Data Management Decision Tree

Legend:

DATA MANAGEMENT

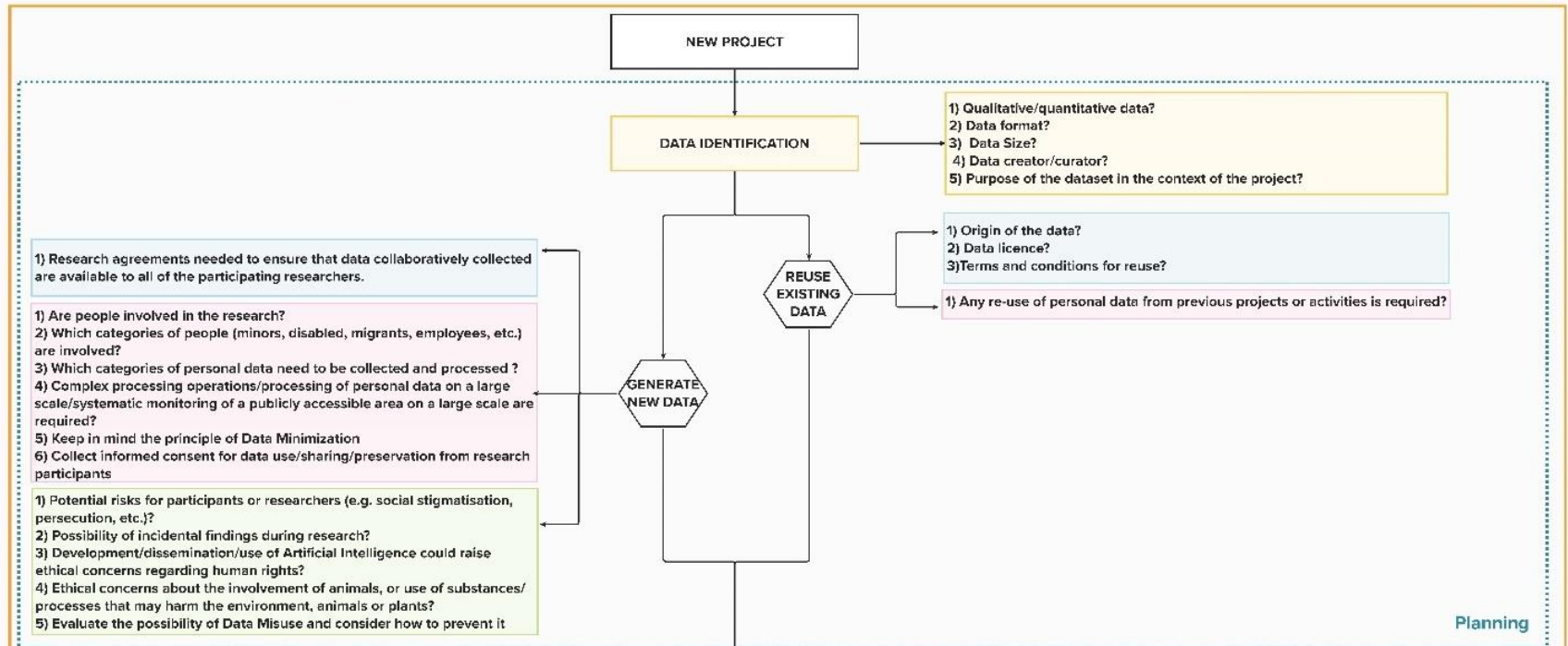
INTELLECTUAL PROPERTY RIGHTS

PRIVACY

ETHICS

DECISION TREE FOR DATA MANAGEMENT

Data management



[Houston, we have



Support is not making its way to those who need it

Almost three-quarters of respondents have never received support with planning, managing or sharing research data.

With the global increase in policies and mandates to share data openly, who researchers are approaching for support becomes a pertinent question.

If respondents stated that they were aware of the concept of a data management plan, they were then asked if they have access to support from specialist data managers and we saw over 50% of our respondents state that they do have access to specialist research data managers in their research setting, but who else has been providing support?

Almost three quarters of respondents **had never received support** with planning, managing or sharing research data. When respondents were asked if they had ever received support with managing or making their data openly available, only 23% said they had. Of that 23%, 61% received support from informal internal sources such as colleagues or supervisors. Two other sources of support ranked highly with our respondents;

Do you have access to support from specialist data managers?



MANCA IL SUPPORTO (POSTO CHE NON SIETE «PERSONALE DI SUPPORTO»)

Graph showing the responses to the question 'Do you have access to support from specialist data managers?'

Gestione = sicurezza

Sweeney L, Crosas M, Bar-Sinai M. Sharing Sensitive Data with Confidence: The DataTags System. *Technology Science*. 2015;101601. October 16, 2015. <http://techscience.org/a/2015101601>



Technology Science

Sharing Sensitive Data with Confidence: The Datatags System 2015

Latanya Sweeney, Mercè Crosas, and Michael Bar-Sinai

Tag Type	Description	Security Features	Access Credentials
Blue	Public	Clear storage, Clear transmit	Open
Green	Controlled public	Clear storage, Clear transmit	Email- or OAuth Verified Registration
Yellow	Accountable	Clear storage, Encrypted transmit	Password, Registered, Approval, Click-through DUA
Orange	More accountable	Encrypted storage, Encrypted transmit	Password, Registered, Approval, Signed DUA
Red	Fully accountable	Encrypted storage, Encrypted transmit	Two-factor authentication, Approval, Signed DUA
Crimson	Maximally restricted	Multi-encrypted storage, Encrypted transmit	Two-factor authentication, Approval, Signed DUA

[DMP]

4 pilastri

andS AUSTRALIAN NATIONAL DATA SERVICE

About ANDS | Events | Contact Us

Australia data service

About us | News and Events | Partners and Communities | Working with data | Online Services | Guides and resources

Findable	Persistent Identifiers (PIDs)	Rich metadata	Indexed data repositories	PIDs in metadata
Accessible	Standard communications protocol	Open, free protocol	Authentication, where necessary	Metadata is always available

FAIR data training

If you run workshops on FAIR data, or include FAIR in training that you are already running check out these ideas and resources.

Search the ANDS Site

Enter Keywords



Digital Curation Center UK

Because good research needs good data

- About
- News
- Events
- Services
- Guidance
- Research



The Digital Curation Centre in collaboration with Research Data Netherlands have developed an online course on **Delivering Research Data Management Services (DRDMS)**.

After two enrolmen the 5th of



Dutch data service

- HOME
- FOR RESEARCHERS
- FOR DATA PROFESSIONALS
- FOR PARTNERS
- PROJECTS
- ABOUT DANS
- NEWS AND EVENTS

Welcome at DANS: the Netherlands institute for permanent access to digital research resources.



Dutch consortium

- Home
- Experts/Mailing List
- Task groups
- RDM Advice & Tips
- RDM in NL
- About us/Meetings

The data support collective



DANS for researchers

DANS for data professionals

DANS for partners

Main Points for Good Data Management

Data acquisition

- Check the type, source of the data and how to gather/collect it
 - Data types (to help define sensitivity of data)
 - Data format (to help define the tools and software)
 - Data size (to help define storage and infrastructure)
- Check the ownership of the collected and processed data
 - Check with the data source about conditions (e.g. licence)
 - Check the need to make a data process on the ownership / access control
 - Are there (own) institutional policies that apply?
 - Can the data be shared with other parties?
- Confidentiality of the data (if applicable):
 - Register crucial information regarding data
 - Ensure security of confidential data (personal data, or data that would harm society with disclosure)
 - Ensure compliance with General Data Protection Regulation (GDPR) / Verordening gegevensbescherming when applicable
 - Ensure there are procedures in place to consult of a privacy advisor/data protection officer

Data collection

- Establish a workflow for data collection
 - How will the data be collected?
 - Who has access to which data in short / long term?
 - What resources are needed for data analysis?
 - How will the data be exchanged / transferred among relevant stakeholders?
- Storage arrangement
 - Check available storage capacity and backup strategy

Data storing / backup

- Create a clear folder structure and consistent file naming convention
- Make a backup strategy where data is stored at least two different physical locations and preferably automatically backed up
- Access control to confidential data
- Apply encryption at disk or folder level if needed
- Create a consistent and standard versioning of the data files
- Determine the minimal documentation of the data that is required to find it, understand it and use it

Data sharing

- Create proper data sharing procedures
 - Consider agreements established in the Data acquisition phase, and evaluate/assess data sharing with other parties
 - Be aware of the permission and consequence of sharing confidential data
- Copyright / Licensing
 - How should others use the data
 - Who should be attributed for creating/gathering the data

Organizational Implications

In addition to the above mentioned actions, there are also a few things to consider to make data management a standard practice in daily operations.

PER FARSI LE DOMANDE GIUSTE

Serve formazione?

[DMP]

The screenshot shows the CESSDA website navigation menu with options: Tools, Training, Strategy & Expertise, Development & Impact, and About. The main content area is titled "Data Management Expert Guide (DMEG)". It features a purple icon of a document with a pencil, followed by text describing the guide's purpose for social scientists. A table of contents is visible, listing chapters from 1. Plan to 8. Contributors. A sidebar on the left lists training resources, with "Data Management Expert Guide" highlighted.

Are you here for the first time?

Take the quiz below and find out which chapters of DMEG will be most useful for you.



Target audience and mission

This guide is written for social science researchers who are in an early stage of practising research data management. With this guide, CESSDA wants to contribute to professionalism in data management and increase the value of research data.

Data Management Expert Guide
<https://dmeg.cessda.eu/>



Data Management Expert Guide (DMEG)

The DMEG is designed by European experts to help social science researchers make their research data Findable, Accessible, Interoperable and Reusable (FAIR).

You will be guided by different European experts who are - on a daily basis - busy ensuring long-term access to valuable social science datasets, available for discovery and reuse at one of the [CESSDA social science data archives](#).

You can [download](#) the full DMEG for your personal study offline (DOI: [10.5281/zenodo.3820473](https://doi.org/10.5281/zenodo.3820473)). PDFs for every [single chapter](#) are also available for being printed as handouts for training.

See also the pilot [interactive game version](#) of the guide!

A search bar with the text "What is the CESSDA Data Management Expert Guide &...". To the right of the search bar are icons for "Guarda più..." and "Condividi".

[DMP]

Con un supporto pratico

ALLA FINE D OGNI MODULO
TROVATE «ADAPT YOUR
DMP» PER APPLICARE I
CONCETTI CHE AVETE
APPENA IMPARATO

Adapt your DMP: part 6

This is the sixth 'Adapt your DMP' section in this tour guide. To adapt your DMP, consider the following elements and corresponding questions:



Deposit your data

- Will the data you produce and/or used in the project be useable by third parties, in particular after the end of the project?
- Which data and associated metadata, documentation and code will be deposited?
- What methods or software tools are needed to access the data?
- Is documentation about the software needed to access the data included?
- Is it possible to include the relevant software (e.g. in open source code)?
- What data quality assurance processes will you apply?

[CESSDA Guide](#)



Data Management Expert Guide Train-the-Trainers

Archived / Training Resources / Train-the-Trainers

Train-the-Trainers Package

General information

Exercises

Name	Description
TTT_EX_GESIS_Re-IDQual_v1.0	Example Exercise by GESIS on re-identification of qualitative data.
TTT_EX_GESIS_Re-IDQual_Ans_v1.0	Sample solutions for TTT_EX_GESIS_Re-IDQual.
TTT_EX_GESIS_Re-IDQuant_v1.0	Example Exercise by GESIS on re-identification of quantitative data.
TTT_EX_GESIS_Re-IDQuant_Data_v1.0	Data belonging to TTT_EX_GESIS_Re-IDQuant.

NELLA SEZIONE «TRAIN THE
TRAINERS» LEZIONI ED
ESERCIZI

Formazione



Essentials4data

Essentials 4 Data Support

ABOUT THE COURSE | START THE COURSE | LOGIN

- I - A bird's-eye view >
- II - Planning phase >
- III - Research phase >
- IV - Harvest phase >
- V - Legislation and policy >
- VI - Data support >
- Closing remarks

Welcome

The Turing way

The Turing Way is an open source community-driven guide to reproducible, ethical, inclusive and collaborative data science.

Our goal is to provide all the information that data scientists in academia, industry, government and the third sector need at the start of their projects to ensure that they are easy to reproduce and reuse at the end.

The book started as a guide to reproducibility, covering version control, testing, and continuous integration. However, technical skills are just one aspect of making data science research "open for all".

In February 2020, The Turing Way expanded to a series of books covering reproducible research, project design, communication, collaboration, and ethical research.



Engaging Researchers with Data Management The Cookbook

CONNIE CLARE, MARIA CRUZ, ELI PAPADOPOULOU, JAMES SAVAGE, MARTA TEPEREK, YAN WANG, IZA WITKOWSKA, AND JOANNE YEOMANS

CASE STUDIES

1. RESEARCH DATA MANAGEMENT POLICY: THE HOLY GRAIL OF DATA MANAGEMENT SUPPORT?
 - 1.1. Are You a Research Data Superhero? One Person Making a Big Difference at Makerere University
 - Changing the Mindset of Researchers
 - Never Overlook an Opportunity to Speak about RDM
 - How Fast Do Things Change?
 - Additional Resources
 - 1.2. Does a Policy Solve Everything? Policy as a Driver for Engagement at Leiden University
 - A Crown Is Merely a Hat that Lets the Rain In
 - Leiden's Use of its RDM Policy to Prompt Discussion
 - 'One Size' Does Not Fit All
 - Why Does this Kind of Engagement Take Time?
 - Continuing the Engagement with a Matrix of Support Services

2020

Home > Learning Hub > Research data management

Research data management

- [Data lifecycle](#)
- [Plan to share](#)
- [Rights in data](#)
- [Collaborative research](#)

<https://ukdataservice.ac.uk/learning-hub/research-data-management/>

The importance of managing and sharing data

In the digital age, the generation of research data has grown exponentially, and data are nowadays very easily stored, kept and exchanged around the world. Digital infrastructures and the internet facilitate both the creation of ever larger amounts of research data, as well

- [Data protection](#)
- [Ethical issues](#)
- [Storing data](#)
- [Formatting data](#)
- [Anonymising data](#)
- [Documenting data](#)

Data management planning overview

ESRC data management plan and policy

Roles and responsibilities

Checklist

Backup

Checksums

Disposal

Encryption



...e le scienze sociali?



Tools ▾

Training ▾

Strategy & Expertise ▾

Development & Impact ▾

About ▾



<https://www.cessda.eu/>

CESSDA consolidates
and strengthens European
social science research.

Find Research Data

The [CESSDA Data Catalogue](#) contains tens of thousands of European studies and data collections held by our Service Providers.



[CESSDA Data Catalogue](#)

Learn Data Management

The [Data Management Expert Guide \(DMEG\)](#) is a comprehensive guide to Research Data Management and FAIR data principles.




[Data Management Expert Guide](#)

...e le scienze umane?

DARIAH-CAMPUS Resources Topics Sources Course Registry About May 2019

DARIAH Pathfinder to Data Management Best Practices in the Humanities

Written by Erzsébet Tóth-Czifra May 03 2019 Source: DARIAH Pathfinders, DARIAH Topics: Data management

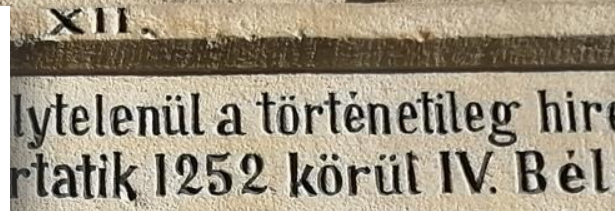


1. Why research data management?

Systematically planning how you will collect, document, organize, manage, share and preserve your data has many benefits. It helps to build a common framework of understanding with your

TABLE OF CONTENTS

1. Why research data management?
2. Data in the Humanities
3. The devil is in the context: a processual view on data curation
4. Sharing your data
 - 4.1. Cite to be cited!
 - 4.2. Be aware of your licensing options
 - 4.3. A case study: different levels of being an open scholar
5. A recipe for your research project: the Data Management Plan
6. Data in publications and data as publications
 - 6.1. The networked publication: interlinking the underlying data with your
 - 6.2. Data journals in humanities



10. THE RISK OF LOSING THE THICK DESCRIPTION: DATA MANAGEMENT CHALLENGES FACED BY THE ARTS AND HUMANITIES IN THE EVOLVING FAIR DATA ECOSYSTEM

Erzsébet Tóth-Czifra

Realising the Promises of FAIR within Discipline-Specific Scholarly Practices

A Cultural Knowledge Iceberg, Submerged in an Analogue World

Legal Problems that Are Not Solely Legal Problems

The Risk of Losing the Thick Description upon the Remediation of Cultural Heritage

The Scholarly Data Continuum

Data in Arts and Humanities — Still a Dirty Word?

The Critical Mass Challenge and the Social Life of Data

The Risk of Losing the *Thick Description* — Again

Conclusion	235
Bibliography	237
	239
	242
	247
	250
	251
	255
	258
	263

PARTHENOS HOME TRAINING MODULES FOR TRAINERS FOR LEARNERS ABOUT PARTHENOS TRAINING

MANAGE, IMPROVE AND OPEN UP YOUR RESEARCH AND DATA

SHARE

About the module

This module will look at emerging trends and best practice in data management, quality assessment and IPR issues

We will look at policies regarding data management and their implementation, particularly in the framework of a Research Infrastructure

Learning Outcomes

By the end of this module, you should be able to:

BROWSE

Introduction to Research Infrastructures

Management Challenges in Research Infrastructures

Introduction to Collaboration in Research Infrastructures

Manage, Improve and Open up your Research and Data

Introduction to Research Data

Data e scienze umane

1. Methodological frameworks and epistemic cultures.

We call for the use of a common methodological terminology in research at the intersection between humanities and data science, and for a wider use of shared research protocols across these domains. We recommend that authors make the methodological framework that they are using explicit in their publications, and we call for inclusive research practices to be fostered across research projects.

2. Best practices in the use and evaluation of computational tools.

We encourage practices that ensure transparency and openness in research, and training programmes that help to choose the most suitable computational tools and processes in humanities research. We also call for computational tools to be evaluated in a dialogue between data scientists and digital humanists.

3. Reproducible and open research.

We promote transparent and reproducible research in the humanities workflows, computational environments, methods, and documentation. Academic institutions should put in place further incentives for humans to publish the digital resources, code, workflows and pipelines they create outputs, e.g. in the form of publications in data journals.

2020

The Alan Turing Institute

Humanities and data science
special interest group

The challenges and prospects of the intersection of humanities and data science:

A white paper from
The Alan Turing Institute

– **History and critique of data science.** This area analyses the characteristics of data science work, sometimes with a focus on the historicity of datasets, and tackles ethical and methodological questions aimed at improving current practices, for example, on issues such as diversity (D'Ignazio and Klein, 2020) and privacy. It also problematizes the very definition of data, considering their complexities, their inherent biases, their contextual and historical natures, in a critical and nuanced way (e.g. Drucker, 2011). Exemplary research in this area includes, among many others, MacKenzie (2017)'s study of the interface between machine learning and critical thought, Kaltenbrunner (2014; 2015)'s studies of infrastructure as a relational and emergent phenomenon that shapes data-driven humanities research and researchers, and the questions they can ask, in complex ways, and Noble (2018)'s study of how white patriarchy and algorithmic bias has resulted in the misrepresentation of women of colour and minorities in search engine results.

– **Algorithmic creativity and cultural innovation in the arts and humanities.** This area focuses on computational creativity, aiming to perform creative tasks with the aid of machines and to explore the plasticity of digital forms for delivering new radical ways of representation and mediation of the arts and humanities. This is the focus of various initiatives such as the Turing AI & Arts group⁶¹ and the Creative Informatics programme in Edinburgh⁶². Individual artists are also increasingly including computational methods in their practice⁶³. Examples of projects undertaken in collaboration with academic researchers, creative industries and the GLAM sector include the King's Digital Lab Digital Ghost Hunt⁶⁴ experience and the AI and Storytelling project⁶⁵.

...e le scienze umane

OPENMETHODS
HIGHLIGHTING DIGITAL HUMANITIES METHODS AND TOOLS

OpenMethods

HOME ABOUT WHO WE ARE JOIN US SUBMIT A CONTENT RSS FEEDS LOG IN

Figure 1: An example of XML data from KYC3 Catalogue. Each snippet is translated to English!

ANALYSIS
The Language Interpretability Tool: Extensible, Interactive Visualizations and Analysis for NLP Models
APRIL 29, 2021 - BY ERZSEBET TÓTH-CZIFRA

ANALYSIS
Cultural Ontologies: the ArCo Knowledge Graph.
MARCH 11, 2021 - BY MARINELLA TESTORI
Introduction: Standing for 'Architecture of Knowledge', ArCo is an open set of resources developed and managed by some Italian institutions, like the MiBAC (Minister

INTERESTED IN BLOGGING ABOUT YOUR RESEARCH? THE DIGITAL HUMANITIES TOOLS AND METHODS BLOG IS FOR YOU!

hypotheses

IN COOPERATION WITH
 DARIAH-EU

EGI-ACE | SERVICES | FEDERATION | USE CASES | BUSINESS

EGI / USE CASES / SCIENTIFIC APPLICATIONS AND TOOLS / DARIAH GATEWAY

DARIAH Gateway

DARIAH Gateway

Cloud applications and services for Arts & Humanities researchers

The **DARIAH Gateway** is a platform that provides access to various digital applications and services for the Arts & Humanities researchers.

The applications made available via the DARIAH Gateway are:

- **Simple Semantic Search Engine (SSE)**: a semantic search engine which allows researchers to search for content in more than 100 languages within the Sci-Gala **e-Infrastructure Knowledge Base**, one of the largest existing databases.
- **Parallel Semantic Search Engine (PSSE)**: a parallelised version of SSE that enables simultaneously search across multiple platforms.
- **DBO@Cloud**: a cloud-based repository made of a 100-years old collection of Bavarian dialects. The datasets are provided by the **Austrian Academy of Science**.

The services made available via the DARIAH Gateway are:

- **Cloud Access**: single-job applications and parameter-sweep applications can be run on the DARIAH virtual organisation clouds without porting efforts.
- **Workflow Development**: workflow applications can be developed and run on all the resources of the DARIAH virtual organisation.
- **File transfer**: enables transferring data from, to and between storage services providing HTTP, HTTPS, SFTP, GSIFTP, SRM, IRODS and S3 protocols.

Ariadne
Home Community Activities & resources Portal Training & TNA News & even

ARIADNEplus

a data infrastructure serving the archaeological community worldwide

[LEARN MORE ABOUT THE PROJECT](#)

Integrate and effectively serve a research community that studies the past to better understand the present with the tools and the methodology of the future, in the service of culture and society

Data management ABC

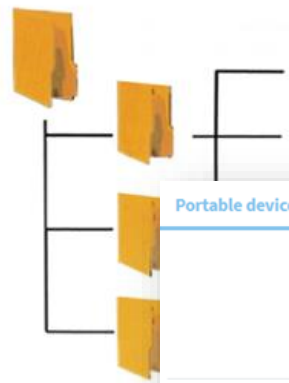
COME NOMINARE I FILE

RECORDS MANAGEMENT

Folder structure

Structuring your data files in folders is important for making it easier to locate and organise files and versions. A proper folder structure is especially needed when collaborating with others.

CESSDA training



STRUTTURA DEI FOLDER

collection and data processing procedures.

Consider the best hierarchy of your files and decide whether a deep or shallow hierarchy is preferable. If you have several independent data collections, it is advisable to create a separate data folder for each. For inspiration, have a look at the examples in the accordion below.

Portable devices Cloud storage Local storage Networked drives



Laptops, tablets, external hard-drives, flash drives and Compact Discs

Advantages Disadvantages/Risks Precautions for (sensitive) personal data

BACK UP E STORAGE

Home > Records Management > Guidance > Managing records

Naming conventions

Make finding electronic records easier.

git --distributed-even-if-your-workflow-isnt Git

Git is a free and open source distributed version control system designed to handle everything from small to very large projects with speed and efficiency.

Git is easy to learn and has a powerful CLI and Clear staging area.

VERSIONING

What are personal data?

FOSTER data protection

Click the plus sign to expand the text box

- + What are personal data?
- + Protecting personal data
- + Legal requirements - EU General Data Protection Regulation (GDPR)
- + Legal requirements - GDPR research exemptions

PROTEZIONE DEI DATI

About
The advantages of Git compared to other source control systems.

Documentation
Command reference pages, Pro Git book content, videos and other material.

Downloads
GUI clients and binary releases for all major platforms.

Community
Get involved! Bug reporting, mailing list, chat, development and more.

Latest source Release
2.31.1
Release Notes (2021-03-26)
Download 2.31.1 for Windows

Data management ABC / essere riproducibili

Reproducible coding

DO's

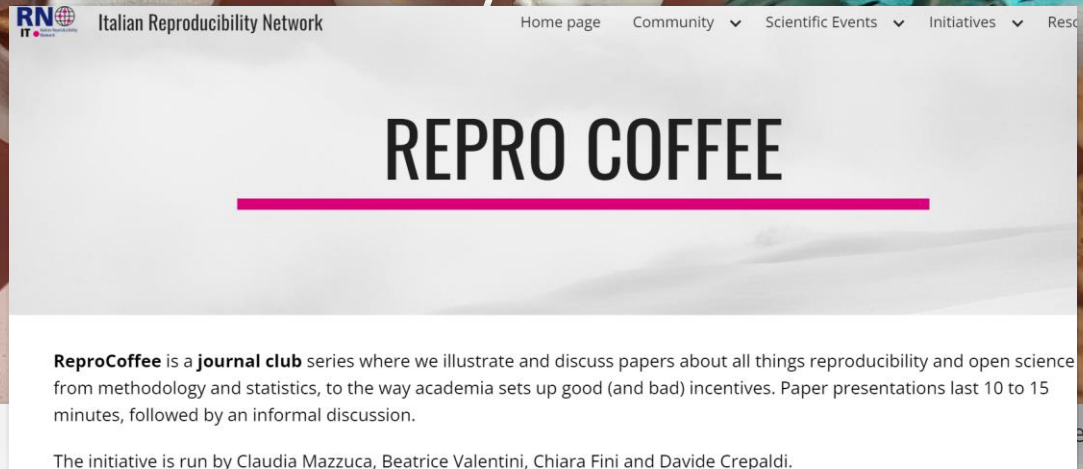
- Comment your code.
- Use versioning systems (such as Git, Mercurial, Subversion, etc.) to keep track of versions, collaborate, and perform automatic backups.
- Use code notebooks (e.g. Jupyter notebooks) for literate programming.
- Consider using emulation and virtualisation tools, such as Docker, to help recreating the original computing environment for reproducibility.
- Use workflow tools (Snakemake, Nextflow, Galaxy, DataJoint) to make your computational methods portable, maintainable, reproducible and shareable, especially if analytical pipelines can involve up to hundreds of successive steps (e.g. bioinformatics).
- Use software licenses (e.g., GNU GPL license)

DONT's



Data management ABC / essere riproducibili

SEMINARI BREVI SU
DIVERSI ASPETTI DI
RIPRODUCIBILITÀ



28 November 2023, [Better Accuracy for Better Science ... Through Random Conclusions](#). by Davis-Storber et al., 2023.

24 October 2023, [Reproducible brain-wide association studies require thousands of individuals](#). by Marek et al., 2022

27 June 2023, [Single case studies are a powerful tool for developing, testing and extending theories](#). by Nickels, L., Fischer-Baum, S. & Best, W. 2022.

28 March 2023, [Increasing Transparency Through a Multiverse Analysis](#) by Steegen et al., 2016 & [Ethics and Statistics: Honesty and Transparency are not Enough](#) by Gelman, 2017.

28 February 2023, [Five Selfish Reasons to Work Reproducibly](#). by Markowitz, 2015

31 January 2023, [Measurement Schmeasurement](#) by Flake & Fried, 2020

14 December 2022, [The Creativity-Verification Cycle in Psychological Science: New Methods to Combat Old Idols](#). by Eric-Jan Wagenmakers et al., 2018

① 23 November 2022, [Is Preregistration Worthwhile?](#) by Aba Szollosi et al., 2020

Data management ABC / documentazione

Organising data

Electronic lab notebooks

MANTRA EDINA

Other tools, such as electronic lab notebooks (ELNs) can support the organisation of research data.

ELNs perform the same function as paper lab notebooks, and also enable better sharing and searching for data.

Some electronic lab notebooks also support integration with other tools used in the lab, and with research infrastructure.

Click on the headings to see the advantages of using an electronic lab notebook.

E-LAB NOTEBOOK SONO
PREZIOSI PER ORGANIZZARE
E DOCUMENTARE LA
VOSTRA RICERCA

Organising your data

Sharing data with others

Integration with other tools

- Data and documents have a unique ID assigned to them at time of creation which stays with them regardless of changes to the filename, preventing mix-ups with similarly named files.
- All files have appropriate metadata including creator, creation date, last modification, file type etc assigned to them at time of creation.
- Files may be tagged according to their content, grouping them together and facilitating their rapid identification and retrieval during searches.
- Powerful searching capabilities allow files to be quickly identified via filters, tags, unique IDs, modification dates, creator etc.
- Files may be organised into notebooks and folders making them easy to locate and browse.
- Files and datasets may be linked within documents allowing results and methodologies for different experiments to be kept together.
- Automatic recording of document versions helps ensure that research is recorded in a way that supports protection of intellectual property.



SciNote



labguru

[dati personali]



CESSDA guide
Data Management Expert Guide

I. Process lawfully, fair and transparent



The participant is informed of what will be done with the data and data processing should be done accordingly.

II. Keep to the original purpose



Data should be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes.

III. Minimise data size



Personal data that are collected should be adequate, relevant and limited to what is necessary.

⊖ Legal Basis

Personal data can only be processed when there is a valid legal basis to do so. The GDPR recognises six bases (grounds):

- consent of the data subject
- necessary for the performance of a contract
- legal obligation placed upon the data controller
- necessary to protect the vital interests of the data subject
- carried out in the public interest or in the exercise of official authority (public task)
- legitimate interest pursued by the data controller

IV. Uphold accuracy



Personal data should be accurate and, where necessary kept up to date. Every reasonable step must be taken to ensure that personal data that are inaccurate are erased or rectified without delay.

V. Remove data which are not used



Personal data should be kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed.

VI. Ensure data integrity and confidentiality



Personal data are processed in a manner that ensures appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss,

[basi legali]

Luca Schirru, corso interno SKILLS4EOSC

Conversazione in corso: Luca...

'Processing shall be lawful only if and to the extent that at least one of the following applies' (art 6(1)):

Consent

- 'the data subject has given consent to the processing of his or her personal data for one or more specific purposes'

Performance of a contract

- 'processing is necessary for the performance of a contract to which the data subject is party or in order to take steps at the request of the data subject prior to entering into a contract'

Compliance with a legal obligation

- 'processing is necessary for compliance with a legal obligation to which the controller is subject'

Protection of vital interests

- 'processing is necessary in order to protect the vital interests of the data subject or of another natural person'

Public interest

- 'processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller'

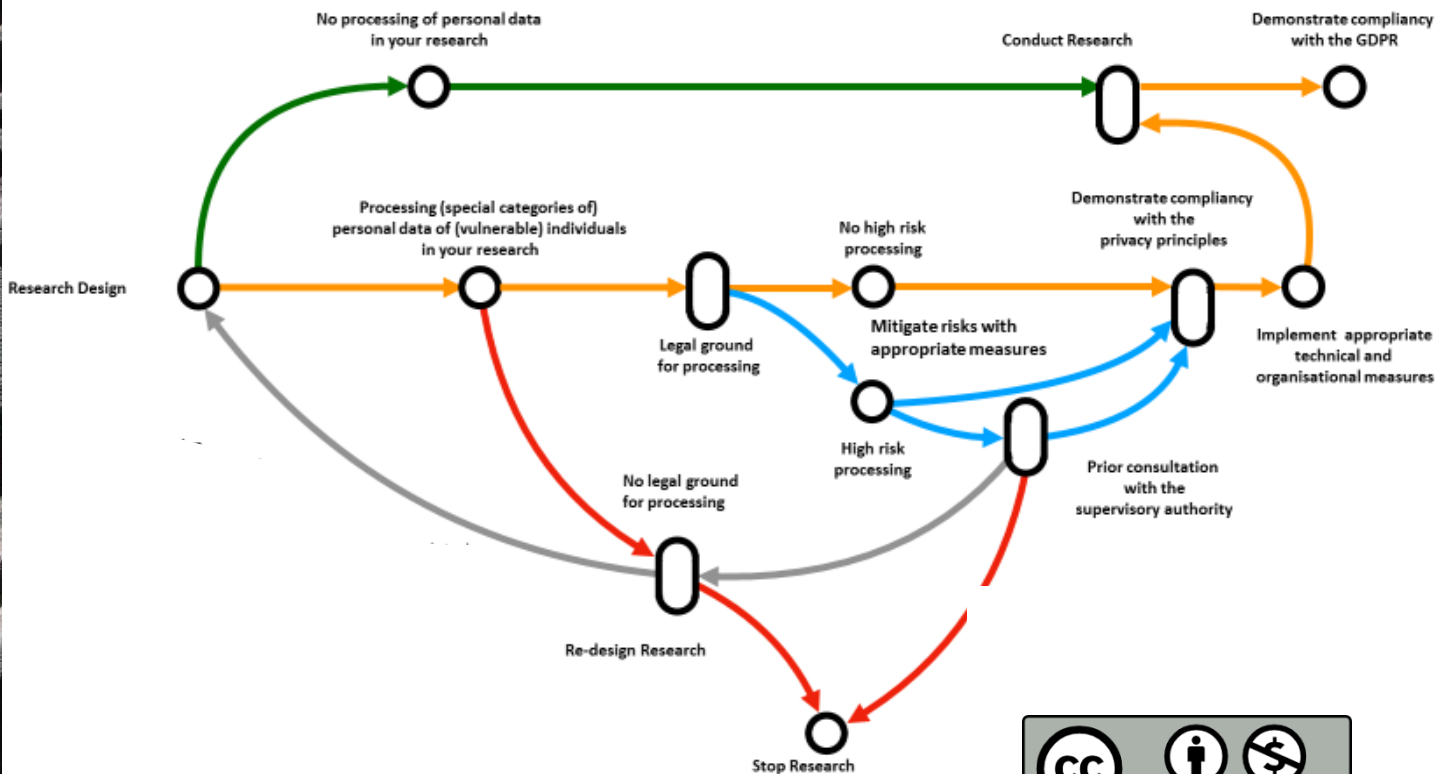
Legitimate interest

- 'processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child.'

Legal Bases

[Data and GDPR]

The Privacy Impact Assessment (PIA) Route Planner for Academic Research Inspired by Harry Beck's London Metro Map



Erasmus University Rotterdam
marlon.domingus@eur.nl
February 2018

The Logic of a Privacy Impact Assessment (PIA) for Academic Research

Q1. Do you process (special categories of) personal data of (vulnerable) individuals in your research?



YES
NO
 Proceed - no measures required for safeguarding privacy.

"Personal Data" (GDPR*, Article 4):
 Any information relating to an identified or identifiable natural person: a name, an identification number, location data, an online identifier, one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person.
"Special Categories of Personal Data (Sensitive Data)" (GDPR, Article 9):
 Data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation.

Action

Records of processing activities (GDPR*, Article 30):
 The university shall maintain a digital record of the processing activities in your research to demonstrate compliancy to the GDPR. This register contains:

1. The name and contact details of the researcher, the research partners and service providers;
2. The purposes of the processing;
3. A description of the categories of data subjects and of the categories of personal data;
4. The categories of recipients to whom the personal data have been or will be disclosed.

Q2. What is the legal ground for this processing?

Lawfulness of Processing (GDPR*, Article 6, 89):

1. The individuals participating in your research have freely given their explicit consent for one or more specific purposes.
2. Your research contributes to a legitimate interest, yet results in no high risks for the individuals participating in the research.
3. Your research has a scientific, historical or statistical purpose, yet results in no high risks for the individuals participating in the research.

Action

Data protection by design and by default (GDPR*, Article 25):
 Implement appropriate technical and organisational measures:

1. **Individual participating in your research (data subject).** Is the participant well informed, aware of possible risks for her/him and aware of the purpose of the research?
2. **Data.** Is the data de-identified and encrypted?
3. **Access Management.** How is access managed and controlled for the PI / team (expanded) / public?
4. **Software / Platform.** Are the *Terms of Service* for used software / platform checked (where is the data and who has access and has which usage rights)?
5. **Devices.** Are devices used safe? Encrypted drive, encrypted communication, strong password / two factor authentication.
6. **Partners.** Are the research partners / service partners trusted and are appropriate legal agreements made, with regards to roles, rights and responsibilities?
7. **Safe and secure collaboration.** Is the ((cross border) communication to, in and from the) collaboration platform end to end encrypted, are roles and permissions defined and implemented, is logging and monitoring implemented?

YES

NO
 Stop research or redefine research.

Q3. Is this processing a high risk processing?

Criteria for high risk processing (WP29 - DPIA Guideline):**

1. Evaluation or scoring
2. Automated-decision making with legal or similar significant effect
3. Systematic monitoring
4. Sensitive data or data of a highly personal nature
5. Data processed on a large scale
6. Matching or combining datasets
7. Data concerning vulnerable data subjects
8. Innovative use or applying new technological or organisational solutions
9. When the processing itself prevents data subjects from exercising a right or using a service or a contract

Action

Prior consultation (GDPR*, Article 36):

1. The Data Protection Officer shall, on behalf of the researcher, consult the supervisory authority, prior to the processing (the research) when the processing would result in a high risk *in the absence of measures* to mitigate the risk.

Action

Principles relating to processing of personal data (GDPR*, Article 5):
 Demonstrate compliancy with the principles: lawfulness, fairness, transparency, purpose limitation, data minimisation, accuracy, storage limitation, integrity, confidentiality and accountability.

YES

NO
 Proceed - measures required for safeguarding privacy.

** Article 29 Data Protection Working Party: *Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679.* Adopted on 4 April 2017. As last Revised and Adopted on 4 October 2017. Online available at: https://ec.europa.eu/newsroom/document.cfm?doc_id=47711

GDPR e ricerca



DARIAH consent wizard



Welcome to the DARIAH ELDAH Consent Form Wizard (CFW)!

Introduction

The GDPR in research, a.o. special categories of personal data, processing in/outside the European Economic Area (EEA), and privacy by design/default.

- > [GDPR in research: introduction](#)
- > [FAQ GDPR in research](#)

Data minimisation

The data minimisation principle comprises that data has to be adequate, relevant and limited to what is necessary for the purposes for which they are processed.

- > [GDPR in research: data minimisation](#)
- > [FAQ data minimisation](#)

Data quality

The data quality principle comprises that data has to be of good quality, i.e. the data has to be accurate and up-to-date.

- > [GDPR in research: data quality](#)
- > [FAQ data quality](#)

Goal setting

In the goal setting, you describe what personal data you process, with which legitimate purpose and for how long.

- > [GDPR in research: goal setting](#)
- > [FAQ goal setting](#)

Minimisation of use

Minimise the processing of and access to personal data, for a pre-defined purpose and period of time, and only by authorised persons.

- > [GDPR in research: minimisation of use](#)
- > [FAQ minimisation of use](#)

Security measures

Make sure that the personal data you collect is well secured. When working with personal data, make use of privacy protection techniques.

- > [GDPR in research: security measures](#)
- > [FAQ security measures](#)

Transparency

The GDPR requires the controller to be transparent to data subjects about the processing of their personal data.

- > [GDPR in research: transparency](#)
- > [FAQ transparency](#)

Rights of data subjects

Fundamental of the GDPR are the right of data subjects concerning the processing of their personal data.

- > [GDPR in research: rights of data subjects](#)
- > [FAQ rights of data subjects](#)

Research Data Management

GDPR in research

Search

HOME PLANNING RESEARCH COLLECTING DATA PROCESSING DATA ARCHIVING DATA **GDPR IN RESEARCH** SUPPORT & TRAINING

Research Data Management > GDPR in research

GDPR in research

As of May 25 2018, the GDPR (General Data Protection Regulation), or AVG (Algemene Verordening Gegevensbescherming) in Dutch, will apply to the entire European Union. The GDPR has its implications for research. Anyone who collects personal data within Radboud University during their research, must follow 8 guidelines following the Privacy by design principle.

The guidelines are only applicable for research with **personal data**. Personal is any data that can lead to the identification of an individual. For example name, birth date, email-address and IP address are direct personal data. But also a combination of data can lead to the identification of an individual and should therefore be treated as personal data. If you **don't process personal data** in your research, then the GDPR is not applicable. This is for instance the case when your research only includes anonymised data (but be aware that pseudonymised data is personal data).



[anonimizzare]

Anonymisation

[UK Data service](#)

Anonymisation is a valuable tool that allows data to be shared, whilst preserving privacy. The process of anonymising data requires that identifiers are changed in some way such as being removed, substituted, distorted, generalised or aggregated.

A person's identity can be disclosed from:

- **Direct identifiers** such as names, postcode information or pictures
- **Indirect identifiers** which, when linked with other available information, could identify someone, for example information on workplace, occupation, salary or age



Amnesia OpenAIRE

High accuracy Data Anonymization.

Perform research and share your results that satisfy GDPR guidelines by using data anonymization algorithms.

GET STARTED



Unlock sensitive data analysis

Use Amnesia to transform personal data to anonymous data that can be used for statistical analysis. Data anonymized with Amnesia are *statistically guaranteed* that they cannot be linked to the original data.



Become GDPR compliant

Create anonymous datasets from personal data that are treated as statistics by GDPR. Anonymous data can be used without the need for consent or other GDPR restrictions, greatly reducing the effort needed to extract value from them.



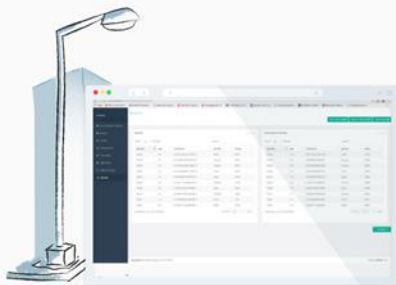
High Usability & Flexibility

Anonymization tailored to user needs through a graphical interface. Guide the algorithm and decide trade-offs with simple visual choices. Developers can incorporate Amnesia anonymization engine to their project through a ReST API.

How it works

Get anonymous data in 3 steps

- ✓ Guarantees no links to original data
- ✓ Offers k-anonymity & l-diversity
- ✓ Allows minimal reduction of data utility



1 Insert your data

Amnesia accepts complex object relational data in delimited text files.

2 Select and Preview the data to anonymize

Visual representations of anonymization parameters and results allow non-expert users to tailor the anonymization process to their needs.

3 Download your data anonymized

The process is completed without any sensitive data leaving your premises!

2. RENDERLI FAIR



FAIR significa [machine readable]

FINDABLE

- IDENTIFICATIVI
- METADATI

INTEROPERABLE

- STANDARD
- ONTOLOGIE

ACCESSIBLE

- DOVE SONO CONSERVATI I DATI E A QUALI CONDIZIONI DI ACCESSO
- NON «OPEN»
- FORMATI APERTI

REUSABLE

- LICENZE
- DOCUMENTAZIONE

SCIENTIFIC DATA

We'd like to understand how you use our websites in order to improve your experience.

Open Access | [FAIR guide](#), Nature, March 2016
Published: 15 March 2016

The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson, Michel Dumontier, [...] Barend Mons

2020

Data Intelligence

Issues Online Early About Submit

Volume 2, Issue 1-2
Winter-Spring 2020

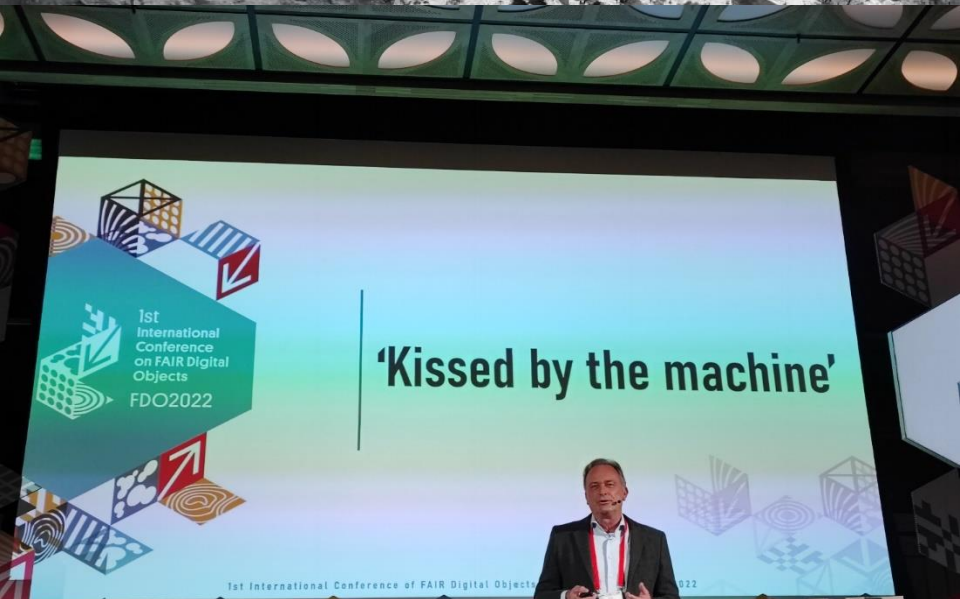
January 01 2020

FAIR Principles: Interpretations and Implementation Considerations

Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Juty, Dominique Batista, Simon Coles, Ronald Cornet, Mélanie Courtoi, Mercè Crosas, Michel Dumontier, Chris T. Evelo, Carole Goble, Giancarlo Guzzardi, Karsten Kryger Hansen, Ali Hasnain, Kristina Hettne, Jaap Heringa, Rob W.W. Hooft, Melanie Imming, Keith G. Jeffrey Rajaram Kalyaperumal, Martijn G. Kersloot, Christine R. Kirkpatrick, Tobias Kuhn, Ignasi Labastida, Barbara Magagnoli, Peter McQuilton, Natalie Meyers, Annalisa Montesanti, Mirjam van Reizen, Philippe Rocca-Serra, Robert Pergi, Susanna-Assunta Sansone, Luiz Olavo Bonino da Silva Santos, Juliane Schneider, George Strawn, Mark Thompson, Andre Waagmeester, Tobias Winkel, Mark D. Wilkinson, Frank W. Willemsen, Rafael W. Winters, Marco Zamboni

DATA
Intelligence

Kissed or missed?



PRINCIPI FAIR SONO
«MACHINE ACTIONABLE»
(DIVERSO DA READABLE)

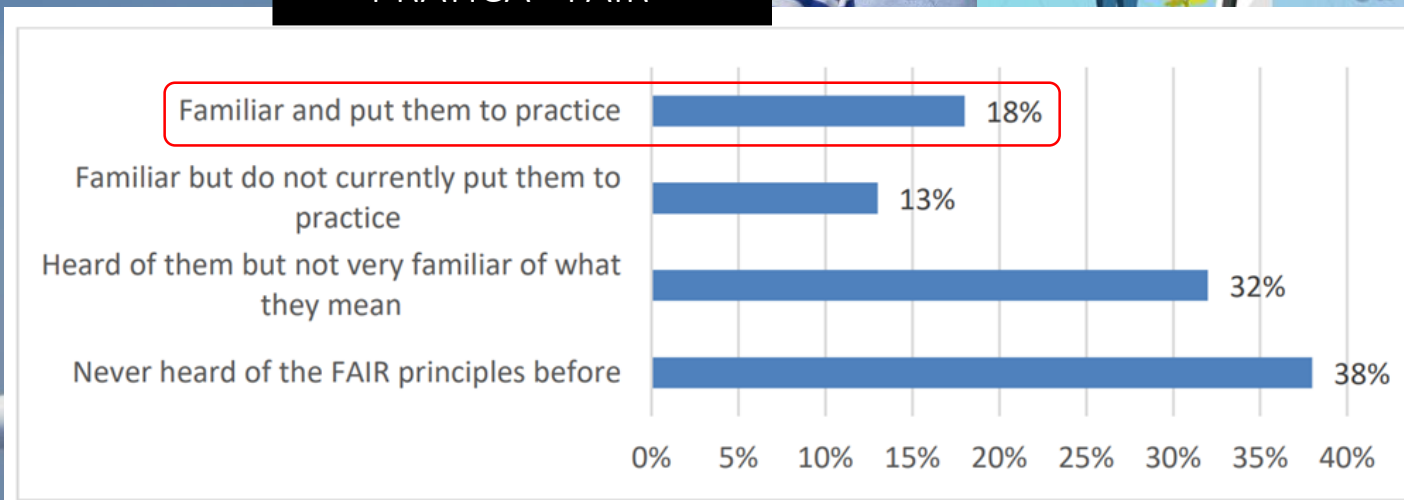
FAIR = FULLY AI READY

SE NO... **YOU'LL BE MISSED (INSTEAD OF KISSED) BY THE MACHINE**

Dati FAIR???



SOLO IL 18% DEI RICERCATORI METTE IN PRATICA «FAIR»



Per iniziare

FARM DATA TRAIN



SMETTIAMO DI USARE
«DATA SHARING» -
USIAMO «**DATA VISITING**»
[I DATI STANNO DOVE SONO, SE SONO
«FAIR» LE MACCHINE LI
TROVANO/USANO]

FAIR principles

To be Findable:

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

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 RE-USABLE:

- R1. meta(data) have 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 their provenance.
- R1.3. (meta)data meet domain-relevant community standards.

Force 11

«ACCESSIBLE»
≠ «OPEN»
= DOVE E A QUALI
CONDIZIONI
I DATI SONO
ACCESSIBILI

9150 FAIR principles

Data Intelligence **2020** Issues Online Early About Submit

Volume 2, Issue 1-2
Winter-Spring 2020



January 01 2020

FAIR Principles: Interpretations and Implementation Considerations

Annika Jacobsen, Ricardo de Miranda Azevedo, Nick Judy, Dominique Batista, Simon Coles, Ronan Miliarie Courtois, Marco Crosas, Michel Dumortier, Chris T. Evelo, Carole Gobbe, Giancarlo Guizzo, Karsten Kroyer Hansen, Ali Hasnain, Kristina Hettne, Jaap Heringa, Rob W.W. Hooff, Melanie Imm, Rajaram Kalyaperumal, Martin G. Kerstoft, Christine R. Kirkpatrick, Tobias Kuhn, Ignasi Labastida, Peter McQuilton, Natalie Moyers, Annalisa Montesanti, Mirjam van Peelsen, Philippe Rocca-Serra, Susana S. Almeida Santos, Işıl Olusan Bostun, Silvia Santos, Juliana Schneider, Geovine Strödel

< Previous Article Next Article >

Article Contents

NO MISTAKES!

- **Findability:** Digital resources should be easy to find for both humans and computers. Extensive machine-actionable metadata are essential for automatic discovery of relevant datasets and services, and are therefore an essential component of the FAIRification process [14].
- **Accessibility:** Protocols for retrieving digital resources should be made explicit, for both humans and machines, including well-defined mechanisms to obtain authorization for access to protected data.
- **Interoperability:** When two or more digital resources are related to the same topic or entity, it should be possible for machines to merge the information into a richer, unified view of that entity. Similarly, when a digital entity is capable of being processed by an online service, a machine should be capable of automatically detecting this compliance and facilitating the interaction between the data and that tool. This requires that the meaning (semantics) of each participating resource – be they data and/or services service – is clear.
- **Reusability:** Digital resources are sufficiently well described for both humans and computers, such that a machine is capable of deciding: if a digital resource *should* be reused (i.e., is it relevant to the task at-hand?); if a digital resource *can* be reused, and under what conditions (i.e., do I fulfill the conditions of reuse?); and *who to credit* if it is reused.



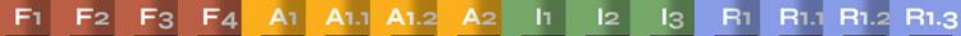
2022

Home FAIR Criteria Qualification Fellowship

The FAIR Guiding Principles

Findable

The first step in (re)using digital resources should be easy to find for humans and machines. Readable metadata are essential for automatic discovery of relevant datasets and services, so they are an essential component of the FAIRification process.



Interpretation of F1

Principle F1 states that digital resources, i.e., data and metadata, must be assigned a globally unique and persistent identifier which serves as a permanent machine interpretable reference. The GO FAIR Foundation emphasises the need for persistence and global uniqueness, as well the property of resolvability of the identifiers (see also A1). Globally unique means that the identifier is guaranteed to unambiguously refer to the intended resources (where 'world' is interpreted as 'universal' as there are described digital assets outside the 'world'). Therefore, it is



Interpretation of A1.2

This principle clearly demonstrates that following the FAIR guiding principles is not equal to making all data 'open'.

ORIGINAL
INTERPRETATION

FAIR for dummies

COSA DEVE FARE IL
RICERCATORE

COSA FA IL
REPOSITORY

Explanation of the [FAIR data principles](#) ²⁰¹⁹

Wilkinson et al. (2016), The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data* 3, [doi:10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18)

Principle	In other words	Researcher's responsibility	Requirements to be fulfilled by the repository	
To be findable: Data and metadata should be easy to find by both, humans and computer systems. Basic machine readable descriptive metadata allows the discovery of interesting data sets and services.	F1. (meta)data are assigned a globally unique and persistent identifier	Each data set is assigned a globally unique and persistent identifier (PID), for example a DOI , ARK , RRID ... These identifiers allow to find, cite and track (meta)data.	Ensure that each data set is assigned a globally unique and persistent identifier. Certain repositories automatically assign identifiers to data sets as a service. If not, researchers must obtain a PID via a PID registration service.	A repository needs to have a predictable way to assign a PID to each component of a dataset (e.g. each file or nanopublication), in order to be able to include these identifiers into the corresponding metadata before the submission.
	F2. data are described with rich metadata (defined by R1 below)	Each data set is thoroughly (see below, in R1) described: these metadata document how the data was generated, under what term (license) and how it can be (re)used, and provide the necessary context for proper interpretation. This information needs to be machine-readable.	Fully document each data set in the metadata, which may include descriptive information about the context, quality and condition, or characteristics of the data. Another researcher in any field, or their computer, should be able to properly understand the nature of your dataset. Be as generous as possible with your metadata (see R1).	Allow researchers to upload metadata for each data set.
	F3. metadata clearly and explicitly include the identifier of the data it describes	The metadata and the data set they describe are separate files. The association between a metadata file and the data set is obvious thanks to the mention of the data set's PID in the metadata.	Make sure that the metadata contains the data set's PID.	Allow researchers to upload metadata for each data set.
	F4. (meta)data are registered or indexed in a searchable resource	Metadata are used to build easily searchable indexes of data sets. These resources will allow to search for existing data sets similarly to searching for a book in a library.	Provide detailed and complete metadata for each data set (see F2).	Request and store part of the metadata in a structured way, for example by providing a form with specific fields to be completed or by providing an XML schema to be used by the researchers. For example the storing of PID's, author names, disciplines, etc. will facilitate the creation of indexes. However, it must remain possible to provide arbitrary metadata in addition.

FAIR/Open



AS EARLY AS POSSIBLE
AS FAIR AS POSSIBLE
AS OPEN AS POSSIBLE

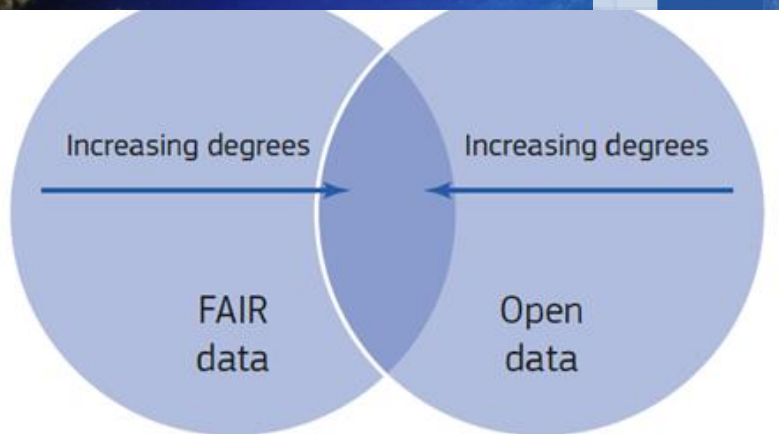
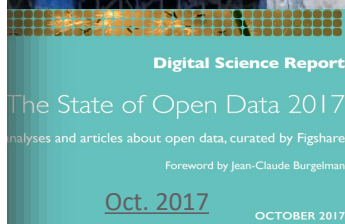


Figure 4. The relationship between FAIR and Open

"Open data is like a renewable energy source: it can be reused without diminishing its original value, and reuse creates new value."



A TENDERE, I DUE INSIEMI SARANNO SEMPRE PIÙ SOVRAPPOSTI. MA ESISTERANNO SEMPRE DATI PERFETTAMENTE FAIR CHE NON POSSONO ESSERE OPEN

Article type: Research Article

Authors: Mons, Barend^{a,b,c,*} | Neylon, Cameron^d | Velterop, Jan^e | Dumontier, Michel^f | da Silva Santos, Luiz Olavo Bonino^{b,g} | Wilkinson, Mark D.^h

4. ...and what FAIR is not

FAIR is not a standard: The FAIR guiding principles are sometimes incorrectly referred to as a 'standard', even though the original publication explicitly states they are not [25]. The guiding principles allow many different approaches to rendering data and services Findable, Accessible, Interoperable, to serve the ultimate goal: the reuse of valuable research objects. Standards are prescriptive, while guidelines are permissive. We suggest that a variety of valuable standards can and should be developed, each of which is guided by the FAIR Principles. FAIR simply describes the qualities or behaviours required of data resources to achieve – possibly incrementally – their optimal discovery and scholarly reuse.

FAIR is not equal to RDF, Linked Data, or the Semantic Web The reference article on Scientific Data [25] emphasises the machine-actionability of data and metadata. This implies (in fact, requires) that resources that wish to maximally fulfil the FAIR guidelines must utilise a widely-accepted machine-readable framework for data and knowledge

FAIR is not just about humans being able to find, access, reformat and finally reuse

data: The official press release for the FAIR Principles states the authors' position clearly: "The release of the FAIR Principles is a data publication autonomously, and not a standard. The FAIR Principles. Computers are now able to process data autonomously. In recent surveys, the time reported to be spent dealing with discovering and reusing data has been pegged at 80% [19]. Were this time spent dealing with FAIR data and services, it would be a significant improvement today. The avoidance of time-wasting is a key aspect of data stewardship. To serve this potential, data and services should be actionable wherever possible.

FAIR is not equal to Open: The 'A' in FAIR stands for 'Accessible under well defined conditions'. There may be legitimate reasons to shield data and services generated with public funding from public access. These include personal privacy, national security, and competitiveness. The FAIR principles, although inspired by Open Science, explicitly and

3. What FAIR is...

FAIR refers to a set of principles, focused on ensuring that research objects are reusable, and **actually will be reused,** and so become as valuable as is possible. They deliberately do not **specify technical requirements,** but are a set of guiding principles that provide for a **continuum of increasing reusability, via many different implementations.** They describe characteristics and aspirations for systems and services to support the creation of valuable research outputs that could then be rigorously evaluated and extensively reused, with appropriate credit, to the benefit of both creator and user.

- PRINCIPI, NON STANDARD [VANNO IMPLEMENTATI]
- NON SOLO PER UMANI
- DIVERSO DA LINKED DATA, RDF...
- DIVERSO DA «OPEN»

FAIR vero o falso

True

"FAIR data" are not necessarily "open data": the principle of accessibility does not say anything about openness.

Both humans and machines are intended as "digesters" of data.

The FAIR principles apply to both data and metadata.

The FAIR principles are not rules or standards to be applied.

False

"FAIR data" are the same as "open data".

The FAIR principles mostly address reusability for humans.

The FAIR principles apply mainly to data.

The FAIR principles are a "standard".

FAIR in sintes

- FAIR data training
- Findable
- Accessible
- Interoperable
- Reusable
- FAIR for Developers
- FAIR data self-assessment tool

f t in +SHARE

F1. (meta)data are assigned a globally unique and eternally binding identifier

There are many resources created by the ARDC on the topic of **metadata**

- Metadata guide
- Data versioning

The ARDC has information on persistent identifiers on three different levels

- Persistent identifiers: awareness level
- Persistent identifiers: working level
- Persistent identifiers: expert level

It is also a provider of services for minting persistent identifiers of many different types (the list of services and the data being identified):

- Digital Object Identifier (DOI) System for research data
- Handle minting Service (Identify My Data)
- International Geo Sample Numbers (IGSN)

Complementary to the assignment of persistent identifiers is their proper



Findable Magnifying glass icon	Persistent Identifiers (PIDs) ID icon	Rich metadata Document icon	Indexed data repositories Database icon	PIDs in metadata ID icon
Accessible Hand pointing icon	Standard communications protocol Network icon	Open, free protocol Dollar sign with slash icon	Authentication, where necessary User and shield icon	Metadata is always available Infinity icon
Interoperable Gears icon	Vocabularies Tree diagram icon	Vocabularies are FAIR Magnifying glass and tree icon	Linked metadata Network diagram icon	
Reusable Recycling icon	Metadata have multiple attributes Document icon	Usage license Open lock icon	Provenance Flowchart icon	Community standards Group of people icon

Responsible

RECOMMENDATIONS

- Clarify all legal issues at the beginning of your research project and include the findings of this process in the data management plan.
- Use checklists adequate to your research topic/discipline.
- Check the resources indicated by DARIAH-CLARIN. (See further reading).
- In the case of personal data ensure that only relevant people can access the data and that these are clearly identified (see GDPR).
- Ask for consent to share anonymised data and establish transparent and well-documented anonymisation routines that consider not just direct identifiers, but also how a combination of indirect identifiers could reveal identities. (See for example the guide on informed consent in the CESSDA data management expert guide).
- Avoid collection of (sensitive and non-sensitive) personal data when possible.
- Get legal support (IPR, copyright, patents, trademarks etc.) from your home institution. If there is no dedicated office for this purpose, try to get information from your university library, as its staff are often confronted with such issues.
- If you need permission from the copyright holder in order to use sources like images for your publication, try to get one that covers both printed and digital copies.
- Finally, check the recommendations in the section on Licences that are closely related to this section.

DISSEMINATION

What it means to disseminate data in the Humanities

IDENTIFY

Research Data in the Humanities



FAIR DATA and the HUMANITIES

DEPOSIT for PRESERVATION, CITE & SHARE
License and Legal aspects
TDRs and PIDs for the Humanities

PLAN
Data Management Plans

COLLECT/PRODUCE & STRUCTURE & STORE
Types and Formats, Metadata and Data Models for the Humanities



Sustainable and FAIR Data Sharing in the Humanities

ALLEA Report | February 2020
February 2020

- If applicable, determine if the body funding your research has particular requirements for a DMP or offers a template for framing your plan. If there is no required template, choose an existing appropriate one (e.g. via DMPOnline).
- Devise a DMP prior to collecting data. Define and plan for your data: all research projects deal with data. If your project includes the analysis of text corpora, for example, then the corpora themselves are data, and you should make sure they are clearly described, documented, and managed according to the FAIR principles so your research is reusable by others.
- Plan documentation of metadata: in order for your data to be comprehensible in the future and/or reusable by others, they will need descriptive metadata created according to a common schema to understand the content/purpose of the research. The richer the metadata, the more intelligible and useful the dataset (see section on Metadata).
- Use standardised terminology to increase interoperability. Consider employing vocabularies or ontologies that follow FAIR principles to increase interoperability and findability (e.g. see [EUBioinformatics](#)).
- Consider the right questions to be answered in your DMP that can account for discipline-specific requirements. The DMP templates suggested by funders are quite high level and provide generic guidance for file naming or versioning conventions, database structuring and can be a good start. Tools like [dmponline.dcc.ac.uk](#) provide discipline specific examples that can be of further reference.
- DMP as living documents: Update your data management plan regularly in order to take into account any potential relevant changes such as using new data types and/or models, technology, new institutional data management policies, reassessing legal aspects or licences for legal compliance etc.
- Depending on the size of the organisation: think of providing institutional support for research data management (RDM); organise information sessions to raise awareness about good research data management, and the risks of not managing it early.
- If possible, consider involving library and/or repository support staff from the initial stages of research data management planning to discuss the best solutions, specifications, standards and protocols along which the repository operates. Repository staff can also assist scholars with understanding any specific data management requirements and associated costs.
- Factor the cost of research data management (time or human resources) into budgetary requirements at the point of application.

RECOMMENDATIONS

- A good starting point is to consult the Metadata Standards Directory, a community-maintained directory hosted by the Research Data Alliance: <https://rd-alliance.github.io/metadata-directory/>.
- Metadata works best when terminology is consistent, e.g. naming conventions are followed, spelling is normalised, and so on. Depending on the complexity and size of your metadata, consider using a tool such as Open Refine to 'clean' your metadata.
- For greater searchability and interoperability, researchers should also consider using controlled vocabularies to identify common terminology when populating metadata fields. The Library of Congress maintains a controlled vocabulary for subject headings: <https://www.loc.gov/standards/subject/>.
- Metadata should include a clear and explicit reference to the dataset with the inclusion of a PID in the metadata (see section on Trustworthy Data Persistent Identifiers).
- Metadata should be as rich as possible in order to better contextualise your data and consider more detailed descriptions, and fuller provenance information, as well as a spectrum of available metadata fields.
- Metadata should be machine-readable.

RECOMMENDATIONS

- Data models go FAIR: the FAIR Guiding Principles, correctly applied, ensure data are findable, accessible, interoperable and reusable. Data modelling should take this into account by using formal, easily accessible languages for knowledge representation, providing persistent identifiers, open standards, well documented Application Programming Interfaces (API), generic user interfaces and rich metadata. The [FAIRification process](#), developed by the GO FAIR initiative offers a system on how to shape the data modelling.
- Use open standards, and wherever possible, standardised technologies and procedures should be used. The World Wide Web Consortium W3C maintains several standards relevant for data models like XML and RDF. Within XML the Text or Music Encoding Initiative (TEI/MEI) or specific expressions of them have become standards for text or music editions. The query language SPARQL and the representation tool for linked data ISON-LD are common standards for RDF (refers to FAIR principle 1).
- Prefer human and machine-readable systems: coding of data models and of the actual data that is both human and machine-readable in a unified way provides better sustainability and long-term accessibility than machine-readable only code (binary codes), that may use different formats for data model description and the actual data. For both, hierarchical data models and graph-based data, various serialisations (file formats) are available that fulfil this condition (XML, TEI/XML, Turtle, N3, RDF/XML), whereas SQL based technologies need bigger efforts.
- Normalise as much as possible: to avoid redundant information, the content of databases should be normalised as far as possible, using for example authority files like VIAF and identifiers like DOI, ARK, ISBN, GND and the like. To foster the exchange of data, standardised vocabularies and ontologies are needed as well, but an overall ontology for the humanities has not yet been established. The ontology CIDOC-CRM and especially some extensions are well on their way to become a reference model for cultural heritage information.
- Data models follow the data management plan (DMP): when establishing a data model, researchers should keep the whole lifecycle of their data in mind, as it should be outlined in a DMP. Therefore, an extensive documentation of the data model, its software and tools are highly relevant and facilitates the transfer of data in a secure and trusted repository in order to keep them accessible. The same is true here: the more you use open standards for your data model, the easier this task becomes.

FAIR è graduale

FAIR = CONTINUUM
«AS FAIR AS POSSIBLE»

Inclusiveness: consider FAIR as a process

If FAIR is not seen as a continuum, we risk losing communities who are not well advanced in sharing their data in a FAIR way, as well as advanced communities for whom the effort to attain optional indicators doesn't outweigh the effort required. In addition to avoiding "mandatory" criteria, using multi-step maturity scales to measure the FAIRness level of a resource, instead of a yes/no evaluation for each criterion, would provide an inclusive system, and a way to set up



Interim recommendations on FAIR Metrics for EOSC

February 2020

Draft for consultation

Feb. 2020

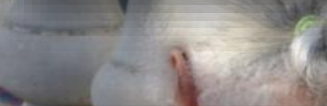
Per sapere se

- DOMANDE
- VERIFICA INTENZIONI
- VERIFICA CONOSCENZA
- FORNISCE INFORMAZIONI



Let's assume you have research data almost ready for uploading to a repository: do you already know how you and the repository can work together to make the data as findable, accessible, interoperable and reusable (FAIR) as possible? By guiding you through the assessment process, the FAIR-Aware tool can help you to better understand the FAIR Principles and how making data FAIR can increase the potential value and impact of your data.

FAIR-Aware is an disciplinary-agnostic online tool developed by the FAIRsFAIR project. Different scientific communities can adapt it to their own use. You should, however, have a target dataset in mind to be able to answer the questions and complete the assessment.



unique persistent and resolvable identifier when deposited with a data repository?

What does this mean?

A **persistent identifier** is a long-lasting reference to a resource. The **data(set)** you deposit in a **data repository** should be assigned a globally unique, persistent and resolvable identifier (PID) so that both humans and machines can find it. Persistent identifiers are maintained and governed so that they remain stable and direct the users to the same relevant object consistently over time. Examples of PIDs include Digital Object Identifier (DOI), Handle, and Archival Resource Key (ARK).

Why is this important?

If your data(set) or metadata does not have a PID, you run the risk of "link rot" (also known as "link death"). When your data(set) or metadata is moved, updated to a new version, or deleted, older hyperlinks will no longer refer to an active page. Without a PID, others will not be able to find or reuse your data(set) or metadata in the long-term.

How to do this?

When you upload your data(set) or metadata to a data repository, the data repository (or other service providers) usually assigns a PID. Repositories ensure that the identifier continues to point to the same data or metadata, according to access terms and conditions you specified.

There are many different types of PIDs, each with their own advantages, disadvantages, and disciplines they are typically used in. Generally speaking, the data repository will have thought about these aspects before deciding which PID type to use. In case you have to choose the PID type yourself, you can visit the Knowledge Hub on the PID Forum for guidance. Some disciplines or organisations also provide tools to help you make this choice, see for example this Persistent Identifier Guide for cultural heritage researchers. Once you have chosen a PID type, you can search for data repositories providing that specific PID in registries such as Re3data or FAIRsharing (see related databases).

Not all data you produce during your research will need a PID. In general, those that underpin published findings or have longer term value are worth assigning a PID. If in doubt about which data should be allocated a PID, speak to your local research data management support team or the data repository.

FINDABLE

1. Are you aware that a data(set) should be assigned a global persistent and resolvable identifier when deposited with a data repository?
2. Are you aware that when you deposit a data(set) in a data repository you will need to provide discovery metadata in order to make your data(set) findable, understandable and reusable to others?
3. Are you aware that the data repository providing access to your data(set) should make the metadata describing your data(set) available in a format readable by machines as well as humans?

ACCESSIBLE

4. Are you aware that access to your data(set) may need to be controlled and that metadata should include licence information under which your data(set) can be reused?
5. Are you aware that metadata should remain available over time even if the data(set) is no longer accessible?



Per sapere quanto siete FAIR

[FAIRassist.org](https://fairassist.org)

<https://fairassist.org/#/>

Help you discover resources to measure and improve FAIRness.

FAIRassist is the new, under development, educational component of the well established FAIRsharing resource.

Resource	Execution Type	Key Features	Organisation	Target Objects	Reading Material
5 Star Data Rating Tool	Manual - questionnaire	Based on rating systems and maturity models	CSIRO OzName	Datasets	
AutoFAIR	Semi-automated	A portal for automating FAIR assessments for bioinfo	Department of Computer		
FAIR enough	Automated			<ul style="list-style-type: none"> 1. Core universal maturity indicators and community compliance tests 2. Stable and fast evaluations execution (less than 1min for most evaluated resources, no commercial license required) 3. Library for defining, publishing and registering new maturity indicators 4. Supports ORCID authentication for creating collections and authoring evaluations 	Maastricht Uni
Data Stewardship Wizard	Predictive; based on a manually filled questionnaire	Helps researchers to design a data stewardship process to achieve the highest reasonable FAIR data.			
FAIR-Aware	Manual - questionnaire			<ul style="list-style-type: none"> 1. Online self-assessment that helps to assess current level of awareness on making datasets FAIR before depositing them in a data repository. 2. Added guidance texts explain the what, why, and how of each FAIR practice. 3. Trainer functionality allows flexible use of the tool for your own purpose 	FAIRsFAIR (D
F-UJI	Automated	The REST API support a programmatic assessment of objects based on a set of core metrics developed by the FAIR community. The metrics specification is available at https://doi.org/10.26434/chemrxiv-2019-08-01			
FAIR-Checker	Automated			FAIR-Checker is a web interface to evaluate FAIR metrics (as implemented through the FAIR Evaluation Service APIs https://fairsharing.github.io/FAIR-Evaluator-FrontEnd/) and to provide developers with technical FAIRification hints. It's also a Python framework aimed at easing the implementation of FAIR metrics.	IFB (ELIXIR)
FAIR Data Self-Assessment Tool	Manual - questionnaire	Educational and Informational purposes			
FAIR Evaluator	Automated	<ul style="list-style-type: none"> 1. Core universal maturity indicators 2. Compliance tests 3. Evaluation tool 			
FAIRdat	Manual - questionnaire			A 5-star rating of the FAIR principles	DANS
FAIRness self-assessment grids	Manual - checklist			<ul style="list-style-type: none"> 1. Assessment grids: quick and extensive 2. Designed as a decision tree 3. Researcher focused 	RDA-SHAR
FAIRshake	Manual - questionnaire			1. FAIR metrics (questions) and rubrics (collection of metrics)	NIH Data Com

...FAIR per enti di ricerca

DO I-PASS FOR FAIR?

Oct. 2020



Policy: Does your organization (institute / university (of applied sciences)) have a FAIR research data policy?

1. Are the FAIR principles explicitly mentioned in the data (or research data) policy of your organization?	The FAIR principles are not explicitly mentioned, but there is a reference to sustainable and long term storage.	The F and A principle are explicitly mentioned.	All FAIR principles are explicitly mentioned (or there is an explicit reference to the overarching Open Science concept).	<input type="checkbox"/> Beginner <input type="checkbox"/> Intermediate <input type="checkbox"/> Advanced
2. Is the general research data policy translated into faculty/institutes specific policies?	Less than 20% of the faculties or sub-	20-80% of the faculties or sub-units have a specific policy.	More than 80% of the faculties or sub-units have	<input type="checkbox"/> Beginner <input type="checkbox"/> Intermediate

Services: Does your organization have a (virtual) DCC which provides services (including infrastructure) to allow comply with FAIR?

3. Is the general research data policy supported by dedicated service units (e.g. DCC) ?	5. Which services does your organization provide in order for researchers to comply with the F principles	We provide or refer to a service to deliver a PID for a data set	We provide or refer to service for PID and adding metadata (including reference to the dataset).	On top of adding PIDs and metadata, we provide or refer to a service to make the data findable through indexed resources.	<input type="checkbox"/> Beginner <input type="checkbox"/> Intermediate <input type="checkbox"/> Advanced
	6. Which services does your organization provide in order for researchers to comply with the A principles?	We provide or refer to a service with a standard and open communication protocol as to access the data.	We provide or refer to a service to enable restricted access to data and meta data (authentication protocol)	Our organization as an archiving policy, both on data and meta data.	<input type="checkbox"/> Beginner <input type="checkbox"/> Intermediate <input type="checkbox"/> Advanced
4. Which means and channels used to communicate the research data policy and services to the researchers	7. Which services does your organization provide in order for researchers to comply with the A principles?	We provide or refer to a service that enriches the dataset metadata, referencing to other datasets and articles (I3).	We provide or refer to a service to attribute (meta)data that use a formal, accessible, shared, and broadly applicable language for knowledge representation.	We provide or refer to a service to attribute (meta)data that use vocabularies that follow FAIR principles in order to facilitate machine readability.	<input type="checkbox"/> Beginner <input type="checkbox"/> Intermediate <input type="checkbox"/> Advanced
	8. Which services you're your organization provide in order for researchers to comply with the R principles?	We provide advice on licenses and provide or refer a service to add licenses in the metadata	We monitor the reuse of our data, both within the institute as well as outside	We perform quality control on the (meta)data, e.g. adequate read me file as to assure that somebody else can actually reuse the data. We assure that (meta)data meet domain-relevant community standards	<input type="checkbox"/> Beginner <input type="checkbox"/> Intermediate <input type="checkbox"/> Advanced <input type="checkbox"/> Not yet

Institute

Supporto / Come essere FAIR

zenodo Search Upload Communities

January 11, 2022 2022 Book Open Access

D7.4 How to be FAIR with your data. A teaching and training handbook for higher education institutions

5 – FAIR lesson plans

6 – Implementing FAIR

- 6.1 Introduction
- 6.2 Getting to FAIR institutional policies
- 6.3 Data management planning

Engelhardt, Claudia; Biernacka, Ka; Demchenko, Yuri; Downes, Stephe; Germer, Kerstin; Helbig, Kerstin; H; Jetten, Mijke; Karimova, Yulia; Kry; Viviana; McCutcheon, Valerie; Mc; Petrus, Ana; Reichmann, Stefan; R; Saenen, Bregt; Schmidt, Birgit; Sch; den Eynden, Veerle; Vandendorpe, Ju; Wuttke, Ulrike; Yeomans, Joanne;

Practical Support for FAIR Data

An overview of how the FAIR Toolkit provides practical support for implementation of FAIR data management through numerous use cases from industry and relevant tools, training and change methods.

Practical Support FAIR Toolkit

The FAIR Toolkit is designed to provide support for management of the FAIR data life cycle as illustrated in Figure 1 below. It places emphasis on the practical aspects of FAIR data management through the leverage of existing resources that are most relevant to the needs of Life Science industry.

FAIR Cookbook

Created by researchers and data managers professionals, the FAIR Cookbook is an online resource for the Life Sciences with recipes that help you to make and keep data Findable, Accessible, Interoperable and Reusable (FAIR).

Turning FAIR into practice

The FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. However, the FAIR Principles are aspirational and generic. The FAIR Cookbook guides researchers and data stewards of the Life Science domain in their FAIRification journey; and also provides policy makers and trainers with practical examples to recommend in their guidance and use in their educational material.

- FOREWORD
- Introduction
- Ethical values of FAIR
- Glossary
- RECIPES
- Findability
- Accessibility
- Interoperability

HOW TO FAIR

How to FAIR

- What is FAIR
- Why FAIR
- How to FAIR
- About
- Quiz

A deep dive into FAIR data

This website will take you on a deep dive into the subject matter of FAIR research data. Over the course of

HOW TO FAIR

- What is FAIR
- Why FAIR
- How to FAIR
- About
- Quiz

- 18 min read Documentation
- 12 min read File formats
- 20 min read Metadata
- 10 min read Access to data
- 7 min read Persistent identifiers
- 5 min read Data licences

The WorldFAIR Project

<https://worldfair-project.eu/>

In the WorldFAIR project, CODATA (the Committee on Data of the International Science Council) and RDA (the Research Data Alliance) work with a set of 11 disciplinary and cross-disciplinary case studies to advance implementation of the FAIR principles and, in particular, to improve interoperability and reusability of digital research objects, including data. Particular attention is paid to the articulation of an interoperability framework for each case study and research domain



WorldFAIR

The WorldFAIR Project
Webinar Series:
Presenting Project
Outputs

Cultural Heritage (WP13)

Cultural Heritage Mapping Report: Practices and Policies supporting Cultural Heritage image sharing platforms

Outline of current practices guiding online digital image sharing by institutions charged with providing care and access to cultural memory, in order to identify how these practices may be adapted to promote and support the FAIR principles for data sharing.

[READ MORE](#)

Cultural Heritage image sharing recommendations report

This report builds on our understanding of what it means to support FAIR in the sharing of image data derived from GLAM collections. This report looks at previous efforts by the sector towards FAIR alignment and presents recommendations designed to be implemented and tested at the DRI that are also broadly applicable to the work of the GLAMs.

[READ MORE](#)

CON
WEBINAR

Geochemistry (WP5)

Formalisation of OneGeochemistry (D5.1)

WorldFAIR Geochemistry sets out to formalise the OneGeochemistry Initiative. With the exponential growth of data volumes and production, better coordination and collaboration is needed within the Earth and Planetary Science community producing geochemical data.

[READ MORE](#)

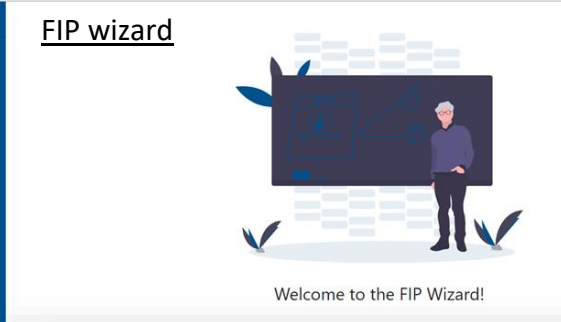
Geochemistry Scientific Content Component (Milestone)

This Milestone describes progress towards developing a methodology designed to assist in defining the individual FERs required to fully describe the minimum scientific and technical variables used to describe any geochemical analysis. It discusses progress towards minimum common variables of samples and how to make best practices for geochemical methods available online. It specifies a set of vocabularies published to describe methodologies.

[READ MORE](#)

Come fare / FAIR Implementation Profiles

- FIP Wizard
- Knowledge Models
- FIPs
- Create a FIP



International Conference on Conceptual Modeling
ER 2020: [Advances in Conceptual Modeling](#) pp 138-147 | [Cite as](#) 2020

Reusable FAIR Implementation Profiles as Accelerators of FAIR Convergence

Authors [Authors and affiliations](#)

Erik Schultes, Barbara Magagna, Kristina Maria Hettne, Robert Pergl, Marek Suchánek, Tobias Kuhn

- FIP Wizard
- Knowledge Models
- FIPs
- Create a FIP

Social Science Survey Research_V1

Questionnaire Metrics Preview Documents

View

Current Phase: Before Submitting the Proposal

I. Background: The FAIR Implementation Profile and FAIR Implementation Community

The FAIR Implementation Profile (FIP) is a collection of FAIR implementation choices made by a FAIR Implementation Community for each of the FAIR Principles. Community-specific FIPs are themselves captured as FAIR datasets and are made openly available to other communities for reuse. To create a FIP, the data steward of a community needs to fill out this questionnaire where the implementation choices are recorded as resources. The questionnaire is structured as follows: the first section is about the FAIR Implementation Community, which is then followed by a number of questions per FAIR principle. The answer to each of the questions should be a FAIR-Enabling Resource. The questionnaire offers to look up the resource in Nanobench. If the resource cannot be found in any of these applications, there is an option at the end of the questionnaire to register a FAIR-Enabling Resource as a nanopublication in Nanobench. The resource will get a PURL which

FAIR Implementation Profile

FAIR principle	Question	FAIR enabling resource types
F1	What globally unique, persistent, resolvable identifiers do you use for metadata records?	Identifier type
F1	What globally unique, persistent, resolvable identifiers do you use for datasets?	Identifier type
F2	Which metadata schemas do you use for findability?	Metadata schema
F3	What is the technology that links the persistent identifiers of your data to the metadata description?	Metadata-Data linking mechanism
F4	In which search engines are your metadata records indexed?	Search engines
F4	In which search engines are your datasets indexed?	Search engines
A1.1	Which standardized communication protocol do you use for metadata records?	Communication protocol
A1.1	Which standardized communication protocol do you use for datasets?	Communication protocol
A1.2	Which authentication & authorisation technique do you use for metadata records?	Authentication & authorisation technique
A1.2	Which authentication & authorisation technique do you use for datasets?	Authentication & authorisation technique
A2	Which metadata longevity plan do you use?	Metadata longevity
I1	Which knowledge representation languages (allowing machine interoperation) do you use for metadata records?	Knowledge representation language
I1	Which knowledge representation languages (allowing machine interoperation) do you use for datasets?	Knowledge representation language
I2	Which structured vocabularies do you use to annotate your metadata records?	Structured vocabularies
I2	Which structured vocabularies do you use to encode your datasets?	Structured vocabularies
I3	Which models, schema(s) do you use for your metadata records?	Metadata schema
I3	Which models, schema(s) do you use for your datasets?	Data schema
R1.1	Which usage license do you use for your metadata records?	Data usage license
R1.1	Which usage license do you use for your datasets?	Data usage license
R1.2	Which metadata schemas do you use for describing the provenance of your metadata records?	Provenance model
R1.2	Which metadata schemas do you use for describing the provenance of your datasets?	Provenance model

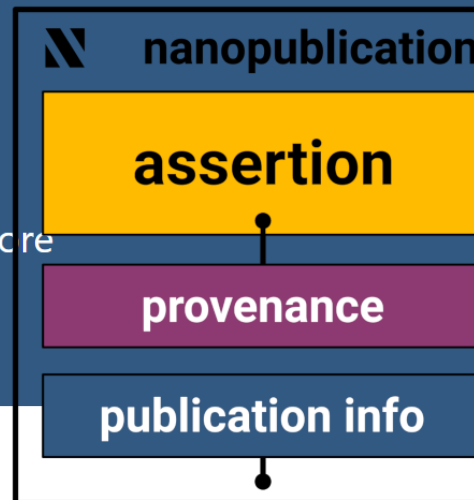
CREARE FAIR
 IMPLEMENTATION
 PROFILES
 RIUSABILI ANCHE
 DA ALTRI NELLA
 COMUNITÀ
 DISCIPLINARE
 KEYWORD:
CONVERGENZA

Slides courtesy of Erik Schultes 'Go FAIR OSF | HS.3PFF.Oct 2021.pdf'



Nanopublications

FAIR data containers for scientific results, and more



What is a Nanopublication?

A nanopublication is a small knowledge graph snippet with metadata that is treated as an independent (scientific) publication. This information in a nanopublication can be about anything, for example a relation between a gene and a disease or an opinion. Nanopublications are expressed in a knowledge graph format that is formal and machine-interpretable. With nanopublications, it is possible to disseminate individual data as independent publications without an accompanying research article. Furthermore, because nanopublications can be attributed and cited, they provide incentives for researchers to make their data available in standard formats that drive data accessibility and interoperability. Nanopublications have the following general structure:

1. **Assertion:** The assertion is the main content of a nanopublication in the form of an small atomic unit of information
2. **Provenance:** This part describes how the assertion above came to be. This can include the scientific methods that were used to generate the assertion, for example a reference to the kind of study that was performed and its parameters.
3. **Publication Info:** This part contains metadata about the nanopublication as a whole, such as when and by whom it was created and the license terms for its reuse.

Nanopublications are implemented in the language RDF and come with an evolving ecosystem of tools and systems. They can be published to a decentralized server network, for example, and then queried, accessed, reused, and linked. To find out more, see the guidelines, the examples, or the published articles about nanopublications.

Come fare / Nanopublications and FAIR connect



FAIR Connect is an Open Access publishing platform for the development and dissemination of good practices for professional FAIR-Data stewardship.

Try our new [Nanopublication Search engine](#)

See our new [Dashboard](#)

JOIN OUR COMMUNITY



<https://fairconnect.pro/>

FAIR Supporting Resources Types

FAIR Data Policy

FAIR Data Management Plan Template

FAIR Implementation Community

Data Steward Professional Profile

FAIR Data Stewardship Event

FAIR Training Material

FAIR Practice

FAIR Specification

FAIR Supporting Service

FAIR Connect Article

FAIR Enabling Resources

...ANCORA IN FASE
DI CREAZIONE

FINDABLE



F = Findable. Metadata



What is metadata?

While generally 'meta-data' is summarised as 'data about data', the following provides examples of what this actually means:

- Metadata is information about an object or resource that describes characteristics of that object, such as content, quality, format, location and access rights.
- Metadata can be used to describe physical objects (e.g. pot shards and specimens) as well as digital objects (e.g. documents, images, data sets and software)
- Metadata can take many different forms, from free text (e.g. a read-me file) to standardised, structured, machine-readable, extensible content.

- Metadata is analogous to any other form of data, in terms of how it is created, managed, linked and stored.
- Metadata is associated with the data it describes. It can be embedded within the data file, or recorded a separate text/ spreadsheet file that is linked to the collection of data files it describes, or contained in a catalogue record that points to the research data collection.
- Metadata enables and enhances the discovery and reuse of data.

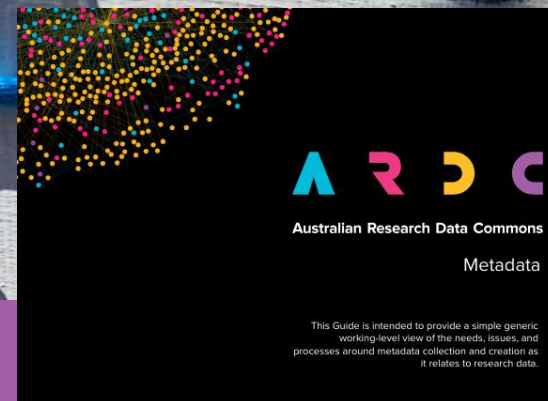
DATI CHE DESCRIVONO
UN OGGETTO
(FISICO O DIGITALE)

Metadata

F = Findable. Metadati

Levels of metadata

METADATI A DIVERSI
LIVELLI
(COLLEZIONE, ITEM,
SUB ITEM)



There are three levels of data groupings: data objects (**Items and Sub-items**) can be collected into groups (**Collections**):

1. **Sub-item level** - sometimes single objects can be made up of component parts: a book has chapters, a documentary has scenes, a relational database has tables, and so on. To search for a particular sub-item component requires not just the item-level metadata but also its sub-item or component level metadata.
2. **Item level** metadata describes individual objects.
3. **Collection level** metadata describes the collection as a whole. For example, Australian Urban Water Collection (<https://researchdata.ands.org.au/australian-urban-water-collection/267201>).

The collection approach and subsequent level of description impacts on discoverability, and cost and effort of management. Metadata ideally should be based on the needs of those for whom the collection is created.

For more information about collections metadata please read ARDC's Research Data Australia Collections policies (<https://documentation.ands.org.au/display/DOC/Research+Data+Australia+Collection+Policies>).

Metadata



F = Findable. M = Metadata.

Types of metadata

Metadata types are often grouped into functional types, but note that some elements will provide multiple functions.

The most common types are:

- **Descriptive metadata**

Information required for discovery and assessment of the collection,

- › e.g. title, contributors, subject or keywords, study description, and the location and dates of the study.

- **Provenance metadata**

This relates to the origins and processing of the data, and enables interpretation and reuse of the data. It ranges from the easily human readable to the highly technical, and usually requires some knowledge of the domain to create.

- **Preservation metadata**

This builds on the history from the Provenance, Rights and Technical metadata, and also includes information to allow the data to be managed for long-term accessibility.

- › e.g. Has there been any restructuring or other changes to the files, e.g. due

DIVERSI TIPI DI METADATI

- **Technical metadata**

Fundamental information for a computer application to read

- › e.g. How is the data set up? What formats, and versions of formats, are used? How is the database configured? How does it relate to other data?

- **Rights and access metadata**

Information to enable access, and licensing or usage rules.

- › e.g. How can someone access the data? Who is allowed to view or modify the data, or the metadata, and under what conditions? Who has some kind of authority over the data? Are there costs associated with access? Under what licence is the data being made



to migration to new file formats? What software has been used to access the data?

- **Citation metadata**

information required for someone to cite the data e.g. Creator(s), Publication Year, Title, Publisher, Identifier.

F Findable – metadati d



Data management stage	Metadata fields	Standardize public resources
Sample	Latitude, longitude, date/time, temperature, biome/ecosystem, depth and/or elevation of sampling site, etc.	Environmental Ontology (ENVO), Minimum Information about x Sequence (MIxS), International Geo/General Sample Number (IGSN)
Preparation	Laboratory protocol(s): DNA extraction, purification, amplification.	Protocols.io, e-laboratory notebook/management software
Data processing	Software tools for QA/QC, assembly, annotation. Include reference (if published), version, and parameters used.	Community guidelines for describing and citing software [23–25]
Feature	E.g., Annotations of sequence data, such as taxonomy or function	NCBI Taxonomy, Genome Taxonomy Database toolkit (GTDB-tk); Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), etc.
FAIR: Findable (i.e., PID metadata)	Data owner(s), organization, keywords	ORCID, Researcher Organization Registry (ROR); keyword selection [26] to enhance search engine optimization (SEO)
FAIR: Accessible	Usage license, privacy protocols, transfer protocols	Creative Commons, HTTP
FAIR: Interoperable	Type and size of data, file formats, etc.	.csv,.tsv, etc.
FAIR: Reusable	See data processing.	Workflow notebooks (e.g. [27])

<https://doi.org/10.1371/journal.pcbi.1010476.t001>

ESEMPIO DI METADATI MINIMI
SECONDO LE DIVERSE FASI DEL
CICLO DELLA RICERCA



F = Findable. Metadata standards

Metadata Standards Catalog

Search Sign in

Metadata standards catalog

Metadata Standards Catalog

The RDA Metadata Standards Catalog is a collaborative, open directory of metadata standards applicable to research data. It is offered to the international academic community to help address infrastructure challenges.

Index of subjects

Multidisciplinary

Science

Atmospheric sciences

Climatology

Meteorology

Biological sciences

Biochemistry

Biochemicals

Proteins

Metabolism

Biology

Physical sciences

Crystallography

Molecular physics

Nuclear physics

Plasma physics

Optics

Image formation

Physics

Scientific approach

Scientific methods

Space sciences

Astronomical systems

Solar system

Crystallography

Found 8 schemes.

×

CIF (Crystallographic Information Framework)

A well-established standard file structure for the archiving and distribution of crystallographic information, CIF is in regular use for reporting crystal structure determinations to Acta Crystallographica and other journals.

Sponsored by the International Union of Crystallography, the current standard dates from 1997. As of July 2011, a new version of the CIF standard is under consideration.

CSMD (Core Scientific Metadata Model)

A study-data oriented model, primarily in support of the ICAT data management infrastructure software. The CSMD is designed to support data collected within a large-scale facility's scientific discipline.

Index of metadata standards

ABCD (Access to Biological Collection Data)

ABCD Zoology

ABCDDNA

ABCDEFG (Access to Biological Collection Databases Extended for Geosciences)

HISPID (Herbarium Information Standards and Protocols for Interchange of Data)

AgMES (Agricultural Metadata Element Set)

AGRIS Application Profile

AVM (Astronomy Visualization Metadata)

Brain Imaging Data Structure (BIDS)

F = Findable - Met

The screenshot shows the OpenRefine website homepage. At the top, there is a navigation bar with links for 'OpenRefine', 'Download', 'Documentation', 'Community', 'Blog', and 'Donate'. The main heading is 'OpenRefine', followed by a description: 'OpenRefine is a powerful free, open source tool for working with messy data: cleaning it; transforming it from one format into another; and extending it with web services and external data.' Below this is a prominent 'Download' button. The page features six feature cards arranged in a 2x3 grid: 'Faceting' (drill through large datasets), 'Clustering' (fix inconsistencies), 'Reconciliation' (match datasets to external databases), 'Infinite undo/redo' (rewind to any previous state), 'Privacy' (data cleaned on your machine), and 'Wikibase' (contribute to Wikidata).

The screenshot shows the Frictionless Data website. The header includes the 'Frictionless Data' logo and navigation links for 'Introduction', 'Projects', 'Universe', 'Adoption', 'People', 'Follows', 'Development', and 'World'. The main heading is 'Data software and Openrefine'. Below the heading is a paragraph: 'Frictionless is an open-source toolkit that brings simplicity to the data experience - whether you're wrangling a CSV or engineering complex pipelines.' There are two buttons: 'Why Frictionless Data?' and 'Get Started'. An illustration of a toolbox is visible on the right side of the page.

How can I use Frictionless?

You can use Frictionless to describe your data (add metadata and schemas), validate your data, and transform your data. You can also write custom data standards based on the Frictionless specifications. For example, you can use Frictionless to:

- easily add metadata to your data before you publish it.
- quickly validate your data to check the data quality before you share it.
- build a declarative pipeline to clean and process data before analyzing it.

Usually, new users start by trying out the software. The software gives you an ability to work with Frictionless using command-line interfaces or programming languages.

As a new user you might not need to dive too deeply into the standards as our software encapsulates its concepts. On the other hand, once you feel comfortable with Frictionless Software you might start reading Frictionless Standards to get a better understanding of the things happening under the hood or to start creating your metadata descriptors more proficiently.

TOOL PER
- DESCRIVERE DATI
- VALIDARE
- IMPOSTARE PIPELINE

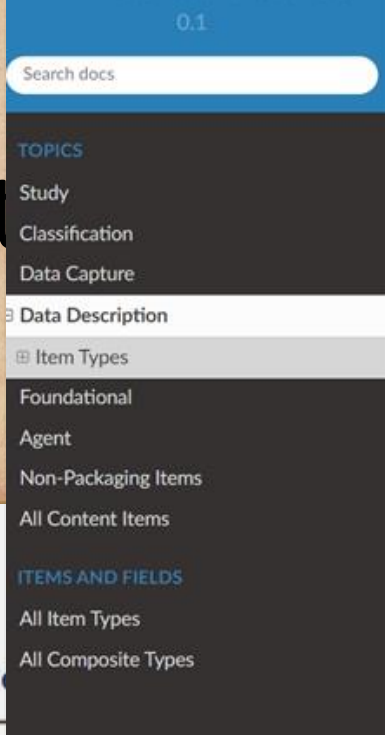
F = Findable - Met

STANDARD PER LE
SCIENZE SOCIALI



<https://ddialliance.org/>

Learn ▾ Products ▾ Membership ▾ Events ▾ Publicati



Data Description

Item Types

DDI has 13 item types related to Data Description

- ConceptualVariable
- RepresentedVariable
- Variable
- PhysicalInstance
- DataRelationship
- ManagedDateTimeRepresentation
- ManagedMissingValuesRepresentation
- ManagedNumericRepresentation
- ManagedScaleRepresentation
- ManagedTextRepresentation
- NCube
- VariableStatistics
- Weighting

Document, Discover and Interoperate

The Data Documentation Initiative (DDI) is an international standard for describing the data produced by surveys and other observational methods in the social, behavioral, economic, and health sciences. DDI is a free standard that can document and manage different stages in the research data lifecycle, such as conceptualization, collection, processing, distribution, discovery, and archiving. Documenting data with DDI facilitates understanding interpretation, and use -- by people, software systems, and computer networks. Use DDI to **D**ocument, **D**iscover, and **I**nteroperate!

Why Use DDI?

- ✓ Generate interactive codebooks
- ✓ Implement data catalogs
- ✓ Build question banks
- ✓ Create concordance mappings
- ✓ Harmonize and compare data
- ✓ Manage longitudinal data sets

Find Out More!



<https://ddialliance.org/>

Learn ▾ Products ▾ Membersh

What is DDI?

What does DDI stand for?

- Data Documentation Initiative

What is it?

- A free suite of standards/products originating from data professionals (within the behavioral, social and economic sciences) who realized
 - **Document** your data to make it more **Discoverable** and **Interoperable** between systems
 - Offer a framework to describe the research data produced and the processes used
 - Are platform-agnostic
 - Facilitate understanding, interpretation, and use by people, software systems, and computer networks
 - Have a great community of researchers who want their research data to be **FAIR** (Findable, Accessible, Interoperable, Reusable)
 - Offer lots of [free training materials](#) to support new and existing users
 - Allow for cross-domain usage (New! See below)

How many DDI standards are there?

- There are three different DDI standards:
 - [DDI Codebook \(DDI-C\)](#)
 - Intended to document simple or single survey data
 - Has the same information describing a dataset that
 - Used in the Dataverse data repository
 - Based on XML
 - [DDI Lifecycle \(DDI-L\)](#)
 - Can be used to document each stage of the lifecycle
 - Can be used for longitudinal datasets
 - Can describe more than one dataset
 - Based on XML
 - [DDI Cross-Domain Integration \(DDI-CDI\)](#)

Training Materials

These training materials were developed by the DDI Alliance Training Working Group, whose mission is to help researchers working with DDI. Only training slides approved by the DDI Alliance are made available here (and are free to use). For DDI training materials created by others, please see the Zenodo [DDI Training Group Community](#).

DDI Basics: What is Metadata?

This is a brief module introducing the basic concept of metadata and describing the different types of metadata. It is a good introduction.

External Link:
<https://doi.org/10.5281/zenodo.5180480>

DDI Basics: What is DDI?

This is a brief general introduction to the major work products (DDI Codebook, DDI Lifecycle, etc.) and their relationship to the result.

External Link:
<https://doi.org/10.5281/zenodo.5180509>

F = Findable DataCite metadata schema

DataCite Metadata Schema Documentation

Version 4.5

DataCite Metadata Working Group

DataCite metadata schema

Mar 12, 2024

Mandatory Properties

The mandatory properties must be supplied with any initial metadata submission to DataCite, together with their relevant sub-properties. **If one of the required properties is unavailable**, please use one of the standard (machine-recognizable) codes listed in [Appendix 3: Standard values for unknown information](#).

Table 1: DataCite Mandatory Properties

ID	Property	Obligation
1	Identifier	M
2	Creator	M
3	Title	M
4	Publisher	M
5	PublicationYear	M
10	ResourceType	M

Table 2: DataCite Recommended and Optional Properties

ID	Property	Obligation
6	Subject	R
7	Contributor	R
8	Date	R
9	Language	O
11	AlternateIdentifier	O
12	RelatedIdentifier	R
13	Size	O
14	Format	O
15	Version	O
16	Rights	O
17	Description	R
18	GeoLocation	R
19	FundingReference	O
20	RelatedItem	O



F = findable. Metadata tools

The CEDAR Workbench is an essential component of open science, ensuring FAIR data and enhancing scientific reproducibility

[Launch Workbench](#)

CEDAR is making data submission smarter and faster, so that scientific researchers and analysts can create and use better metadata. Through better interfaces, terminology, metadata practices, and analytics, CEDAR improves metadata from provider to end user.

940
users

51,875
resources

120,829
metadata

CEDAR

Better metadata means better science

[WATCH VIDEO](#)

[LEARN MORE](#)

What CEDAR does...

CREATE FORMS

Create user-friendly, shareable forms for collecting metadata, with nested form elements.

COMPOSE LIBRARIES

Build reusable form components that can be shared and published in libraries.

COLLABORATE

Collaborate with other users and groups of users to build forms and fill in forms with metadata.

PUBLISH FORMS

Publish your forms for others to use to enter metadata.

SUBMIT METADATA

Submit data and metadata to NCBI repositories directly.

VERSION CONTENT

Version your templates to maintain history.



Findable — Metadata creation

FAIR cookbook

FAIR Cookbook

- FAIR Cookbook
- Introduction
- Assessing FAIR
- Infrastructure for FAIR
- Improving Findability
- Improving Accessibility
- Improving Interoperability
- How to interlink data from different sources?
- Identifier mapping with BridgeDB
- Which vocabulary to use?
- Requesting terms addition to terminology artefacts

Creating a Metadata Profile

<p>Recipe metadata</p> <p>identifier: <i>RX.X</i> version: <i>v1.0</i></p>	<p>Difficulty level</p> <p>🔥🔥🔥</p>	<p>Reading Time</p> <p>🕒 20 minutes</p> <p>Recipe Type</p> <p>🖨 Hands-on</p> <p>Executable Code</p> <p>▶ Yes</p>	<p>Intended Audience</p> <p>👤 Principal Investigator</p> <p>🗄 Data Manager</p> <p>🔧 Data Scientist</p>
---	---	---	---

How to generate a metadata template

The following steps are intended as a starting point to guide the generation of a metadata template.

Step 1: Define competency questions

- What are the questions you would like to address with the template? Without a set of a competency questions, important variables may easily be forgotten. It is equally possible to collect too much metadata, making the resulting metadata model opaque and difficult to navigate. Competency questions serve as a guide to identify the most relevant experimental factors.

Step 2: Define a Minimal Set Of Metadata (MSOM) according to these questions

- Compile metadata from different sources
- Generate consolidated view on metadata by merging attributes as far as possible
- Differentiate metadata available for most of the studies from metadata occurring rarely (sparse matrix)
- Identify gaps in the metadata available for most of the studies comprising data that is considered important but has not been captured in the past
- Define a MSOM to be captured in the future from the metadata that is available for most of the studies and the metadata considered to be important
- Identify available community standards regarding minimal sets of metadata
- Add metadata attributes from those community standards to the MSOM, if they are not yet included
- Assign cardinality to the MSOM (identify mandatory metadata and how many times the attributes may be reported. Some metadata might not be mandatory but are still important to capture, if available)
- Identify appropriate ontologies representing your data and establish an application ontology (see recipe 4 of UC3)
- Assign, as far as possible, ontologies to the MSOM and the sparse matrix

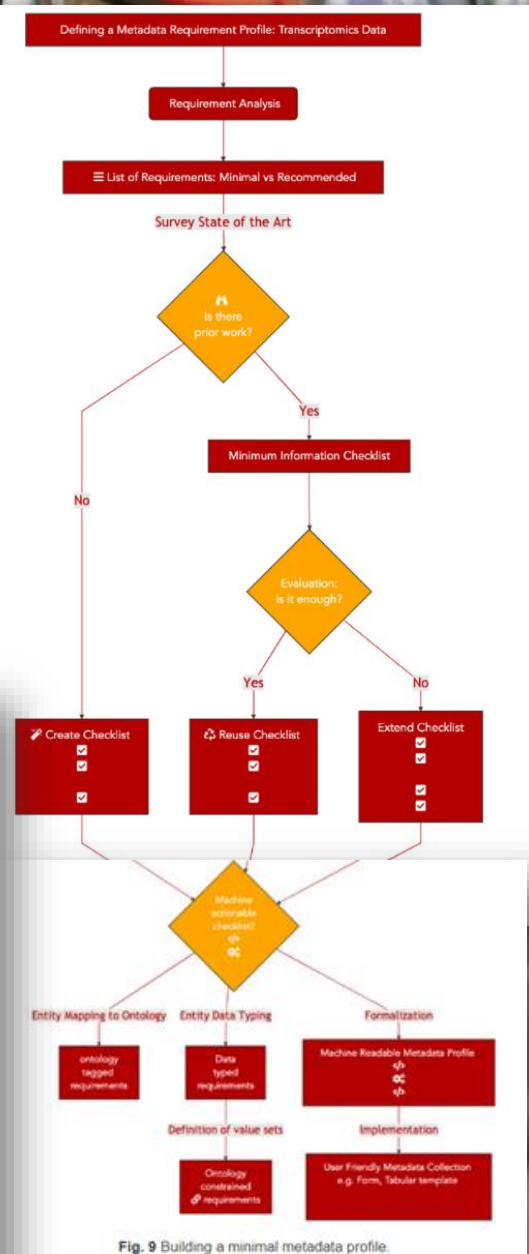
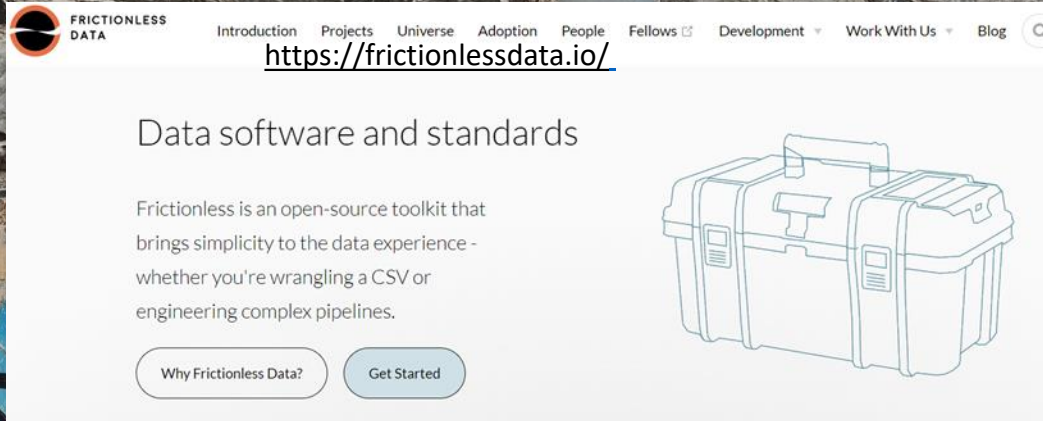


Fig. 9 Building a minimal metadata profile.

F = Findable. Metadata tools



The screenshot shows the homepage of Frictionless Data. At the top left is the logo, a black circle with a white 'F' and the text 'FRICTIONLESS DATA'. To the right is a navigation menu with links: Introduction, Projects, Universe, Adoption, People, Fellows, Development, Work With Us, and Blog. Below the navigation is the URL <https://frictionlessdata.io/>. The main heading is 'Data software and standards'. Below this is a paragraph: 'Frictionless is an open-source toolkit that brings simplicity to the data experience - whether you're wrangling a CSV or engineering complex pipelines.' To the right of the text is a line-art illustration of a toolbox. At the bottom are two buttons: 'Why Frictionless Data?' and 'Get Started'.

- TOOL PER
- DESCRIVERE I DATI
 - VALIDARLI
 - CREARE PIPELINE

How can I use Frictionless?

You can use Frictionless to describe your data (add metadata and schemas), validate your data, and transform your data. You can also write custom data standards based on the Frictionless specifications. For example, you can use Frictionless to:

- easily add metadata to your data before you publish it.
- quickly validate your data to check the data quality before you share it.
- build a declarative pipeline to clean and process data before analyzing it.

Usually, new users start by trying out the software. The software gives you an ability to work with Frictionless using visual interfaces or programming languages.

As a new user you might not need to dive too deeply into the standards as our software encapsulates its concepts. On the other hand, once you feel comfortable with Frictionless Software you might start reading Frictionless Standards to get a better understanding of the things happening under the hood or to start creating your metadata descriptors more proficiently.

Metadati e dati:



RO-crate

RO-Crate 1.1

Background

Community

Examples

Outreach and Publications

Profiles

RO-Crate In Use

Specification

Tools

Search Research Object Crate

Research Object Crate (RO-Crate)

Permalink: <https://w3id.org/ro/crate>

TABLE OF CONTENTS

- 1 What is RO-Crate?
- 2 Where did RO-Crate come from?
- 3 Who is it for?
- 4 When can I use it?
- 5 How can I use it?
- 6 RO-Crate in use
- 7 Current
- 8 Contribute to RO-Crate community
- 9 Cite RO-Crate
 - a Cite RO-Crate as project/approach
 - b Cite RO-Crate Specification (any version)
 - c Other citations



RO-Crate

Other citations

News: RO-Crate Metadata specification 1.1 released

What is RO-Crate?

RO-Crate is a community effort to establish a lightweight approach to packaging research data with their metadata. It is based on schema.org annotations in JSON-LD, and aims to make best-practice in formal metadata description accessible and practical for use in a wider variety of situations, from an individual researcher working with a folder of data, to large data-intensive computational research environments.

Where did RO-Crate come from?

RO-Crate is the marriage of [Research Objects](#) with [DataCrate](#). It aims to build on their respective strengths, but also to draw on lessons learned from those projects and similar research data packaging efforts. For more details, see [background](#).

Who is it for?

The RO-Crate effort brings together practitioners from very different backgrounds, and with different motivations and use-cases. Among our core target users are: a) researchers engaged with computation and data-intensive, workflow-driven analysis; b) digital repository managers and infrastructure providers; c) individual researchers looking for a straight-forward tool or how-to guide to "FAIRify" their data; d)

STRUMENTO PER
«IMPACCHETTARE» DATI E
METADATI E RENDERE I DATI
FAIR

Findable - Metadata

Packaging Entities with Machine actionable Metadata



Enabling reproducible, transparent research.



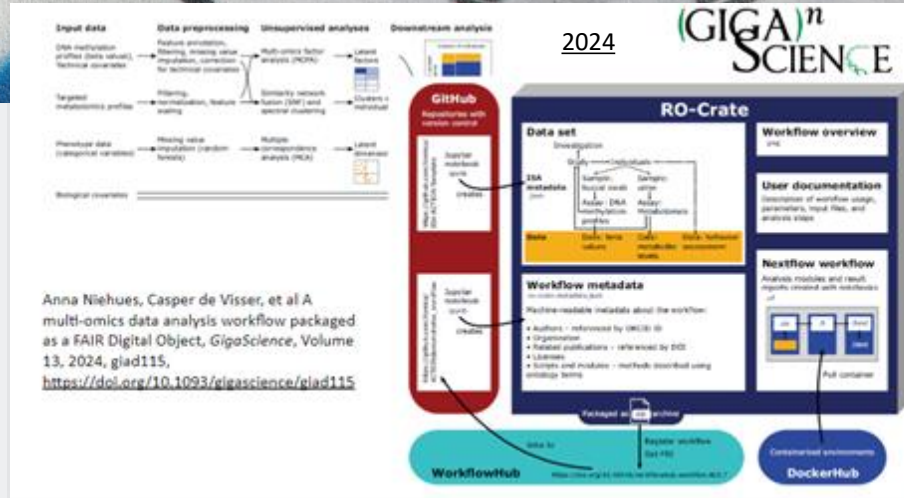
Lightweight, developer friendly, infrastructure independent, extensible packaging approach.

Uses off the shelf web methods.

Aggregate files and/or any URI-addressable content, with contextual information into a machine actionable, metadata rich, structured archive

Human readable, search engine accessible.

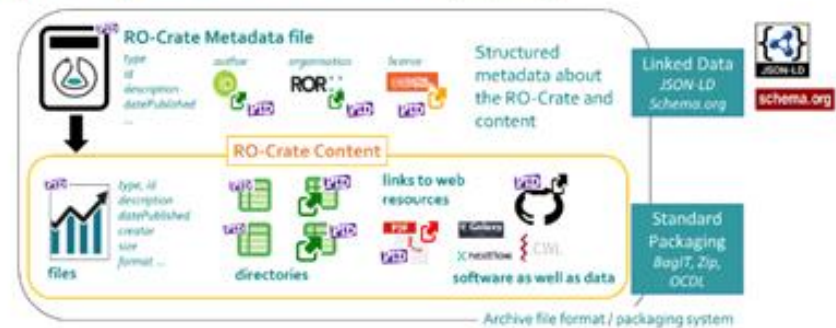
<http://www.researchobject.org/ro-crate/>



Anna Niehues, Casper de Visser, et al A multi-omics data analysis workflow packaged as a FAIR Digital Object, GigaScience, Volume 13, 2024, giad115, <https://doi.org/10.1093/gigascience/giad115>

Object Exchange & Archive Format

Structured self-describing, machine readable, metadata objects
Wraps data, metadata, software and references in single package



Packaging research artefacts with RO-Crate
Data Science <https://doi.org/10.3333/DS-230053>

RO-Crate Specification 1.1
<https://w3id.org/ro/crate/1.1>

RO-CRATE PER CREARE «PACCHETTI» DI RESEARCH OBJECTS CON MACHINE ACTIONABLE METADTA

«una» soluzione

Practical webby FDOs with RO-Crate and FAIR Signposting

2024

Experiences and lessons learned

Stian Soiland-Reyes^{1,2} , Peter Sefton³ , Simone Leo⁴ , Leyla Jael Castro⁵ ,
Claus Weiland⁶ , and Herbert Van de Sompel⁷ 

Abstract: Research Object Crate (RO-Crate) is a lightweight method to package research outputs along with their metadata. Signposting provides a simple yet powerful approach to navigate scholarly objects on the Web. Combining these technologies form a "webby" implementation of the FAIR Digital Object principles which is suitable for retrofitting to existing data infrastructures or even for ad-hoc research objects using regular Web hosting platforms. Here we give an update of recent community development and adoption of RO-Crate and Signposting. It is notable that programmatic access and more detailed profiles have received high attention, as well as several FDO implementations that use RO-Crate.

F = Findable. Identificativo
persistente



<https://ror.org/>

ABOUT SCOPE FACTS SUPPORTERS RESOURCES

Open Funder Registry (OFR)

<https://www.crossref.org/services/funder-registry/>

Home > Find a service > Open Funder Registry (OFR)

The Open Funder Registry (OFR, formerly FundRef) and associated funding metadata allows transparency into research funding and its outcomes. It's an open and unique registry of providing organizations around the world.



About us Services Res

Welcome to the Research Organization Registry Community

ROR is a community-led project to develop an open, sustainable, usable, and unique identifier for every research organization in the world.

WELCOME TO DATA CITE

with the leading global pro

Learn more

ORCID
Connecting Research and Researchers

FOR RESEARCHERS FOR ORGANIZATIONS ABOUT HELP SIGN IN

SIGN IN REGISTER FOR AN ORCID ID LEARN MORE

6,055,250 ORCID IDs and counting. See more

We need your feedback! Please tell us about your understanding and perceptions of ORCID and your experience of using your ID by completing our community survey. Thank you!

DISTINGUISH YOURSELF IN THREE EASY STEPS

ORCID provides a persistent digital identifier that distinguishes you from every other researcher and, through integration in key research workflows such as manuscript and grant submission, supports automated linkages between you and your professional activities ensuring that your work is recognized. Find out more

- 1 REGISTER** Get your unique ORCID Identifier Register now!
Registration takes 30 seconds.
- 2 ADD YOUR INFO** Enhance your ORCID record with your professional information and link to your other identifiers (such as Scopus or ResearcherID or LinkedIn).

<https://orcid.org/>

LATEST NEWS
Tue, 26 Feb 2019
Construyendo una Infraestructura para Apoyar a los Investigadores - Una entrevista

-PER LE COSE:
ASSEGNATE DOI
DIGITAL OBJECT
IDENTIFIER

- PER GLI AUTORI: USATE
ORCID

- ENTI: ROR

- FUNDERS: OFR



Search our registry to find datasets, software, images, and other research material.



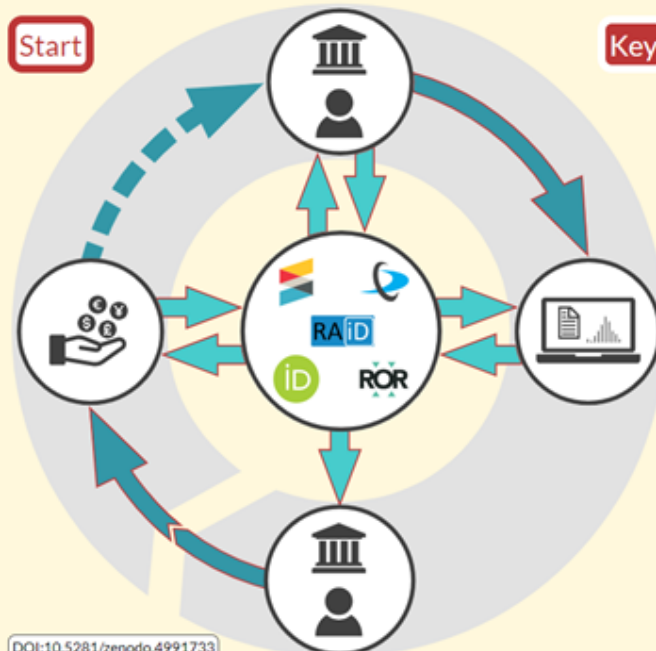
Find an appropriate repository to access and deposit research data with re3data.org

<https://www.datacite.org/>

Il ciclo della ricerca migliora con i PID

MOREBRAINS

The PID-optimised research cycle



DOI:10.5281/zenodo.4991733

Actors and entities

- = Funding organisations
- = Research-performing organisations
- = Research contributors
- = Research output platforms/publishers (articles, data, algorithms, etc)

Priority PIDs and registries

- = Crossref
- = Datacite
- = ORCID
- = Digital Object Identifier (DOI)
- = Research Activity Identifier (RAiD)
- = Research Organisation Registry (ROR)

<https://doi.org/10.5281/zenodo.4991733>

FAIR vero o falso / Findable

Findable means...

...that the data can be discovered by both humans and machines, e.g.

- by exposing meaningful machine-actionable metadata and keywords to search engines and research data catalogues.
- by providing to data unique and persistent identifiers (e.g. DOIs or Handles)
- by including the DOI in the dataset's metadata.

Valentina Pasquale 2024

Your data are not findable if...

... you backup them in a portable drive that you keep in a closet

... you don't assign them a unique persistent digital object identifier ("I'll make them available in my personal website")

... you don't provide in your articles the DOI of supporting datasets ("Data will be available upon request to the corresponding author")

... you name your files in an unintelligible way and you don't provide good searchable keywords to describe their content

ACCESSIBLE



A = Accessible

ACCESSIBLE ≠ OPEN
«ACCESSO» PUÒ ANCHE ESSERE
RISERVATO O SOTTO EMBARGO

- **Open access**

Data that can be accessed by any user whether they are registered or not.
Data in this category should not contain personal information unless consent is given (see '[Informed consent](#)').

- **Access for registered users (safeguarded)**

Data that is accessible only to users who have registered with the archive.
This data contains no direct identifiers but there may be a risk of disclosure through the linking of indirect identifiers.

- **Restricted access**

Access is limited and can only be granted upon request. This access category is for the most sensitive data that may contain disclosive information.

Restricted access requires the long-term commitment of the researcher or person responsible for the data to handle the upcoming permission requests.

- **Embargo**

Besides offering the opportunity for restricted access 'for eternity' most data repositories allow you to place a temporary embargo on your data. During the embargo period, only the description of the dataset is published. The data themselves will become available in open access after a certain period of time.

A = Accessible

Repositories



Upload

Communities

Featured communities

Why use Zenodo?

- **Safe** – your research is stored safely for the future in CERN's Data Centre for as long as CERN exists.
- **Trusted** – built and operated by CERN and OpenAIRE to ensure that everyone can join in Open Science.
- **Citeable** – every upload is assigned a Digital Object Identifier (DOI), to make them citable and trackable.
- **No waiting time** – Uploads are made available online as soon as you hit publish, and your DOI is registered within seconds.
- **Open or closed** – Share e.g. anonymized clinical trial data with only medical professionals via our restricted access mode.
- **Versioning** – Easily update your dataset with our versioning feature.
- **GitHub integration** – Easily preserve your GitHub repository in Zenodo.
- **Usage statistics** – All uploads display standards compliant usage statistics

The Dataverse Project

Open source research data repository software

Researchers Enjoy full control over your data. Receive web visibility, academic credit, and increased citation counts. A personal Dataverse collection is easy to set up, allows you to display your data on your personal website, can be branded uniquely as your research discoverable to the research community, and satisfies data management policy for your personal Dataverse collection?

Journals Seamlessly manage the submission, review, and publication of articles. Establish an unbreakable link between articles in your journal and the open data movement by using a Dataverse policy or list of repository recommendations. Want to find a journal that supports open data collections?

Institutions Establish a research data management solution for your organization. Dataverse repositories worldwide for increased discoverability and visibility in the drive to set norms for sharing, preserving, citing, and archiving research data. Want to install a Dataverse repository?

Developers Participate in a vibrant and growing community that is helping to develop, preserve, cite, explore, and analyze research data. Contribute to the development, documentation, testing, and/or standards. Integrate research data management tools, or other research and data archival systems with the Dataverse ecosystem.

DRYAD **DRYAD** Explore Data | About | Help | Login

for your research data

Submit Now

figshare **Figshare** Search on figshare ... Log in

store, share, discover **research**

get more citations for all of the outputs of your academic research over 30,000 citations of figshare content to date

ALSO FOR INSTITUTIONS & PUBLISHERS

Our curators will check through your submission to ensure the data are usable.

Review

Cite

Cite and promote your data publication!

POSSIBILE CREARE «COMUNITÀ»

The place to share your research

OSF is a free, open platform to support your research and enable collaboration.

Get started

Discover public research

Discover projects, data, materials, and collaborators on OSF that might be helpful to your own research.

Search discipline, author...

AMBIENTE INTEGRATO CONTIENE DATI, CODICE, PROTOCOLLI E PERMETTE PUBBLICAZIONE PREPRINT [SCEGLIERE SERVER AMBURGO PER GDPR]

Search and Discover Find papers, data, and materials to inspire your next research project. Search public projects to build on the work of others and find new collaborators.

Design Your Study Start a project and add collaborators, giving them access to protocols and other research materials. Built-in version control tracks the evolution of your study.

Collect and Analyze Data Store data, code, and other materials in OSF Storage, or connect your Dropbox or other third-party account. Every file gets a unique, persistent URL for citing and sharing.

Publish Your Reports Share papers in OSF Preprints or a community-based preprint provider, so others can find and cite your work. Track impact with metrics like downloads and view counts.

Open Science Framework

A = Accessible FAIR enabling r



Online storage, sharing and publishing of research data

Search for datasets in DataverseNL

Search

<https://dataverse.nl/>



DataverseNL is a publicly accessible data repository platform, open to researchers of affiliated institutes and their collaborators to deposit and share research data openly with anyone. It facilitates making your research data FAIR (Findable, Accessible, Interoperable, Reusable).



DataverseNL supports the creation of custom terms of use and restrictions in order to control access to your research data. DataverseNL facilitates long-term access, persistent identifiers, and preservation by storing a backup copy for safekeeping.



Receive academic credit and recognition by making your data more discoverable to the research community online. Collaborate in teams and track changes as Dataverse provides increased user control over managing changes to a project.

What are the benefits of using DataverseNL for sharing your research data? ▾

- Your dataset receives a DOI (a persistent identifier) and it will be easy for others to refer to your data with the provided citation information.
- You can manage access and reuse of your own data.
- DataverseNL provides safe storage for your data.
- Sharing your data increases the impact and visibility of your research.
- You can link your dataset to the related publications.
- You can meet grant requirements by depositing your research data in DataverseNL.
- You can update your stored dataset during your research, and keep track of your changes with version control.
- By using the Guestbook-feature you can check the use of your data by others. Dataverse also displays download statistics per dataset and per file.

What are the costs for depositing my data in DataverseNL? ▾

Whether storage costs are charged to the researcher can differ per institute. You can contact your local DataverseNL contact for more information. See this [list for contact information per institute](#).



For institutes it is possible to get a CoreTrustSeal certification for their dataverses within DataverseNL, as Tilburg University did. A connection to the DANS Data Vault secures Long Term Preservation of the

**DATAVERSE NL ABILITA LA
FAIRIFICAZIONE DEI DATI /
FEDERABILE**

A = Accessible – FAIR enabling repositories

DataverseNO: A National, Generic Repository and its Contribution to the Increased FAIRness of Data from the Long Tail of Research

Philipp Conzett

OTTIMO ESEMPIO DI REPOSITORY NAZIONALE «FAIR ENABLING»

Dataverse. This article presents the organization and operation of DataverseNO, and investigates how the repository contributes to the increased FAIRness of small and medium sized research data. Sections 1 to 3 present background information about the FAIR Data Principles (section 1), how FAIR may be turned into reality (section 2), and what these principles and recommendations imply for data from the so-called long tail of research, i.e. small and medium-sized datasets that are often heterogenous in nature and hard to standardize (section 3). Section 4 gives an overview of the key organizational features of DataverseNO, followed by an evaluation of how well DataverseNO and the repository application Dataverse as such support the FAIR Data Principles (section 5). Section 6 discusses how sustainable and trustworthy the repository is. The article is rounded up in section 7 by a brief summary including a look into the future of the repository.

Table 2: The implementation of Findability in Dataverse and its adoption in DataverseNO. Adapted from Crosas (2020).

Principle	Implementation in Dataverse	Applied in DataverseNO
R1		
R1.1	Included in metadata: data use license/waiver; data access and use terms. But, licenses other than CC0 are not predefined and by default not machine-readable.	Yes. Almost all datasets are published under default license CC0.
	By default no support for explicit information about metadata license	Terms for reuse of metadata described on website
R1.2	Rich citation metadata including information about data authors and other contributors, providers, distributors, related data (input data)	Yes
	Versions with changes documented automatically	Yes
	W3C PROV support	No
R1.3	DDI for social science data	Partially/in some datasets
	Metadata blocks for other community standards	Partially/in some datasets
	Ongoing work on support for more domains.	No
	Custom metadata	No
	FITS for astronomy data	N/A (so far)
	File format conversion to reusable formats (tabular)	Partially/in some datasets
		Data in preferred file formats
		Datasets include ReadMe file.

Principle	Implementation in Dataverse	Applied in DataverseNO
F1	Support for DOI and Handle	Yes (DOI)
	Always at the dataset level	Yes
	Optionally at file level	Yes
F2	Metadata standards in human- and machine-readable formats: Dublin Core; Documentation Data Initiative (DDI); DataCite; Schema.org	Yes
	Optional custom metadata	No
F3	Dataset PID is part of metadata record presented on Dataset landing page.	Yes
	File PID is part of metadata record presented on File landing page.	Yes
	PIDs are included in exported metadata files.	Yes
F4	DataCite metadata is harvested and indexed by DataCite Search.	Yes. In addition: B2FIND and VLO.
	Schema.org metadata is indexed by Google Dataset Search.	Yes

A = Accessible. Cercate un archivio?






REGISTRY OF RESEARCH DATA REPOSITORIES

Checklist: is it the right repository for your data?

The checklist that follows addresses the five key questions posed in this guide:

1. is the repository reputable?
2. will it take the data you want to deposit?
3. will it be safe in legal terms?
4. will the repository sustain the data value?
5. will it support analysis and track data usage?

[DCC checklist](#)



2,000 Data Repositories and Science Europe's Framework for Discipline-specific Research Data Management

By offering detailed information on more than 2,000 research data repositories, re3data has become the most comprehensive source of reference for research data infrastructures globally. Through the development and advocacy of a framework for discipline...

[Read more](#)

Three new DOI Fabrica features to simplify account management

Last month we launched DOI Fabrica, the modernized version of the DataCite Metadata Store (MDS) web frontend. It is the one place for DataCite providers and their clients to create, find, connect and track every single DOI from their organization...

[Read more](#)

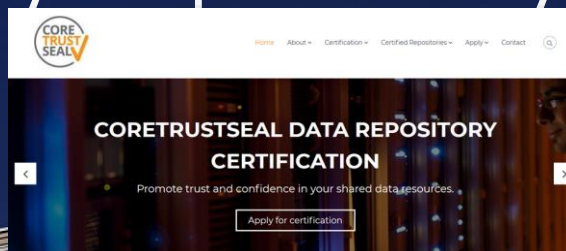
One step closer towards instant DOI search results

Art Art? You might be wondering, what this pink and green picture illustrates? A few months ago we couldn't show you this picture; the data that we used to create it, did not exist. And the answer to what this illustrates – this is simply a distorted...

[Read more](#)

Criteria per la selezione di un «trustworthy repository»

HORIZON EUROPE HA UNA SUA DEFINIZIONE



TRUSTWORTHY REPOSITORIES

Trustworthy repositories should meet the following minimum criteria:

- 1. Provision of Persistent and Unique Identifiers (PIDs)**
 - a. Allow data discovery and identification
 - b. Enable searching, citing, and retrieval of data
 - c. Provide support for data versioning
- 2. Metadata**
 - a. Enable finding of data
 - b. Enable referencing to related relevant information, such as other data and publications
 - c. Provide information that is publicly available and maintained, even for non-published, protected, retracted, or deleted data
 - d. Use metadata standards that are broadly accepted (by the scientific community)
 - e. Ensure that metadata are machine-retrievable
- 3. Data access and usage licences**
 - a. Enable access to data under well-specified conditions
 - b. Ensure data authenticity and integrity
 - c. Enable retrieval of data
 - d. Provide information about licensing and permissions (in ideally machine-readable form)
 - e. Ensure confidentiality and respect rights of data subjects and creators
- 4. Preservation**
 - a. Ensure persistence of metadata and data
 - b. Be transparent about mission, scope, preservation policies, and plans (including governance, financial sustainability, retention period, and continuity plan)

A = Accessible. Data repositories

Where to keep research data

Domain/discipline-specific data repository, data centre or 'scientific database'

- Pros - most likely to offer both the specialist domain knowledge and data management expertise needed to ensure your data collection is properly kept and used
- Cons - most likely to be selective, requiring advance planning of the effort needed to meet high standards for metadata and documentation

General-purpose data repository: e.g. Dryad, Figshare, Zenodo

- Pros - most likely to offer useful search, navigation and visualisation functionality
- Cons - requires scrutiny of terms and conditions to ensure consistency with your funder, journal or institution's policies on cost recovery, copyright/IP, long-term preservation

Institutional data repository

- Pros - most likely to accept any data of value, especially if no suitable home can be found for it elsewhere, and to ensure that policy requirements for long-term access are met
- Cons - unlikely to be as well-resourced as either general-purpose or domain repositories

Journal supplementary material service

- Pros - most likely to comply with the journal or publisher's requirements
- Cons - may be costly, unlikely to offer a data repository's functionality or long-term solution

Departmental, project or personal web page

- Pros - might provide functionality tailored to your data collection and/or your existing data users and peer network
- Cons - least likely to make your data collection visible to new users and contacts, or to sustain long-term access to your data collection

CONDIVISIONE

CONSERVAZIONE

Institutional data archive or vault

- Pros - most likely to have considered the total costs of long-term storage, and to ensure that policy requirements for long-term access are met
- Cons - may be less likely to offer the same ease of use as third-party storage or archiving services

Safe centres or havens

- Pros - most likely to meet stringent security requirements for handling sensitive data, and to ensure that legal requirements for data protection are met
- Cons - may be less likely to offer similar levels of digital preservation as a data archiving third-party service or institutional data archive

Cloud storage third-party services

- Pros - most likely to offer easy to use file store and share functionality
- Cons - long-term reliability and costs of data retrieval may be unpredictable; terms and conditions need careful scrutiny to ensure it complies with policy requirements for long-term access and other legal requirements, e.g. a data centre location within the European Union

Data archiving third-party services

- Pros - likely to offer cost-effective long-term storage with guarantee of accessibility, including data that may not be shareable for confidentiality reasons
- Cons - less likely to offer administrative interface to manage access and preservation policies (although some services offer

A = Accessible, Data journals

Title	URL	Charge	Notes for authors (N.B. we suggest checking in particular for policy on submission of data already published)	Publisher	Notes on Subject Area
Journal of Open Archaeology Data	http://openarchaeologydata.metajni.com/		http://openarchaeologydata.metajni.com/about/submissions	Ubiquity Press	Archaeology
Open Health Data	http://openhealthdata.metajni.com/		http://openhealthdata.metajni.com/about/submissions#authorGuidelines	Ubiquity Press	Public Health
Journal of Open Behavioural Data	http://openpsychologydata.metajni.com/		http://openpsychologydata.metajni.com/about/submissions#onlineSubmissions	Ubiquity Press	Psychology

[UCL Home](#) » / [Open@UCL Blog](#) » / [Data journals and data reports – don't miss out](#)

Data journals and data reports – don't miss out on this useful publishing format!

Aug. 2021

By Kirsty, on 17 August 2021

Guest post by [James Houghton](#) – [Research Data Support Officer](#)

Why not publish a data report article?

Publishing with a data journal offers several benefits. First, a data report article is more formal than a publication of data files in a repository and is a peer reviewed publication which then contributes to a researcher's publication record which is important for CVs and advancement for many. Second, they allow a more detailed explanation of a dataset and any analysis or code related to it than is usually otherwise possible. Third, the appearance of an article in a recognised journal can help to drive visibility of a dataset for other researchers. In practice it may often be the case that a repository will be used to host material which is discussed at length in a paper.

[nature.com/sdata/for-authors](https://www.nature.com/sdata/for-authors)

[nature.com/sdata/for-authors#data-deposition](https://www.nature.com/sdata/for-authors#data-deposition)

Dataset Description

Object Name

- *walkers* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for records made by individual walkers during stage-one fieldwalking.
- *counts* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for potsherds counted during stage-one fieldwalking.
- *pottery* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for the main pottery database, assembled various artefact specialists.
- *petrography* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for those sherds sampled for thin section petrography.
- *lithics* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for the main lithics database.
- *other* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for the main database of all non-ceramic and non-lithic finds.
- *struts* – three files providing the data, metadata and field type definitions (.csv, .txt, .csvt respectively) for the main database of all standing remains, except for terraces.
- *coast* – a vector polygon dataset (.shp and associated files) with the shape of Antikythera's coastline.
- *geology* – a vector polygon dataset (.shp and associated files) with the main bedrock units on Antikythera.
- *tracts* – a vector polygon dataset (.shp and associated files) with the main stage-one survey units.
- *grids* – a vector polygon dataset (.shp and associated files) with the main stage-two survey units.
- *terraces* – vector line dataset (.shp and associated files) with all observable agricultural terraces (i.e. the location

Data journals

Panayiota Polydorotou

- *other*
- *assis*
- *geol*

Alexander Technological Educational Institute of Thessaloniki

Repository

UK Ar
10.5284

*European Commission Workshop
Alternative Open Access Publishing Models: Exploring New Territories in
Communication
Brussels, 12 October 2015*

Publication

05/02/2012

Language

English (a Greek language summary of the project methods and results can be found at www.ucl.ac.uk/asp/ or www.tuarc.trentu.ca/asp/).

License

Creative Commons CC-BY 3.0

Reuse Potential

Due to their unusual coverage of an entire landscape, these datasets would provide a good basis for developing a tutorial on survey, GIS and/or spatial analysis in archaeology. They also lend themselves to the comparative analysis of evidence from other intensive Mediterranean surveys that are in the public domain (e.g. <http://dx.doi.org/10.5284/1000271>, <http://dx.doi.org/10.5284/1000208>, <http://dx.doi.org/10.5284/1000103> and, to a lesser extent, also <http://dx.doi.org/10.5284/1000351>), albeit with due attention to the fact that the intensive methods used are not identical. The ASP data is particularly reusable because artefact locations, dates and identifications are recorded individually in the database rather than in aggregate. The standing structures and terraces from Antikythera are also the kinds



Type DANS format

• Preferred format(s)

• Non-preferred format(s)

Text documents

- PDF/A (.pdf)

- ODT (.odt)
- MS Word (.doc, .docx)
- RTF (.rtf)
- PDF (.pdf)

Plain text

- Unicode text (.txt)

- Non-Unicode text (.txt)

Markup language

- XML (.xml)
- HTML (.html)
- Related files: .css, .xslt, .js, .es

- SGML (.sgml)

Spreadsheets

- ODS (.ods)
- CSV (.csv)

- MS Excel (.xls, .xlsx)
- PDF/A (.pdf)
- OOXML (.docx, .docm)

Databases

- SQL (.sql)
- SIARD (.siard)
- DB tables (.csv)

- MS Access (.mdb, .accdb) (v. 2000 or later)
- dBase (.dbf)
- HDF5 (.hdf5, .he5, .h5)

Statistical data

- SPSS Portable (.por)
- SPSS (.sav)
- STATA (.dta)
- DDI (.xml)

- SAS (.7dat; .sd2; .tpt)
- R (* under examination)



Appendix A: Tables of File Formats

Quick Links

Computer Aided Design	Digital Audio	Digital Moving Images
Digital Cinema	Digital Video	Digital Still Images
Digital Photographs	Scanned Text	Digital Posters
Geospatial Formats	Presentation Formats	Textual Data
Structured Data Formats	Email	Web Records
Calendars	Navigational Charts	

Geospatial Formats		
Preferred Formats	Format Versions	Format Specifications
Geospatial Tagged Image File Format	1.8.2	Geo TIFF Format Specification: (https://geotiff.maptools.org/spec/geotiffhome.html)
Geographic Markup Language	2.0 through 3.2	ISO 19136:2007 & Version 3.2, OGC document 07-036: (http://www.opengispatial.org/standards/gml)
Topologically Integrated Geographic Encoding and Referencing Files	2006 Second Edition	2006 Second Edition TIGER/Line®: (https://www.census.gov/programs-surveys/geography/technical-documentation/complete-technical-documentation.html)
Keyhole Markup Language	2.2	Open Geospatial Consortium Inc. OGC 07-147r2: (http://www.opengispatial.org/standards/kml)
Acceptable Formats	Format Versions	Format Specifications
Vector Product Format		MIL-STD-2407: (http://earth-info.nga.mil/publications/specs/)

- DICOM (.dcm) (by mutual)

A = Accesible - Formati

12. Converting from proprietary to open format

FAIR cookbook converting

Recipe Overview

Reading Time
20 minutes

Executable Code
Yes

Difficulty
🔥🔥🔥

Converting from proprietary to open format

Recipe Type
Hands-on

Audience
Principal Investigator, Data Manager, Data Scientist

Maturity Level & Indicator
DSM-3-R4

Cite me with FCB029

12.3. FAIRification Objectives, Inputs and Outputs

Actions.Objectives.Tasks	Input	Output
formatting	Waters MS format	mzML
text annotation	PSI-MS	annotated text

12.4. Table of Data Standards

Data Formats	Terminologies	Models
mzML	PSI-MS	

12.5. Ingredients

Tools and Software:

- github
- docker
- python

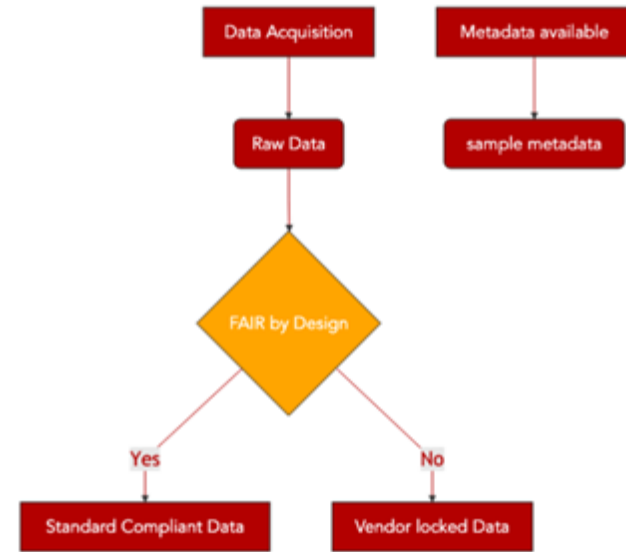


Fig. 12.1 Converting to an open standard file format.

INTEROPERABLE



DATI CON
SOFTWARE/MACCHINE

DATI CON ALTRI DATI

I = Interoperable

- Findability
- Accessibility
- Interoperability
 - 1. Registering SwissLipids identifiers in Wikidata
 - 2. Interlinking data from different sources
 - 3. Mapping Identifiers with BridgeDb
 - 4. Introducing terminologies and ontologies
 - 5. Selecting terminologies and ontologies
 - 6. Requesting new terms from terminologies and ontologies
 - 7. Introducing ontology-related tools and services
 - 8. Building an application ontology with ROBOT
 - 9. Mapping Ontologies with QxO, EBI Ontology Xref

Interoperability

This chapter is dedicated to FAIRification processes which focus on improving Interoperability.

Data usually need to be integrated with other data, and be interoperable with applications or workflows for analysis, storage, and processing.

Data objects can be interoperable only if:

- (Meta) data is machine-actionable.
- (Meta) data formats utilize shared vocabularies and/or ontologies.
- (Meta) data within the Data Object should thus be both syntactically parseable and semantically machine-accessible.

This chapter is dedicated to the **standards**, **tools**, **services** and other **resources** necessary to make data interoperable.

Browse existing recipes, but bear in mind that this is a **live resource**, and recipes are added and improved, iteratively, in an open manner.

If you want to contribute follow the **help** provided, or contact us at faircookbook-ed@elixir-europe.org.

<< 2. Downloading data with Aspera

1. Registering SwissLipids identifiers in Wikidata >>

I = Interoperable. Standards



PARTHENOS HOME TRAINING MODULES FOR TRAINERS FOR LEARNERS

WHAT ARE STANDARDS?

Even perfect metadata may not allow data to become interoperable if a different standard is used. A "standard" refers to a system that structures what types of information are captured in an item in a collection. In our .mp3 library system, a standard is expressed in the header categories such as 'name,' 'time,' 'artist,' and 'album' are listed, with every entry having the fields filled in. Standards are used to ensure that metadata is as useful as possible for organizing a collection, ensuring that common questions (how many songs are there on the album "Big") can be easily and accurately answered.

How Many Standards Are There and Who Decides Which One To Use?

Different standards have arisen in different kinds of cultural heritage institution: the most common standards in museums are different from those in archives, and those common in libraries are different again.

<http://training.parthenos-project.eu/sample-page/intro-to-ri/interoperability/what-are-standards/>



PARTHENOS HOME TRAINING MODULES FOR TRAINERS FOR LEARNERS

WHAT ARE KNOWLEDGE REPRESENTATION SYSTEMS AND 'ONTOLOGIES'?

In addition to metadata and standardised metadata schemas, research infrastructures can also use other forms of "knowledge representation system" to enhance the researcher's experience of the interoperable data they present. When we talk about 'Knowledge Representation Systems' in research infrastructures, we usually mean a specific category of hierarchical systems of terms known more commonly as an 'ontology'. Before the digital age, philosophers referred to an ontology as "the study of the kinds of things that exist". Ontologies are similar to taxonomies, another knowledge organisation framework you probably remember from early lessons in biology.



PARTHENOS TRAINING: Ontologies Explained (in 5 minutes ...)

Guarda più tardi Condividi

What is Metadata?

What are Standards?

What Are Knowledge Representation Systems and 'Ontologies'?

Sustainability

Methods and Tools

NETWORKS

I = Interoperable. Standards



» Prefer human and machine-readable systems: coding of data models and of the actual data that is both human and machine-readable in a unified way provides better sustainability and long-term accessibility than machine-readable only code (binary codes), that may use different formats for data model description and the actual data. For both, hierarchical data models and graph-based data, various serialisations (file formats) are available that fulfil this condition (XML, TEI/XML, Turtle, N3, RDF/XML), whereas SQL based technologies need bigger efforts.

» Normalise as much as possible: to avoid redundant information, the content of databases should be normalised as far as possible, using for example authority files like VIAF and identifiers like DOI, ARK, ISNI, GND and the like. To foster the exchange of data, standardised vocabularies and ontologies are needed as well, but an overall ontology for the humanities has not yet been established. The ontology CIDOC-CRM and especially some extensions are well on their way to become a reference model for cultural heritage information.

» Data models follow the data management plan (DMP): when establishing a data model, researchers should keep the whole lifecycle of their data in mind, as it should be outlined in a DMP. Therefore, an extensive documentation of the data model, its software and tools are highly relevant and facilitates the transfer of data in a secure and trusted repository in order to keep them accessible. The same is true here: the more you use open standards for your data model, the easier this task becomes.

NORMALIZZARE,
USARE AUTHORITY
FILES (es. VIAF)

- ✓ Easy tagging
- ✓ Independent from spelling
- ➔ Linked data and unique identifiers: Use **authority files** (Normdaten)!
- ➔ BARTOC (Basel Register of Thesauri, Ontologies & Classifications): <https://bartoc.org/>
 - **GND** (German National Library)
 - **ORCID** (Open Researcher and Contributor ID)
 - **GeoNames**
 - **Wikidata**
 - **Getty Union List of Artist Names**
 - **VIAF** (Virtual International Authority File)

I= Interoperable – Iavorare FAIR

Home About Download Support Development Publications Contacting Us Register



Download

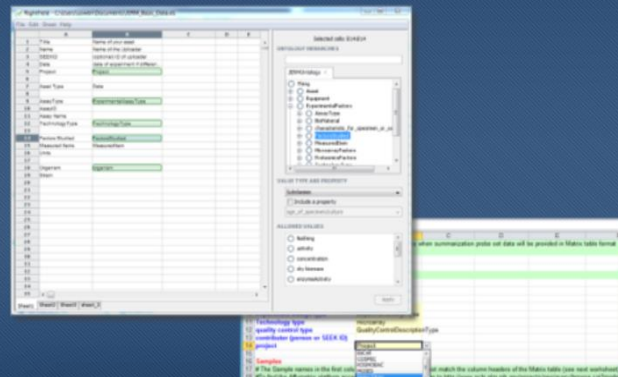
<https://rightfield.org.uk/>

RightField



Semantic data
annotation by
Stealth

[Find out more](#)



FAIRDOM

<https://fair-dom.org/>

[Twitter](#)

[GitHub](#)

[Search FAIRDOM](#)

- About
- FAIRDOM-SEEK
- FAIRDOMHub
- FAIRDOM In Use
- FAIRDOM Hubs
- Communities
- Support
- Outreach
- News
- Events
- Contact Us

A Consortium of Services for Research Data Management and More



Findable



Accessible



Interoperable



Reusable

• Collect, Manage, Store and Publish your data, models and operating procedures in a FAIR manner.

I = Interoperable - Ontologie

RAPPRESENTAZIONI
FORMALI IN CUI I
CONCETTI HANNO UNA
GERARCHIA

4.2. Controlled terminology or ontology: what's the difference?

Faircookbook Ontologies

A **controlled terminology** is a **normative** collection of terms, the spelling of which is fixed and for which additional information may be provided such as a **definition**, a set of **synonyms**, an **editor**, a **version**, as well as a **license** determining the condition of use. The set of information about a specific controlled terminology term is designated as **term metadata**. In a controlled terminology, terms appear as a **flat list**, meaning that no relationship between any of the entities the controlled terminology represents is captured in any formal way. This is the main drawback and limitation of **controlled terminologies**, which are often developed to support a data model or an application.

An **ontology** on the other hand, is a **formal representation of a domain knowledge where concepts are organized hierarchically**. The qualifier **formal** refers to a set of axioms and rules based on logic (e.g. **first order logic**) to structure, organize and check the consistency of the term hierarchy. As one can sense right away, ontologies are often a more sophisticated artefact, supported by more advanced theoretical frameworks and dedicated tools to develop them (e.g. Protégé, TopBraid Composer, OBO foundry INCAtools or Robot tool).

I = Inteoperable.

FORMAL ONTOLOGIES: A COMPLETE NOVICE'S GUIDE

[Formal ontologies](#)

Formal ontologies are proposed as a solution to the data heterogeneity problem because they describe broad ranges of data in a manner that is intellectually consistent and able to cover both general and particular levels of knowledge. Put another way, a formal ontology attempts to build a schema that applies to anyone and no one in a particular field.

A successful formal ontology must:

- accurately represent the most common information points of interest to a particular domain and the relationships that users want to trace between entities.
- offer sufficient abstract classes and relations in order to allow the representation of characteristic states of lack of knowledge. That is to say, create information structures which allow representing not just highly accurate information but also more general, uncertain information.

PARTHENOS HOME TRAINING MODULES FOR TRAINERS FOR LEARNERS

THE DATA HETEROGENEITY PROBLEM

[Data heterogeneity](#)

however, they reach their technical limits for large scale data integration and another solution is needed: a formal ontology.

Thesauri and Authority Files – why do we need them?

Before turning to formal ontologies, however, it is useful to quickly point to the role of **thesauri** and **authority files** in the process of standardization at the data level. Regardless of the standardization method chosen for the schema level, data integration is only fully achieved when harmonization is carried out also on the data value level. Enabling such standardization are thesauri and authority files. These are curated lists of either **controlled terminologies** or **controlled references**.

Controlled thesauri are generally curated by a specific community and provide a list of terms and their (un)official spellings for those concepts that are recognized and used for describing some aspect of reality. A classic example is the [Getty Art and Architecture Thesaurus](#).

+ Thesaurus Exercise (click to expand)

- Basic Formal Ontology: originally used in modelling of medical data, provides a complete methodology for data modelling
- DOLCE: was constituted to aid in modelling common sense natural language
- CIDOC CRM: originally designed in the museological community to account for cultural heritage and e-sciences data

On the other hand, other ontologies are designed to address very specific needs, ignoring the general aim of interoperability in favour of a more focused problem level. Examples of such focussed ontologies include:

- FOAF: an ontology for tracking social relations
- SPAR: for organizing citation data, article structure and content
- NeMO: for tracking scholarly process

CIDOC CRM Class Declarations
E1 CRM Entity
E2 Temporal Entity
E3 Condition State
E4 Period
E5 Event
E6 Destruction
E7 Activity
E8 Acquisition
E9 Move
E10 Transfer of Custody
E11 Modification
E12 Production
E13 Attribute Assignment
E14 Condition Assessment
E15 Identifier Assignment
E16 Measurement
E17 Type Assignment
E18 Physical Thing
E19 Physical Object
E20 Biological Object
E21 Person
E22 Human-Made Object
E24 Physical Human-Made Thing
E25 Human-Made Feature
E26 Physical Feature
E27 Site

I = Interoperable. Ontologies

The Publishing Workflow Ontology (PWO)

IRI:

<http://purl.org/spar/pwo>

Version IRI:

<http://purl.org/spar/pwo> <https://sparontologies.github.io/pro/current/pro.html>

Date:

01/06

PRO, the Publishing Roles Ontology

Current version:

2.4

IRI:

<http://purl.org/spar/pro/>

Version IRI:

<http://purl.org/spar/pro/2017-09-04>

Date:

04/09/2017

Current version:

1.5.4

Previous version:

<https://svn.code.sf.net/p/semublishing/code>

Authors:

David Shotton
Silvio Peroni

Imported Ontologies:

<http://purl.org/spar/foaf> (visualise it with LODE)
<http://purl.org/spar/tvc> (visualise it with LODE)

Previous version:

<https://purl.org/spar/pwo>

Authors:

Aldo
David
Silvio

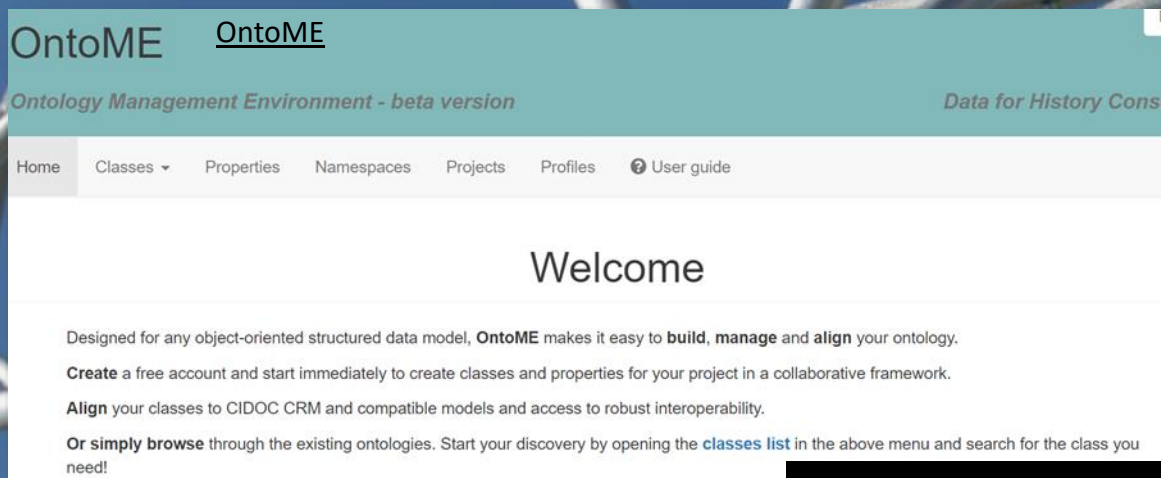
Imported Ontologies:

<http://purl.org/spar/foaf>
<http://purl.org/spar/tvc>
<http://purl.org/spar/foaf>
<http://purl.org/spar/tvc>

The screenshot shows the website <http://www.sparontologies.net/>. The navigation bar includes links for SPAR Ontologies, Ontologies, Examples, Publications, Uptake, Contacts, About, and News. The main content area is titled "Using" and contains the text: "Use the SPAR Ontologies for creating and publishing machine-readable RDF metadata concerning the publishing domain". There is a "See examples" button. Below this, there is a section titled "Describing the publishing domain" which states: "The Semantic Publishing and Referencing Ontologies, a.k.a. SPAR Ontologies, form a suite of orthogonal and complementary OWL 2 DL ontology modules for the creation of comprehensive machine-readable RDF metadata for every aspect of semantic publishing and referencing: document description, bibliographic resource identifiers, types of citation".

This work is distributed under a Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>).

I = Interoperable. Ontologies



The screenshot shows the OntoME website interface. At the top, there is a teal header with the text "OntoME" and a link to "OntoME". Below the header, it says "Ontology Management Environment - beta version" and "Data for History Conso". A navigation menu includes "Home", "Classes", "Properties", "Namespaces", "Projects", "Profiles", and "User guide". The main content area features a "Welcome" heading and three bullet points: "Designed for any object-oriented structured data model, **OntoME** makes it easy to **build, manage** and **align** your ontology.", "Create a free account and start immediately to create classes and properties for your project in a collaborative framework.", and "Align your classes to CIDOC CRM and compatible models and access to robust interoperability." A fourth point says "Or simply browse through the existing ontologies. Start your discovery by opening the **classes list** in the above menu and search for the class you need!"

ONTOME
ESEMPIO DI APPLICAZIONE
ALLE SCIENZE STORICHE



The screenshot shows an article page from IOS Press. The header includes the IOS Press logo, "IOS Press Content Library", and the year "2021". A navigation bar has "Home" and "Journals". The article title is "A challenge for historical research: Making data FAIR using a collaborative ontology management environment (OntoME)". The author is "Francesco Beretta" and the affiliation is "Laboratoire de recherche historique Rhône-Alpes, CNRS – Université de Lyon, 14 avenue Berthelot, 69363 Lyon cedex 07, France".

Abstract

This paper addresses the issue of interoperability of data generated by historical research and heritage institutions in order to make them re-usable for new research agendas according to the FAIR principles. After introducing the *symogih.org* project's ontology, it proposes a description of the essential aspects of the process of historical knowledge production. It then develops an epistemological and semantic analysis of conceptual data modelling applied to factual historical information, based on the foundational ontologies *Constructive Descriptions and Situations* and DOLCE, and discusses the reasons for adopting the CIDOC CRM as a core ontology for the field of historical research, but extending it with some relevant, missing high-level classes. Finally, it shows how collaborative data modelling carried out in the ontology management environment OntoME makes it possible to elaborate a communal fine-grained and adaptive ontology of the domain, provided an active research community engages in this process. With this in mind, the *Data for history* consortium was founded in 2017 and promotes the adoption of a shared conceptualization in the field of historical research.

I = Interoperable

1. Interlinking data from different sources

2. Identifier mapping with BridgeDb

3. Introduction to terminologies and ontologies

4. Selecting terminologies and ontologies

5. Requesting new terms

6. Ontology-related tools and services

7. Building an application ontology with ROBOT

8. Creating a data/variable dictionary

9. Creating a metadata profile

10. Converting from proprietary to open format

11. An inventory of tools for converting your data to RDF

12. File format validation,

1. Interlinking data from different sources

Recipe Overview

- Reading Time: 30 minutes
- Executable Code: No
- Difficulty: 4 stars

Interlinking data from different sources

- Recipe Type: Background information
- Audience: Principal Investigator, Data Manager, Data Scientist

Cite me with FCB016

1.1. Main Objectives

The FAIR principles, under Interoperability state that:

13. (Meta)data include qualified references to other (meta)data

[FAIR cookbook](#)

[FAIRsharing. Per essere interoperabili]

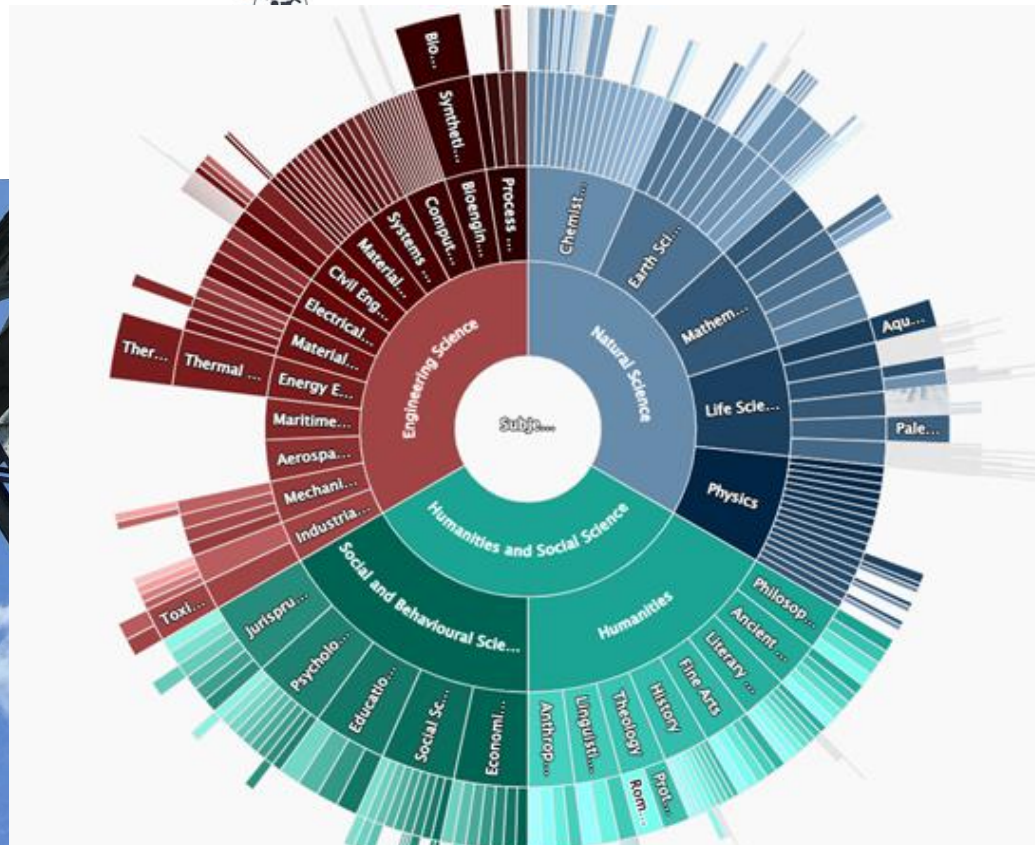
FAIRsharing.org standards, databases, policies search through all content STANDARDS DATABASES POLICIES COLLECTIONS ADD CONTENT STATS LOGIN

A curated, informative and educational resource on data and metadata standards, inter-related to databases and data policies.

We guide consumers to discover, select and use these resources with confidence, and producers to make their resource more discoverable, more widely adopted and cited.

RESEARCHERS DEVELOPERS & CURATORS JOURNAL PUBLISHERS LIBRARIANS & TRAINERS SOCIETIES & ALLIANCES FUNDERS

<https://fairsharing.org/>



FAIRSHARING
REGISTRO DI
STANDARD
SUDIDIVISO PER
DISCIPLINE

[FAIRsharing. Per essere interoperabile]

FAIRsharing.org
standards, databases, policies

search through all content

<https://fairsharing.org/educational>

STANDARDS

DATABASES

POLICIES

COLLECTIONS

ORGANISATIONS

ADD CONTENT

FACTSHEETS,
TUTORIALS...

Educational

Infographics and factsheets by the FAIRsharing Team and our Community Champions, covering content and functionalities for all stakeholders

FAIRsharing in a nutshell



doi [10.5281/zenodo.7737366](https://doi.org/10.5281/zenodo.7737366)

FAIRsharing content: standards overview

Definitions, types, benefits and need-to-know facts



doi [10.5281/zenodo.7737795](https://doi.org/10.5281/zenodo.7737795)

FAIRsharing content: databases overview

Definitions, types, benefits and need-to-know facts



doi [10.5281/zenodo.7737841](https://doi.org/10.5281/zenodo.7737841)

FAIRsharing content: policies overview

Definitions, types, benefits and need-to-know facts



Coming Soon

FAIRsharing content: collections overview

Coming Soon

FAIRsharing for you: researchers



doi [10.5281/zenodo.7733025](https://doi.org/10.5281/zenodo.7733025)

FAIRsharing for you: journal publishers



doi [10.5281/zenodo.7737417](https://doi.org/10.5281/zenodo.7737417)

FAIRsharing for you: funders



doi [10.5281/zenodo.7737551](https://doi.org/10.5281/zenodo.7737551)

Computer Science > Databases

[Submitted on 10 Jan 2023 (v1), last revised 11 Jul 2023 (this version, v4)]

Extending FAIR to FAIRer: Cognitive Interoperability and the Human Explorability of Data and Metadata 2023

Lars Vogt

Making data and metadata FAIR (Findable, Accessible, Interoperable, Reusable) has become an important objective in research and industry, and knowledge graphs and ontologies have been cornerstones in many going-FAIR strategies. In this process, however, human-actionability of data and metadata has been lost sight of. Here, in the first part, I discuss two issues exemplifying the lack of human-actionability in knowledge graphs and I suggest adding the Principle of human Explorability to extend FAIR to the FAIRer Guiding Principles. Moreover, in its interoperability framework and as part of its GoingFAIR strategy, the European Open Science Cloud initiative distinguishes between technical, semantic, organizational, and legal interoperability and I argue to add cognitive interoperability. In the second part, I provide a short introduction to semantic units and discuss how they increase the human explorability and cognitive interoperability of knowledge graphs. Semantic units structure a knowledge graph into identifiable and semantically meaningful subgraphs, each represented with its own resource that instantiates a corresponding semantic unit class. Three categories of semantic units can be distinguished: Statement units model individual propositions, compound units are semantically meaningful collections of semantic units, and question units model questions that translate into queries. I conclude with discussing how semantic units provide a framework for the development of innovative user interfaces that support exploring and accessing information in the graph by reducing its complexity to what currently interests the user, thereby significantly increasing the cognitive interoperability and thus human-actionability of knowledge graphs.

As atomic entities of the FAIR ecosystem, FAIR Digital Objects take a central role in the interoperability framework of EOSC (6) by providing, along with appropriate tools and controlled vocabularies such as ontologies, the metadata necessary to achieve interoperability. When embedded in a FAIR ecosystem, the interoperability framework aims to provide a general standard in science and industry and has the potential to substantially contribute to a solution for the reproducibility crisis in science (8) and the question of the trustworthiness of information in general (see also TRUST Principles of Transparency, Responsibility, User Focus, Sustainability, and Technology (9)).

All this, however, depends on the availability of readily and reliably usable tools for the technical realization of a FAIR ecosystem. In this context, ontologies and knowledge graphs (KGs), when employed while at the same time consistently applying adequate semantic data models, provide a promising technology for the development of a FAIR ecosystem (10–15), which is a main prerequisite for establishing technical and semantic interoperability. The graph-based abstractions used in KGs are superior to relational or other NoSQL models in terms of (i) their potential to provide an intuitive way for modelling relationships in a way that is similar to mind-maps with which most human users are familiar. (ii) the ability to postpone specifications of definitions for data schemata so that these can





Scholarly Knowledge. Comparable.

The Open Research Knowledge Graph (ORKG) aims to describe research papers in a structured manner. With the ORKG, papers are easier to find and compare. [▶ Play video](#)

☰ Browse by research field <https://orkg.org/>

Search for fields... ▾

Arts and Humanities

443 papers - 34 comparisons

Engineering

3438 papers - 357 comparisons

Life Sciences

4178 papers - 210 comparisons

Physical Sciences & Mathematics

15838 papers - 746 comparisons

Social and Behavioral Sciences

951 papers - 174 comparisons

Click to activate the tour guide

Got it



SciNEXT Training

Research is a fundamental pillar of societal progress. Yet, scientific communities face great difficulties in sharing their findings. With approximately 2.5 million newly published scientific articles per year, it is impossible to keep track of all relevant knowledge. Even in small fields, researchers often find themselves drowning in a publication flood, contributing to major scientific crises such as the reproducibility crisis, the deficiency of peer-review and ultimately the loss of knowledge.

The underlying problem is that we never updated our methods of scholarly communication to exploit the possibilities of digitalization. This is where the **O**pen **R**esearch **K**nowledge **G**raph comes into play!

The ORKG makes scientific knowledge human- and machine-actionable and thus enables completely new ways of machine assistance. This will help researchers find relevant contributions to their field and create state-of-the-art comparisons and reviews. With the ORKG, scientists can explore knowledge in entirely new ways and share results even across different disciplines.

The Open Research Knowledge Graph (ORKG)



00:00 / 01:50



1x

TIB




[un esempio]

Welcome to Polifonia

Playing the soundtrack of our history

From the soundscape of Italian historical bells, to the influence of French operas on traditional Dutch music, European cultural heritage hides a goldmine of unknown encounters, influences and practices that can transport us to experience the past, understand the music we love, and imagine the soundtrack of our future. Polifonia will organise this journey by using ten pilot use cases as drivers, addressing preservation, management, study, and interaction with musical heritage.



Discovery similar tunes across different traditional European music corpuses with the 'Patterns User Interface'

The Polifonia project formally ended on April 30, which means that the tools and software developed within this 4-year-project are released and ready for use. Today we look at 'Patterns UI'.

Technology

<https://polifonia-project.eu/>

- Semantic Web technologies and good practises to fill the FAIR gap in music data (knowledge graphs)
- Novel Data Science methods to extract information on music patterns and music object intrinsic features to enrich Musical Heritage knowledge graphs
- Linguistic corpora on Musical Heritage covering multilingual and multidisciplinary texts as well as texts from different centuries
- Knowledge extraction techniques from text to extract structured data on cultural, social, and historical aspects of Musical Heritage, including tangible and intangible heritage assets, temporal and spatial entities
- Innovative Human-Machine Interaction ways to study and experience Musical Heritage

FAIR vero o falso / Interoperable

Interoperable means...

...that the data can be exchanged and used across different applications and systems — also in the future, for example, by using open file formats (see recommended formats).

It also means that the data can be integrated with other data from the same research field or data from other research fields.

This is made possible by using metadata standards, standard ontologies, and controlled vocabularies as well as meaningful links between the data and related digital research objects.

Your data are not interoperable if...

... you share them in formats that can be only read by using proprietary software

... you do not include any documentation about variables' content and units of measurement

... you do not follow any standard for metadata and/or controlled vocabulary in your datasets' documentation

... you do not include in your datasets links to the corresponding publication, or deposited protocol, or any other related "object" (including your ORCID or funding grant ID)

REUSABLE



R = Reusable. Documentazione

DOCUMENTAZIONE (README FILE) PER
- EVITARE USO SCORRETTO/CATTIVE
INTERPRETAZIONI DEI VOSTRI DATI
- MANTENERE INTEGRITÀ



Project-level documentation



Project-level documentation explains the aims of the study, what the research questions/hypotheses are, what methodologies were being used, what instruments and measures were being used, etc. In the accordion the questions which your project-level documentation should answer are stated in more

detail:

- ⊕ 1. For what purpose was data created
- ⊕ 2. What does the dataset contain
- ⊕ 3. How was data collected
- ⊕ 4. Who collected the data and when
- ⊕ 5. How was the data processed
- ⊕ 6. What possible manipulations were done to the data
- ⊕ 7. What were the quality assurance procedures
- ⊕ 8. How can data be accessed

Data-level documentation

Data-level or object-level documentation provides information at the level of individual objects such as pictures or interview transcripts or variables in a database. You can embed data-level information in data files. For example, in interviews, it is best to write down the contextual and descriptive information about each interview at the beginning of each file. And for quantitative data variable and value names can be embedded within the data file itself.



⊖ Quantitative data

Variable-level annotation should be embedded within a data file itself. If you need to compile an extensive variable level documentation that can be created by using a structured metadata format.



Data-level documentation for quantitative data

For quantitative data document the following:

- **Information about the data file**
Data type, file type and format, size, data processing scripts.
- **Information about the variables in the file**
The names, labels and descriptions of variables, their values, a description of derived

R = Reusable. Documentazione

✓ protocols.io

Make your science more reproducible
protocols.io is the #1 open access repository for science methods

The screenshot shows the RStudio interface with a protocol editor. The protocol title is 'Fixation of yeast cells for RNA-FISH'. It includes a description, guidelines, materials, and a list of steps. Step 1: 'Around 10am, start a cell culture in a 50ml tube... Grow for 8-10 hours in a shaker at 30 °C.' Step 2: 'Measure OD in the evening and dilute into 250ml.' Step 3: 'Transfer to 50ml falcon tubes. Add 5ml of Formaldehyde. Invert a few times...'

What is an Open Notebook?

Open Notebooks are documents that contain equations, visualisations, narrative text and live code that can be executed independently and interactively, with output visible immediately beneath the input.

They bring together analysis descriptions and results, which can be executed to perform the data analysis in real time.

The screenshot shows the Jupyter Notebook interface with a code editor and a plot of two overlapping circles. The text 'The Jupyter Notebook' is visible at the top right.

Notebook web application

The screenshot shows the LabArchives website header with the logo and navigation menu: Products, Enterprise, Integrations.

Lab notebook for Humanities

Using an ELN for Humanities

by Olivia Sisson | May 21, 2020 | Data Management, Research | 0 comments

The screenshot shows a notebook entry titled 'The Age of Steel' with a video player showing a character from Star Wars. The interface includes a search bar, user profile, and various notebook management options.

web application enables users to:

the browser, with automatic syntax highlighting, indentation, and tab introspection.

in the browser, with the results of computations attached to the code which they run.

is of computations with rich media representations, such as HTML, LaTeX, and images.

se interactive JavaScript widgets, which bind interactive user interface controls to reactive kernel side computations.

ive text using the Markdown markup language.

ematical equations using LaTeX syntax in Markdown, which are rendered in the browser using MathJax.

The screenshot shows a Jupyter Notebook with Python code for plotting a waveform. The code includes imports for numpy and matplotlib, and a function to plot a Gaussian waveform. The output shows a plot of the waveform with a peak at 1.0 and a time window of 3e-07 s.

...WHY NOT?

- PROTOCOLS.IO PER I PROTOCOLLI
- OPEN LAB NOTEBOOK TIENE TRACCA DI TUTTO IL LAVORO (DIFFICILE LA PRIMA VOLTA, POI...)

R= Reusable. License

Copyright: protects the STRUCTURE, selection or arrangement of their contents" (Art. 3) NOT THE DATA

Sui generis database right: protects the «substantial effort» in OBTAINING data [NOT «CREATING»]... the right owner often is the institution

Database=a collection of independent works, data or other materials arranged in a systematic or methodical way (Art.1)



KEEP CALM AND

RICORDA: NESSUN COPYRIGHT SUI DATI (NON CREATIVI)

DIRECTIVE 96/9/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 11 March 1996

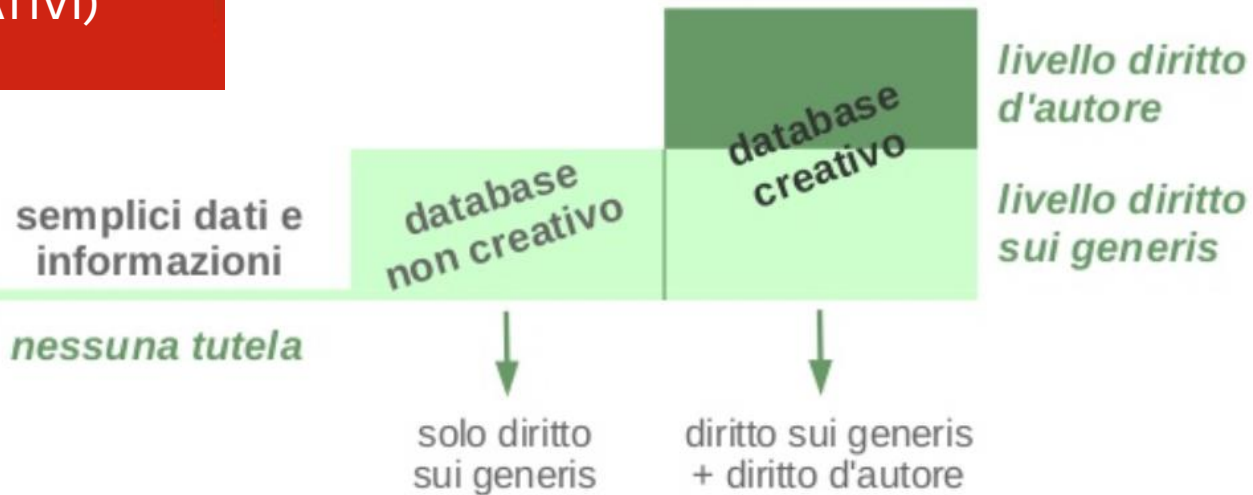
on the legal protection of databases

COUNCIL OF THE EUROPEAN UNION, in the Community, and in particular Article 57 (2), 66 and 100a thereof,

Simone Aliprandi

2014

la QUALI DIRITTI SUI DATI?



R = Reusable – Aspetti legali

1. THE PROTECTION OF DATA, DATA SETS AND DATABASES

European Union (EU) law defines “databases”, but not data sets or, at least for copyright purposes, data. Databases that meet the legal definition^① can be protected by copyright if they are original. Data sets, if they correspond to the definition of database, are protected by copyright otherwise not. Data as such are normally excluded from copyright protection [2,3]. It is important to understand that copyright protects original expressions in the “literary and artistic” domain^②, an expression that has historically included works such as books, musical works, choreographies, cinematographic works, drawings, etc [4]. Ideas, procedures, methods of operation or mathematical concepts as such, news of the day and miscellaneous facts are excluded from copyright protection [4,5,6].



MIT Press Direct



2020

Data Intelligence

Volume 2, Issue 1-2

Winter-Spring 2020



< Previous Article Next Article >

Article Contents

Abstract

1. THE PROTECTION OF DATA, DATA SETS AND DATABASES

2. SUITABLE OPTIONS FOR LICENSING DATA AND DATABASE RIGHTS

January 01 2020

Licensing FAIR Data for Reuse

Ignasi Labastida, Thomas Margoni

> Author and Article Information

Data Intelligence (2020) 2 (1-2): 199–207.

https://doi.org/10.1162/dint_a_00042



Cite



PDF



Permissions



Share

Abstract

The last letter of the FAIR acronym stands for Reusability. Data and metadata should be made available with a clear and accessible usage license. But, what are the choices? How can researchers share data and allow reusability? Are all the licenses available for sharing content suitable for data? Data can be covered by different layers of copyright protection making the relationship between data and copyright particularly complex. Some research



R = Reusable - Aspetti legali

Legal Compliance

Guidelines for Researchers: a Checklist

Phase1
Research Proposal

Phase2
Research Implementation

Phase3
Research review

Check whether there is background information, data and intellectual property rights brought into the project. More specifically

Clarify who brings what

Identify the member state law applicable to each background material. Pay particular attention to the territorial applicability of each right

Make sure to secure clearance for each use by

- Obtaining any authorisation needed for re-use and protection of derived data and other intellectual outputs,
- Agree on rules of ownership attribution on derived data and other intellectual outputs.

Aim at avoiding secrecy and at allowing re-use

Define Clearly

The ownership and/or co-ownership of each research output stemming from

- The use and re-use of pre-existing background information, data and IPRs,
- Single or joint research activities within the framework of the project,
- Single or joint research activities partially within OR outside the framework of the project, if building or depending on project activities.

R = Reusable

Hum

DEPOSIT, PRESERVE and SHARE

Legal Aspects

Data generated or collected in the e-humanities may potentially be subject to copyright in whole or in part. In addition, particular challenges may arise when some of the items within a data set themselves are subject to third party rights.

In practice, digital humanists can make good use of some checklists to determine whether and how data relevant for their research are subject to legal regulation. Some important questions that have to be solved concern topics like:

- Which national legislation applies to other researchers' work I use in my project?
- Do I have the right to collect, preserve and provide access to the data of my project?
- Is there sensitive information that could connect to some privacy issues?
- Are there risks of exposing the identity of human participants in my study?
- Am I allowed to digitally reproduce material and (re-)publish it in a digital reproduction?



RECOMMENDATIONS

- » Clarify all legal issues at the beginning of your research project and include the findings of this process in the data management plan.
- » Use checklists adequate to your research topic/discipline.
- » Check the resources indicated by DARIAH, CLARIN. (see further reading).
- » In the case of personal data ensure that only relevant people can access the data and that these are clearly identified (see GDPR).
- » Ask for consent to share anonymised data and establish transparent and well-documented anonymisation routines that consider not just direct identifiers, but also how a combination of indirect identifiers could reveal identities. (See for example the guide on informed consent in the CESSDA data management expert guide).
- » Avoid collection of (sensitive and non-sensitive) personal data when possible.



Reusable / licenses



RECOMMENDATIONS

ALLEA Report p. 26

Researchers are “prosumers” who produce and consume information and knowledge of other researchers. This section focuses on their role of producing knowledge and on ways to foster its diffusion by clear legal boundaries. In the humanities, texts are quite often closely intertwined with underlying data, which form an indispensable part of digital publications. Traditional conceptions of copyright like “All Rights Reserved” raise obvious problems for data sharing in the context of publications. In general: if machine readable data is to be shared, the recipient, in order to use the data effectively, will most likely need to make a local copy for analysis, or for merging with other data sets, or to extract some subset of the data. For this reason, our recommendation is to avoid applying any legal restrictions that do not embrace the principle of openness. The Reusability FAIR

» Proper entitlement: first of all, identify who owns the data, i.e. whether you are entitled to license your work. You may only attribute a licence to a work of which you are the copyright holder. If there are co-authors, you have to agree with them on the licence. Furthermore, you are not allowed to license the works of the public domain. You should also be aware of whether there are any licensing requirements from the funding organisation or the data repository.

» Determine the necessary and sufficient level of access restrictions. Some data cannot be shared openly but can still be shared under certain restrictions while at the same time protecting the data. See for instance the [CLARIN licensing framework](#) for language data or the [CESSDA access categories](#) for qualitative and quantitative data (interviews, survey data etc).

» Use free and standardised licences: In order to benefit from the possibility of sharing data since the digital turn and to foster Open Science, use a licence as free as possible. The Open Knowledge Foundation and the [Open Access Scholarly Publishers Association](#) only

» For editors of journals and repositories managers: Avoid applying more restrictive licences like NC (non-commercial) or ND (no derivatives) just to be ‘on the safe side’. NC can produce unintendedly limiting side-effects to potential re-users, as it is not quite clear whether the setting of a re-used work has commercial aspects or not. ND originates from the creative sector and is thought of as an instrument to protect the integrity of a work of art, such as a music composition. Many humanities scholars also want to protect their works from misuse and therefore are in favour of a ND licence. However, the risk of misuse through derivatives in the humanities is often quite low, so one has to balance this potential risk against the potentially unintended constraints imposed by ND, such as restrictions against reuse of publications in text and data mining procedures. Keep in mind that anybody deliberately deriving original content and thoughts by other scholars with misleading intention violates ethical scientific behaviour, whether a work is put under and ND licence or not.

R = Reusable - Aspetti legali



OpenAIRE How do I know SERVICES SUPPORT

Guides for Researchers

How do I know if my research data is protected?

Learn more about what is research data and their protection by intellectual property rights



OpenAIRE SERVICES SUPPORT

How do I license

Guides for Researchers


How do I license my research data?

Learn more about licenses for research data and how to apply it

- WHAT IS RESEARCH DATA?
- PROTECTION OF RESEARCH DATA
- SUI GENERIS DATABASE RIGHT (SGDR)
- COPYRIGHT
- TRAINING MATERIALS

What is Research Data?

Research data are the evidence that underpins the answer to the research question, and can be used to validate findings regardless of its form (e.g. print, digital, or physical). These might be quantitative information or qualitative statements collected by researchers in the course of their work by experimentation, observation, modelling, interview or other methods, or information derived from existing evidence. Data may be raw or primary (e.g. direct from measurement or collection) or derived from primary data for subsequent analysis or interpretation (e.g. cleaned up or as an extract from a larger data set), or derived from existing sources where the rights may be held by others. Data may be defined as 'relational' or 'functional' components of research, thus signalling that their identification and value lies in whether and how researchers use them as evidence for claims. They may include, for example, statistics, collections of digital images, sound recordings, transcripts of interviews, survey data and fieldwork observations with appropriate annotations, an interpretation, an artwork, archives, found objects, published texts or a manuscript.



OpenAIRE SERVICES SUPPORT

Licenses for Research Data

LICENSES FOR RESEARCH DATA

HOW TO APPLY LICENSES FOR RESEARCH DATA

SPECIFICATIONS OF LICENSING RESEARCH DATA

TRAINING MATERIALS

What licence should be applied to the research data?

It depends on what rights protect your research data, if at all. In the light of what is explained in the guide "[How do I know if my research data is protected?](#)":

- o If your research data qualifies as a work (literary work such as a journal article or a software), then CC BY 4.0 is usually the best choice. The use of the Share Alike (SA) is also compatible with the Open Access definition and reinforced in Plan S licensing guidance for publications. Non-commercial should be avoided as it is not Open Access compliant. Non-derivative is a tricky issue and should be avoided, especially if you do not know what you are doing. That said, it may not be incompatible with the Open Access definition.
- o If your research data is a database or a dataset (unstructured data that do not meet the database definition) usually the best option is a CC0, which waives all your rights in the database.

Keep in mind that CC licences only deal with copyright and copyright related matter. Personal data are not included in CC and are analysed separately.

What is a Creative Commons licence?



OpenAIRE Can I use SERVICES SUPPORT

Guides for Researchers

Can I reuse someone else's research data?

Learn more on how to reuse research data

- How can a protected dataset be used? +
- Where are licences found? +
- Interoperability and stacking +
- What happens if I use 'Share Alike' (SA) licensed material in my work? Does that mean I have to make my work available under the same SA licence? +
- Can a dataset be used if there is no licence? +
- What are the risks of using a dataset without a licence? +
- Training materials +

Creative Commons

CC Factsheet  creative commons UK

FACT SHEET ON CREATIVE COMMONS & OPEN SCIENCE v0.1

This information guide contains questions and responses to common concerns surrounding open science and the implications of licensing data under Creative Commons licences. It is intended to aid researchers, teachers, librarians, administrators and many others using and encountering Creative Commons licences in their work.

LICENZA CC0:
LEGALMENTE LA PIÙ
CORRETTA

What is Open Science?

[Open Science](#) is the movement to make scientific research and data accessible to all for knowledge dissemination and public reuse.

How should I licence my data for the purposes of Open Science?

We recommend you use the [CC0 Public Domain Dedication](#), which is first and foremost a waiver, but [can act as a licence](#) when a waiver is not possible.

CC ZERO LICENCE, 'NO RIGHTS RESERVED' LOGO



By applying CC0 to your data you enable everyone to freely reuse your data as they see fit by waiving (giving up) your copyright and related rights in that data.

You should keep in mind that there are many situations in which data is **not** protected as a matter of law. Such data can include facts, names, numbers – things that are considered 'non-original' and part of the public domain thus not subject to copyright protections. Similarly, your database (which is a structured collection of data) might be considered 'non-original' and thus ineligible for copyright, and it might additionally be excluded

from other forms of protection (like the [EU sui generis database right](#), also known as the 'SGDR', for non-original databases).

In these cases, using a Creative Commons licence such as a CC BY could signal to users that you claim a copyright in the non-original data despite the law, and perhaps despite your real intention.

Finally, if your data is in the public domain worldwide, you might state simply and obviously on the material that no restrictions attach to the reuse of your data and apply a [Public Domain Mark](#).

PUBLIC DOMAIN MARK LOGO



When in doubt, consider which use may be appropriate according to the chart below:

CC0 & PUBLIC DOMAIN LICENCES WHICH LICENSE TO USE AND WHEN



'Creative arrangement' of data is original, but any copyright has been waived and content is made available copyright-free



'Creative arrangement' of data is not original; the author acknowledges this and communicates the data is in the public domain

Commons e Open

But I would like attribution when others use my dataset. In that case, shouldn't I use a CC BY licence?

We recommend that you avoid using a CC BY licence. Here's why:

While attribution is a genuine, recognisable concern, not only might using a CC BY licence be legally unenforceable when no underlying copyright or SGDR protects the work, but it may also communicate the wrong message to the world. A better solution is to use CC0 and [simply ask for credit](#) (rather than require attribution), and provide a citation for the dataset that others can copy and paste with ease. Such requests are consistent with scholarly norms for citing source materials.

Legally speaking, datasets that are *not* subject to copyright or related rights (and are thus in the public domain) cannot be the object of a copyright licence. Despite this, agreements based in contract law may be enforceable. Creative Commons licences, however, are copyright licences. Therefore, where the conditions for a copyright or related right are not triggered, copyright licences, such as the CC BY licence, [are unenforceable](#).

In some cases, however, rights may exist (like the *sui generis* database right previously mentioned), and permission for others to use your dataset may be legally required. These rights are meant to protect the maker's investment, rather than originality. As such, database rights do not include the moral right of attribution. So by using a CC BY licence, you signal to users that you restrict access to your dataset beyond the protections provided by the law. We are not saying that this cannot be done, we are just saying that if you choose to do this, you should make sure you fully understand what it entails.

USARE UNA CC0

- NON SIGNIFICA DIVENTARE ACCADEMICAMENTE MALEDUCATI
- LA FONTE VA CITATA SEMPRE
- USATE LA CC0 E ASSOCIATE UNA FORMULAZIONE DELLA CITAZIONE CHE RICHIEDETE (DA COPIARE/INCOLLARE)

cannot be done, we are just saying that if you choose to do this, you should make sure you fully understand what it entails.

I'm uncomfortable with others using my research for commercial purposes. Should I use a non-commercial licence for my dataset?

We recommend you avoid using a non-commercial licence. Here's why:

For legal purposes, drawing a line between what is and is not 'commercial' can be tricky; it's not as black and white as you might think. For example, if you release a dataset under a non-commercial licence, it would clearly prohibit an organisation

I'm uncomfortable permitting use of my research for any and all purposes. Should I use a 'No Derivatives' (ND) licence for my dataset?

We recommend you avoid using a 'No Derivatives' licence. Here's why:

Similar to how a non-commercial licence might restrict meaningful reuse of your dataset, a ND licence can have the same effect: it may prevent someone from recombining and reusing your data for new research. For data to be truly Open Access, it must permit these important types of reuse.

It sounds like you're really pushing for the use of CC0 for open science datasets.

Exactly. Data is only open if anyone is free to use, reuse, and distribute it. This means it must be made available for both commercial and non-commercial purposes under non-discriminatory conditions that allow for it to be modified.

When data is made available for all reuse, others can create new knowledge from combining it. This leads to the enrichment of open datasets and further dissemination of knowledge. Accordingly, CC0 is ideal for open science as it both protects and promotes the unrestricted circulation of data.

And remember, it's bad science not to cite the source of data you use. To help others cite your data [include a citation](#) that users can copy and paste to give you credit for your hard work.

FAIR research software

The FAIR4RS Principles are:

F: Software, and its associated metadata, is easy for both humans and machines to find.

F1. Software is assigned a globally unique and persistent identifier.

- 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.

F2. Software is described with rich metadata.

F3. Metadata clearly and explicitly include the identifier of the software they describe.

F4. Metadata are FAIR, searchable and indexable.

A: Software, and its metadata, is retrievable via standardized protocols.

A1. Software is retrievable by its 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 software is no longer available.

I: Software interoperates with other software by exchanging data and/or metadata, and/or through interaction via application programming interfaces (APIs), described through standards.

I1. Software reads, writes and exchanges data in a way that meets domain-relevant community standards.

I2. Software includes qualified references to other objects.

R: Software is both usable (can be executed) and reusable (can be understood, modified, built upon, or incorporated into other software).

R1. Software is described with a plurality of accurate and relevant attributes.

- R1.1. Software is given a clear and accessible license.
- R1.2. Software is associated with detailed provenance.

R2. Software includes qualified references to other software.

R3. Software meets domain-relevant community standards.

FAIR RESEARCH
SOFTWARE

Table 1: The FAIR Principles for Research Software

FAIR software

Practical guide to Software Management Plans 2022

Core requirement (Section 5.1)	Example SMP question(s) (Section 6.1)
Purpose	Please provide a brief description of your software, stating its purpose and intended audience.
Version control	How will you manage versioning of your software?
User documentation	How will your software be documented? Please provide a link to the documentation if available.
Deployment documentation	How will the installation requirements of your software be documented? Please provide a link to the installation documentation if available.
Software licencing and compatibility	What licence will you give your software? How will you check that it respects the licences and dependencies it uses?

Core requirement (Section 5.1)	Example SMP question(s) (Section 6.1)
Purpose	Please provide a brief description of your software, stating its purpose and intended audience.
Version control	How will you manage versioning of your software?
Repository	How will you make your software publicly available? If you do not plan to make it publicly available you should provide a justification.
User documentation	How will your software be documented for users? Please provide a link to the documentation if available. How will you document your software's contribution guidelines and governance structure?
Software licencing and compatibility	What licence will you give your software? How will you check that it respects the licences and dependencies it uses?
Deployment documentation	How will the installation requirements of your software be documented? Please provide a link to the installation documentation if available.
Citation	How will users of your software be able to cite your software? Please provide a link to your software citation file (CFF) if available.
Developer documentation	How will your software be documented for future developers?
Testing	How will your software be tested? Please provide a link to the (automated) testing results if available.
Software Engineering quality	Do you follow specific software quality standards? If yes, which ones?
Packaging	How will your software be packaged and distributed? Please provide a link to your packaging information (e.g. entry in a package registry, if available).
Maintenance	How do you plan to procure long term maintenance of your software?

Core requirement (Section 5.1)	Example SMP question(s) (Section 6.1)
Purpose	Please provide a brief description of your software, stating its purpose and intended audience.
Version control	How will you manage versioning of your software?
Repository	How will you make your software publicly available? If you do not plan to make it publicly available, you should provide a justification.
User documentation	How will your software be documented for users? Please provide a link to the documentation if available. How will you document your software's contribution guidelines and governance structure?
Software licencing and compatibility	What type of licence will your software have? How will you check that it respects the licences of libraries and dependencies it uses?
Deployment documentation	How will the installation requirements of your software be documented? Please provide a link to the installation documentation if available. This documentation should include a complete and unambiguous description of dependencies to other software, datasets, and hardware.
Citation	How will users of your software be able to cite your software? Please provide a link to your software citation file (CFF) if available.
Developer documentation	How will your software be documented for future developers?
Testing	How will your software be tested? Please provide

Table 1. Core requirements of an SMP and examples of a low level of software management.

...YOU NEED SOFTWARE MANAGEMENT PLANS!

3. SE POSSIBILE, RENDERLI OPEN

Digital Science Report

The State of Open Data 2021

The longest-running longitudinal survey and analysis on open data

Foreword by Natasha SIMONS, Australian Research Data Commons (ARDC)

2021

November 2021

Open data saves lives. The global pandemic has highlighted
beyond anything that came before it the importance of data sharing
in solving the big challenges of our time. COVID-19 data may be the

Perché i dati aperti?

Oct. 2017

Digital Science Report

The State of Open Data 2017

A collection of analyses and articles about open data, curated by Figshare

Foreword by Jean-Claude Burgelman

OCTOBER 2017

"Open data is like a renewable energy source: it can be reused without diminishing its original value, and reuse creates new value."



Wilma van Wezenbeek

@wvanwezenbeek

Following



#osc2018 @sjDCC I really like what Sarah said just now "There is more risk in losing your data than sharing your data #openscience

Traduci il Tweet

11:14 - 13 mar 2018

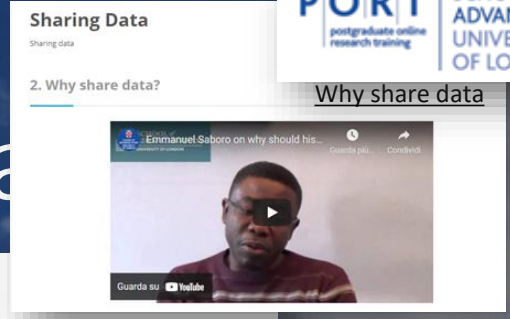
10 Retweet 10 Mi piace



<https://twitter.com/wvanwezenbeek/status/973502457115537408>

- 1) COME LE ENERGIE RINNOVABILI: IL LORO USO NON FA PERDERE VALORE, LO CREA
- 2) C'È PIÙ RISCHIO A PERDERLI CHE A CONDIVIDERLI

Il valore degli Open data



Better research

- Demonstrates research integrity, as there is transparency and accountability in the production of the data being released
- Encourages research enquiry and debate
- Promotes innovation and potential new data uses
- Encourages the improvement of research methods
- Prevents research fraud

UNA RICERCA MIGLIORE

- INTEGRITÀ
- DIBATTITO
- RIUSO INEDITO

Better impact

- Enables peer scrutiny of the research findings, validating the work carried out
- Increases the visibility of the research
- Provides credit for the creation of the data in its own right
- Can lead to new collaborations
- Produces a public record of the research

UN MAGGIORE IMPATTO

- VISIBILITÀ
- CREDITO
- COLLABORAZIONI

Better value

- Avoids duplication of effort in data creation
- Provides resources for use in teaching and learning
- Meets funder requirements
- Ensures data can be re-visited for future research
- Maximises return on research investment
- Preparing data for sharing also prepares it well for preservation

PIÙ VALORE

- EVITA DUPLICAZIONI
- MASSIMO RITORNO SU INVESTIMENTI

Il valore degli Open data / 3

PENSATE ALLA RICERCA CLINICA: QUANTO SI PERDE SE NON SI CONDIVIDONO I DATI E LE RICERCHE NON SONO RIPRODUCIBILI?

- AL MEGLIO: FONDI SPRECATI PER RICERCHE INUTILI
- ALLA PEGGIO: RISULTATI NON VALIDI IMMESSI NEL CIRCUITO CLINICO

Great values lost by not sharing data

Lack of reproducibility well known problem in medical research.

Investigations in the US: Up to 50% of studies not reproducible. 25% of this caused by unavailability of data.

At best: Expensive research is of little or no value.

At worst: Results of invalid research are put into clinical use.



People will contact me to ask about stuff

Christopher and Alex (C&A) say: "This is usually objection of people who feel overworked and th [data sharing] isn't part of their job..." I would a learning from each other – if a researcher is opp datasets, collaborating with others, and genera they should be outed by their community as a p

COMUNICAZIONE
SCIENTIFICA=
GRANDE
CONVERSAZIONE

People will misinterpret the data

C&A suggest this: "Document how it should be interpreted. Be prepared to help and correct such people; those that mis... the help." From the UK Data Archive: "P contextual information for your res... chers to correctly use and understand your o

IMPOSSIBILE, SE
NELLA «R» DI FAIR
AVETE
DOCUMENTATO

It's worth mentioning, however, a s... actually Freedom be useful to counter willful misrep... of Information legislation), as one c... refute the wrong interpretation."

My data is not very interesting

C&A: "Let others judge how intere... people that care about them." I'd... dataset has value to future resear... "climate change" was a research t... documenting and understanding

EHM... ALLORA
PERCHÉ USA
FONDI PUBBLICI?

I might want to use it in a research paper

Anyone who's discussed data sharing with a researcher is familiar with this excuse. The operative word here is *might*. How many papers have we all considered writing, only to have them shift to the back bur... is a real concern.

EMBARGO
PERFETTAMENTE
COMPATIBILE
CON FAIR

C&A suggest the embargo route: "One embargo; require people to archive the public after X months. You could even g things that are no longer cared about b eventually everything can become ope of their datasets, but I would caution to have any restrictions default to sharing. That is, after X months the data are automatically made open by the repository.

I would also add that, as the original collector of the data, you are at a huge advantage compared to others that might want to use your dataset. You have knowledge about your system, the conditions during collection, the nuances of your methods, et cetera that could never be fully described in the best metadata.

I'm not sure I own the data

My data is too complicated.

C&A: "Don't be too smug. If it turns out it's not that comp professional [standing]." I would add that if it's too comp complicated to reproduce, which means it's arguably not can be solved by more documentation.

IMPOSSIBILE, SE
NELLA «R» DI FAIR
AVETE
DOCUMENTATO

My data is embarrassingly bad

C&A: "Many eyes will help you imp accept your data for what it is." I a making the sausage. We know it's Plus it helps you strive will be at m collection phase.

MA COME SI FA A
FARE RICERCA
CON DATI
«BRUTTI»???

It's not a priority and I'm busy

SEMPRE MAGGIOR NUMERO DI FINANZIATORI E
DI RIVISTE LO CHIEDE...
STA DIVENTANDO UNA PRIORITÀ

...eppure

Digital Science Report

The State of Open Data 2021

The longest-running longitudinal survey and analysis on open data

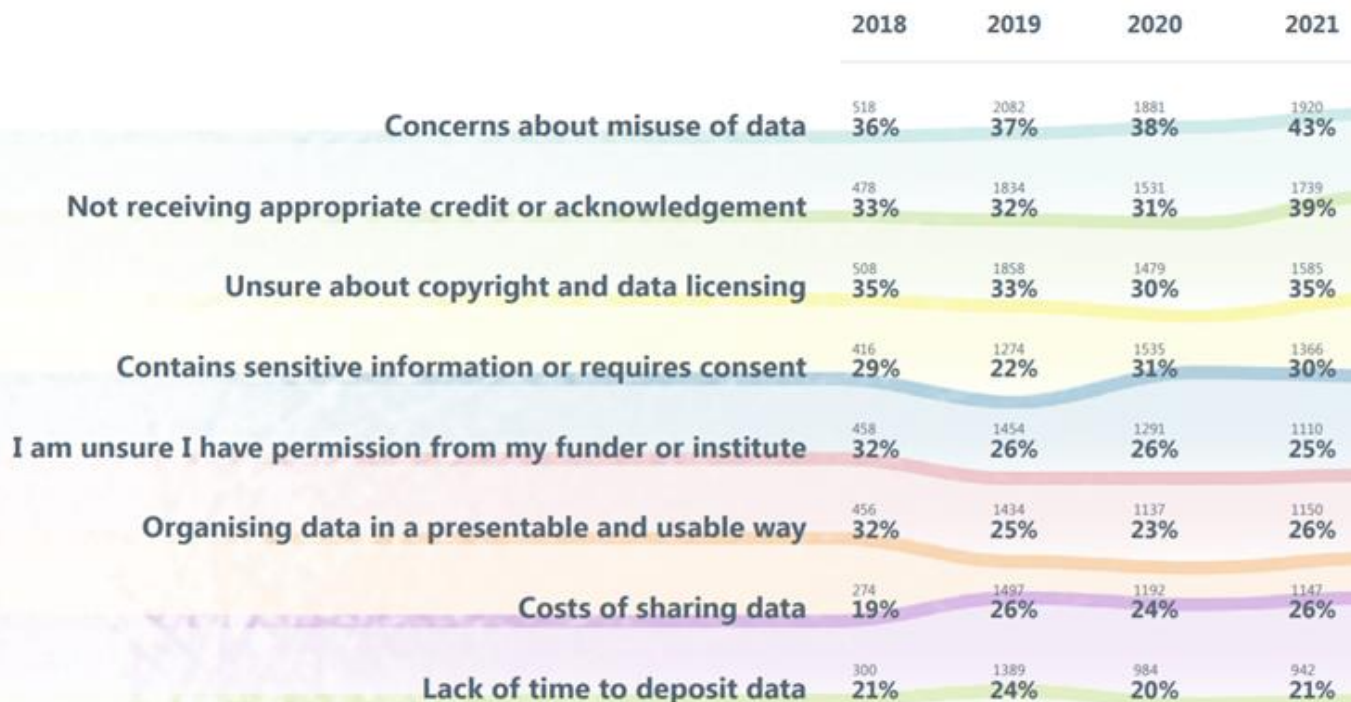
Foreword by Natasha SIMONS, Australian Research Data Commons (ARDC)

Nov. 29, 2021

November 2021

What is most striking about this year's State of Open Data report is that while researchers' familiarity and compliance with the FAIR data principles is greater than ever before, there is also more concern about sharing datasets than ever before. In their article on the three key findings of this year's State of Open Data report, Dr. Greg Goodey and Megan Hardeman stress that concern has risen in several key areas, one of which is not receiving enough credit or acknowledgement for data sharing. This points to the uncomfortable tension between the increasing ubiquity of data management and data availability policies and the rareness of rewards and recognition for data sharing. Clearly, the reward and recognition structures of academia are misaligned with

transparency of research

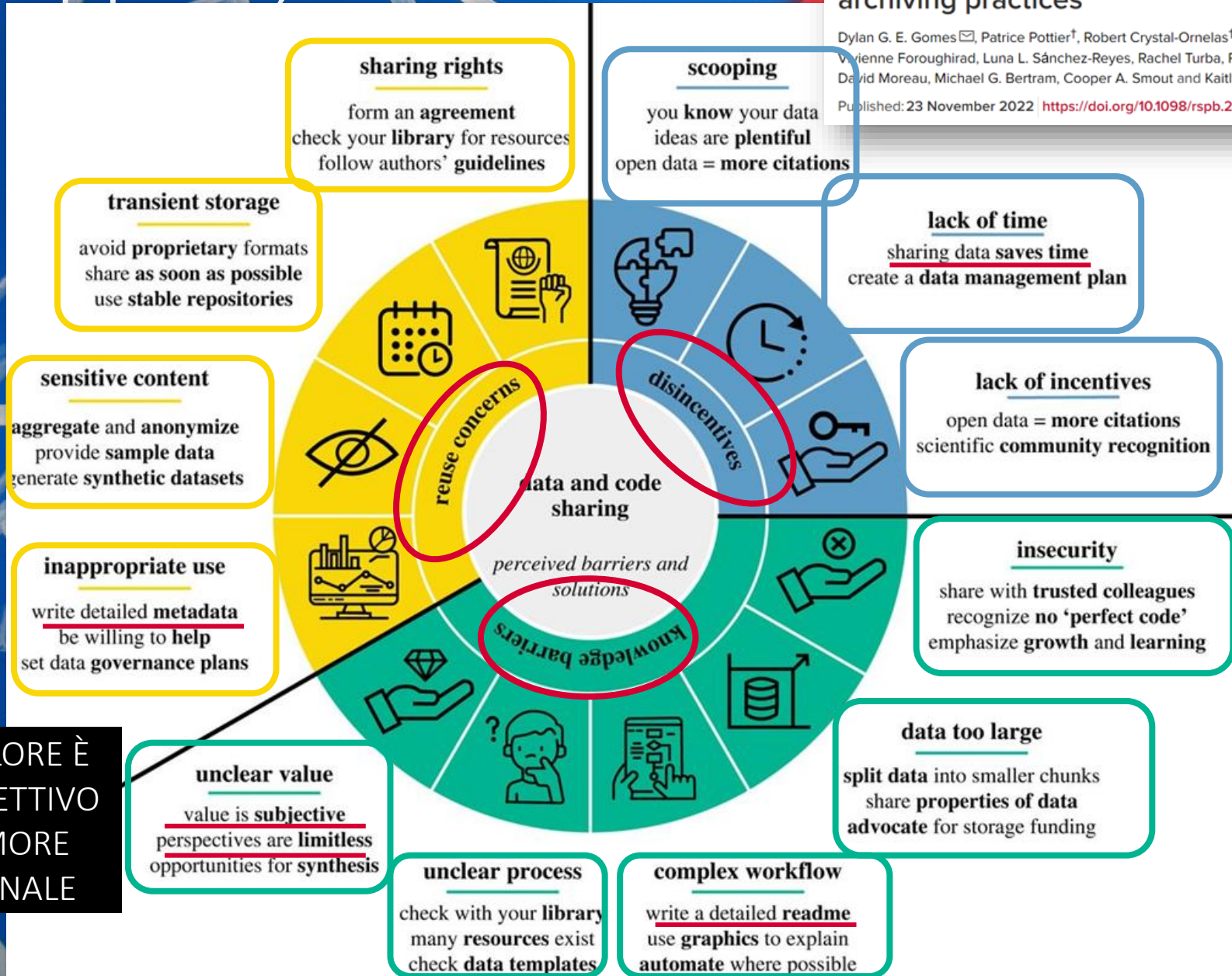


- PREOCCUPAZIONI
- USO IMPROPRIO
 - NESSUN CREDITO
 - CONSENSO

Why don't we share data and code? Perceived barriers and benefits to public archiving practices

Dylan G. E. Gomes, Patrice Pottier, Robert Crystal-Ornelas, Emma J. Hudgins, Vivienne Foroughirad, Luna L. Sánchez-Reyes, Rachel Turba, Paula Andrea Martini, David Moreau, Michael G. Bertram, Cooper A. Smout and Kaitlyn M. Gaynor
Published: 23 November 2022 | <https://doi.org/10.1098/rspb.2022.1113>

...eppure, ci sono resistenze



IL VALORE È
SOGGETTIVO
RUMORE
/SEGNALE

Pro e d

	REASONS NOT TO SHARE DATA	REPLIES OR ARGUMENTS IN FAVOUR OF SHARING
1	My data is not of interest or use to anyone else.	It is! Researchers want to access data from all kinds of studies, methodologies and disciplines. It is very difficult to predict which data may be important for future research. Who would have thought that amateur gardener's diaries would one day provide essential data for climate change research? Your data may also be essential for teaching purposes. Sharing is not just about archiving your data but about sharing them amongst colleagues.
2	I want to publish my work before anyone else sees my data.	Data sharing will not stand in the way of you first using your data for your publications. Most research funders allow you some period of sole use, but also want timely sharing. Also remember that you have already been working with your data for some time so you undoubtedly know the data better than anyone coming to use them afresh. If you are still concerned you can embargo your data for a specific period of time.
3	I have not got the time or money to prepare data for sharing	It is important to plan data management early in the research data lifecycle. Data management ideally becomes an integral part of your research practice, reduces time and financial costs and greatly enhancing the quality of the data for your use too.
4	If I ask my respondents for consent to share their data then they will not agree to participate in the study.	Don't assume that participants will not participate because data sharing is discussed. Talk to them - they may be less reluctant than you might think, or less concerned over data sharing! Make it clear that it is entirely their decision, whereby they can decide whether their data can be shared, independent of them participating in the research. Explain clearly what data sharing means, and why it may be important. But they are still free to consent or not. You can always explain what data archiving means in practice for their data. If you have not asked permission to share data during the research, then you can always return to gain retrospective permission from participants.
5	I am doing highly sensitive research. I cannot possibly make my data available for others to see.	The first thing is to ask respondents and see if you can get consent for sharing in the first instance. Anonymisation procedures can help to protect identifying information. If these first two strategies are not appropriate then consider controlling access to the data or embargoing for a period of time. Also data that is held in the UK Data Archive is not publically available. Only registered researchers can gain access to the data.
6	I am doing quantitative research and the combination of my variables discloses my participant's identity.	Quantitative data can be anonymised through processes of aggregation, top coding, removal of variables, or controlled access to certain variables (i.e. postcodes).
7	I have collected audiovisual data and I cannot anonymise them, therefore I cannot share these data.	Visual data can be anonymised through blurring faces or distorting voices, but this can be time consuming and costly to carry out. It can mean losing much of the value of the data. It is better to ask for consent to share data from participants in an unanonymised form,
8	I have made promises to destroy my data once the project finishes.	Why were such promises made? Always avoid making unnecessary promises to destroy data. There is usually no legal or ethical need to do so, except in the case of personal data. But that certainly would not apply to research data in general. Also consider where you have received this advice from? You may need to negotiate with research ethics committee or ethics boards about this agreement.

CI SONO BUONI
ARGOMENTI CHE
SMONTANO LE
PAURE RISPETTO
ALLA
CONDIVISIONE

A wooden bench with a sign on it. The sign is made of white-painted wooden planks and has the text "IF YOU ARE NOT DOING WHAT YOU LOVE, YOU ARE WASTING YOUR TIME." written on it in black, bold, sans-serif capital letters. The bench is made of light-colored wood and is set on a brick-paved surface. The background shows a brick wall and a concrete curb.

**“IF YOU ARE NOT
DOING WHAT
YOU LOVE,
YOU ARE
WASTING
YOUR TIME.”**

... grazie e... ora tocca a voi!