


Research Data Management and FAIR Principles

More better, Open Science

module 1

Emma Lazzeri  0000-0003-0506-046X

23 February 2026



WARNING: this is going to be an
interactive lesson!

Please **stop** looking at your
email, chats, messages!



What is Open Science?

- Go to www.menti.com and insert code:
4758 1389

OR

- use your mobile/tablet to scan the QR code





Mentimeter Answers: What is Open Science? 1/4

- Open Science è dare la possibilità ad altri ricercatori di collaborare e contribuire
- Scienza aperta
- la possibilità di scambiare dati ed informazioni scientificamente validate
- scienza accessibile a tutti, con le conseguenti operazioni necessarie alla comprensibilità e fruibilità
- sharing data and informations
- Informazioni e Dati della ricerca disponibili a tutti
- Sono fruibili da tutti, i dati disponibili possono essere consultati
- Un grande contenitore in cui esistono diversi oggetti a cui tutti possono accedere
- condividere le informazioni, i risultati delle ricerche scientifiche con la comunità scientifica e il pubblico in generale
- scienza aperta
- Condividere
- It's an approach that aims at making research results accessible to the widest possible public
- Un 'insieme di: dati aperti, articoli accessibili collaborazione aperta fra scienziati e società
- è la condivisione dei dati e risultati della ricerca con modalità che siano accessibili, riusabili, interoperabili sia dalle persone sia dalle macchine



Mentimeter Answers: What is Open Science? 2/4

- Approccio che rende la ricerca aperta, accessibile e trasparente ed ha come obiettivo quello di condividere conoscenza scientifica come bene pubblico
- Si riferisce alla pratica di rendere la ricerca scientifica, i dati e le metodologie accessibili a tutti, con l'obiettivo di migliorare la trasparenza, la riproducibilità e la collaborazione.
- è un approccio alla ricerca scientifica che promuove trasparenza, accessibilità e condivisione di conoscenze, dati e risultati, lungo tutto il ciclo della ricerca
- Informazioni/ valori numerici che consentono di rispondere a una domanda di ricerca: fonti della formulazione della/e domanda/e di ricerca; dati e procedure (metodologia); dati e informazioni empiric
- It is a way of conducting research that includes among its objectives the sharing and reusability of research results and their sharing with citizens.
- Scienza aperta e disponibile per creare nuova conoscenza.
- Scienza aperta
- Condivisione, trasparenza, riproducibilità
- Diffusione e condivisione di dati e risultati della ricerca
- Condivisione di metodi e materiali per accrescere la conoscenza e la consapevolezza in campo scientifico
- Open Science is a movement and set of practices that make the processes, outputs, and tools of research as transparent, accessible, and reusable as possible.



Mentimeter Answers: What is Open Science? 3/4

- Credo che si tratti di un metodo di ricerca che pone accessibili a tutti
- Rendere i dati condivisibili per favorire la replicabilità dei risultati
- Ricerca collaborativa attraverso accesso pubblico a dati e metodi di ricerca
- Scienza Aperta come non solo rendere i dati e l'output della ricerca FAIR, ma anche promuovere il riutilizzo dei dati e favorire la diffusione dei risultati della ricerca
- Un approccio alla ricerca scientifica che promuove trasparenza accessibilità e condivisione di conoscenze
- Trasparenza, condivisione accessibilità della ricerca scientifica
- Movement to make research, data and dissemination accessible to everyone. Transparency and knowledge following FAIR principles
- E' una modalità diretta di condivisione dell'intero processo di ricerca con la società, rendendola potenzialmente fruibile a tutti
- Accesso a dati pubblici
- Un poeta siciliano, Ignazio Buttitta, scriveva (traduco, male ahimè!!): Se io mi appoggio a te e tu ti appoggi a me, non c'è vento e tempesta che possa farci cadere. Ecco la forza "l'open science".
- It is science with linked open data.
- Scienza aperta
- Il principio per cui si condivide ciò che si scopre e come lo si è scoperto



Mentimeter Answers: What is Open Science? 4/4

- é un sistema di condivisione e riuso "aperto" dei dati scientifici
- Conoscenza trasparente accessibile e condivisa
- condividere risultati e metodi della ricerca
- Obiettivo di rendere la ricerca scientifica accessibile a tutti
- È la tempestiva condivisione di dati e risultati della ricerca scientifica, che ne consente l'uso alla comunità scientifica e a qualsiasi stakeholder. Massimizza lo sfruttamento dei risultati.
- Un approccio, rivolto a dare libero accesso e facilitare lo scambio del sapere scientifico
- Principio operativo di condivisione delle scienza nella società, e.g. condivisione di risultati e/o di database di ricerca
- Letteralmente è la scienza aperta necessaria per condividere la conoscenza e i risultati della ricerca
- un approccio che consente la condivisione della conoscenza
- Scienza che mette a disposizione dati in modo verificabile e riutilizzabile
- è un approccio che promuove la condivisione libera e trasparente di dati, metodi, pubblicazioni e risultati scientifici. Rende la conoscenza scientifica accessibile a tutti.
- E' una pratica che consente di condividere dati e risorse a chiunque sia interessato
- Rendere i dati e le informazioni liberamente disponibili a tutti

Open Science is...

not only open access to
publications!

image credit: Components of Open Science, UNESCO 2020,
https://en.unesco.org/sites/default/files/open_science_brochure_en.pdf



Definition of Open Science

Open Science is defined as an **inclusive construct** that combines various movements and practices aiming to make **multilingual scientific knowledge openly available, accessible and reusable for everyone**, to increase **scientific collaborations** and **sharing of information** for the **benefits of science and society**, and to **open the processes** of scientific knowledge creation, evaluation and communication to societal actors **beyond the traditional scientific community**.

It comprises **all scientific disciplines and aspects of scholarly practices**, including basic and applied sciences, natural and social sciences and the humanities, and it builds on the following **key pillars**:

- I. open **scientific knowledge**,
- II. open **science infrastructures**,
- III. science **communication**,
- IV. open **engagement of societal actors** and open dialogue with **other knowledge systems**.



Course Contents

1

RDM

Stories to understand why RDM is so important and what RDM is

3

Open Data and Legal Aspects

Open Data, data protection and legal aspects

2

FAIR principles

What they are

4

DMP

What it is and how to build it



Open Science is about collaboration!

Not only publications

Science is also:

- data
- software
- protocols
- negative results
- lab notes
- project deliverable
- and much more...



26 billion €

**are lost every year in Europe for
not managing the data properly**

“Publishing research
without **data** is simply
advertising, not
science”

Graham Steel

What happens if...

We do not manage
and share research
data in the correct
way?



Data can be lost...

JAMA journal retracts paper when author can't produce original data

In July 2017, a *JAMA* journal called for an investigation into a 2013 paper it had published after concluding that the article had “scientific and ethical concerns.” Now the journal, *JAMA Otolaryngology – Head & Neck Surgery*, is retracting the paper.

The article, “Dexamethasone for the prevention of recurrent laryngeal nerve palsy and other complications after thyroid surgery: a randomized double-blind placebo-controlled trial,” came from a group in Italy led by Mario Schietroma, of the Department of Surgery at the University of L'Aquila, in Abruzzo, Italy. Schietroma, who in December admitted to us that a retracted 2015 paper of his in the *Journal of the American College of Surgeons* suffered from “misinterpretation of the statistical data,” now has four retractions.



Neither [the original dataset and the approved protocol] have been provided by Dr Schietroma, and the university has informed us that “without those pieces of information the results of the papers under investigation cannot be validated.”

<https://retractionwatch.com/2018/10/25/jama-journal-retracts-paper-when-author-cant-produce-original-data/>

The importance of depositing research data

MENU ▾

nature

Carlisle has kept going. This year, he warned about dozens of anaesthesia studies by an Italian surgeon, Mario Schietroma at the University of L'Aquila in central Italy, saying that they were not a reliable basis for clinical practice⁶. Myles, who worked on the report with Carlisle, had raised the alarm last year after spotting suspicious similarities in the raw data for control and patient groups in five of Schietroma's papers.



Bottled oxygen, used by anaesthetists during surgery. Credit: Mark Thomas/Alamy

The challenges to Schietroma's claims have had an impact in hospitals around the globe. The World Health Organization (WHO) cited Schietroma's work when, in 2016, it issued a recommendation that anaesthetists should routinely boost the oxygen levels they deliver to patients during and after surgery, to help reduce infection. That was a controversial call: anaesthetists know that in some procedures, too much oxygen can be associated with an increased risk of complications – and the recommendations would have meant hospitals in poorer countries spending more of their budgets on expensive bottled oxygen, Myles says.

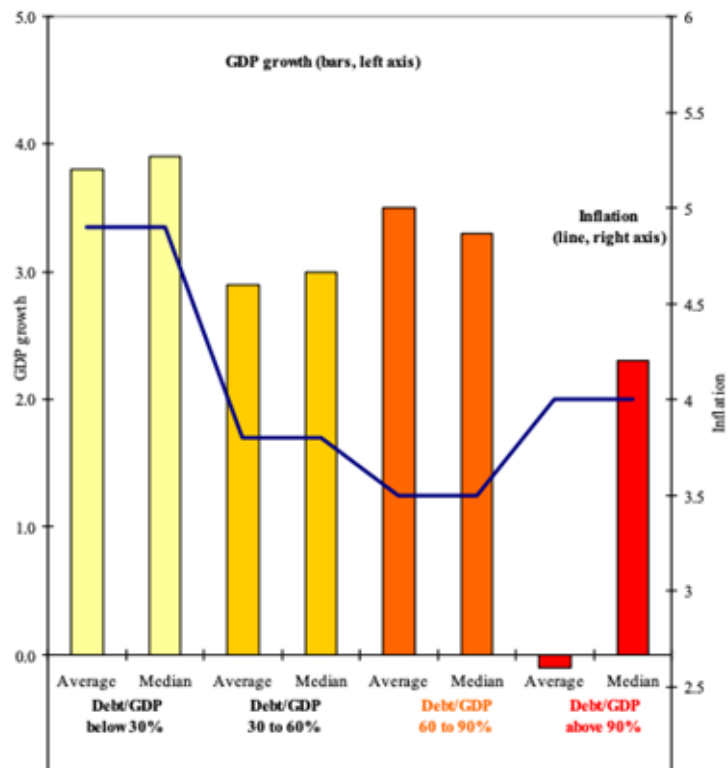
The five papers Myles warned about were quickly retracted, and the WHO revised its recommendation from 'strong' to 'conditional', meaning that clinicians have more freedom to make different choices for various patients. Schietroma says his calculations were assessed by an independent statistician and through peer review, and that he purposely selected similar groups of patients, so it's not surprising if the data closely match. He also says he lost raw data and documents related to the trials when L'Aquila was struck by an earthquake in 2009. A spokesperson for the university says it has left enquiries to "the competent investigating bodies", but did not explain which bodies those were or whether any investigations were under way.

Errors can get undetected

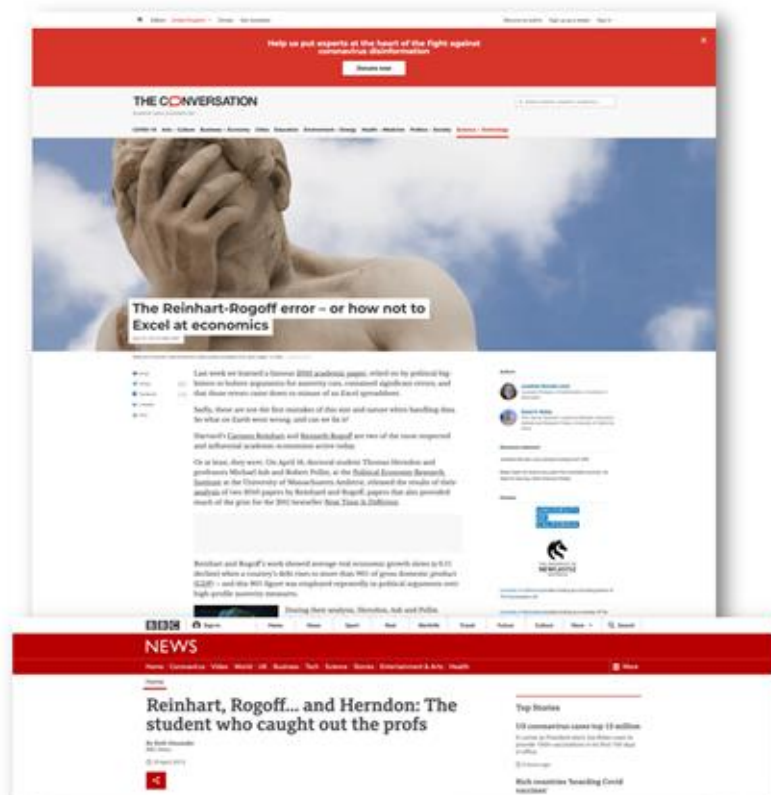
The case with **austerity theory**.

- The thesis: economic growth slows down dramatically when the size of a country's debt exceeds 90% of gross domestic product.
- The results shown in the paper were used **to support public austerity policies** during the recent economic crisis.
- But some considerations were based on wrong calculations.
- This was discovered by a student who could not replicate the results and asked the authors for the original dataset.

Figure 2. Government Debt, Growth, and Inflation: Selected Advanced Economies, 1946-2009



Errors and miscalculations



Does High Public Debt Consistently Stifle Economic Growth? A Critique of Reinhart and Rogoff

Thomas Herndon*

Michael Ash

Robert Pollin

April 15, 2013

[Herndon, 2013](#)

JEL CODES: E60, E62, E65

Abstract

We replicate Reinhart and Rogoff (2010a and 2010b) and find that coding errors, selective exclusion of available data, and unconventional weighting of summary statistics lead to serious errors that inaccurately represent the relationship between public debt and GDP growth among 20 advanced economies in the post-war period. Our finding is that when properly calculated, the average real GDP growth rate for countries carrying a public-debt-to-GDP ratio of over 90 percent is actually 2.2 percent, not -0.1 percent as published in Reinhart and Rogoff. That is, contrary to RR, average GDP growth at public debt/GDP ratios over 90 percent is not dramatically different than when debt/GDP ratios are lower.

We also show how the relationship between public debt and GDP growth varies significantly by time period and country. Overall, the evidence we review contradicts Reinhart and Rogoff's claim to have identified an important stylized fact, that public debt loads greater than 90 percent of GDP consistently reduce GDP growth.

Data can be manipulated

Nikolai Ivanovich Yezhov was a Soviet secret police official under Joseph Stalin who was head of the NKVD from 1936 to 1938, during the height of the Great Purge. After he fell from Stalin's favour he was executed. Among art historians, he also has the nickname "The Vanishing Commissar" because after his execution, his likeness was retouched out of an official press photo; he is among the best-known examples of the Soviet press making someone who had fallen out of favour "disappear".



The Newseum (1 September 1999). "The Commissar Vanishes" in The Vanishing Commissar".
Archived from the original on 8 February 2007.
https://en.wikipedia.org/wiki/Nikolay_Yezhov

Data can be manipulated

Largely used drug in surgery due to evidence in scientific papers.

- An investigation by Jacob-Liebig-University into the research conduct of J. Boldt.
- They found that data included in his student's MD theses were significantly different from the data in the published research papers where J. Boldt was the responsible author
- Findings in the investigation include changed mean values and standard deviations providing considerably more positive results



The screenshot displays the Association of Anaesthetists website. At the top, the logo and name 'Association of Anaesthetists' are visible. Below this is a search bar. A prominent orange banner reads 'PUBLICATIONS'. The main content area features the journal title 'Anaesthesia' with the subtitle 'Peri-operative medicine, critical care and pain'. A notice states 'THIS ARTICLE HAS BEEN RETRACTED' with a 'Free Access' link. The retracted article is titled 'Retracted: Volume therapy with hypertonic saline hydroxyethyl starch solution in cardiac surgery'. Below this, a link for 'Retraction(s) for this article' is provided. The authors listed are J. Boldt, D. Kling, C. Herold, F. Dapper, and G. Hempelmann. The journal information at the bottom indicates 'Anaesthesia 2021, 76, 563' and the DOI 'doi:10.1111/anae.15378'. A section titled 'Retraction' contains a detailed text explaining the retraction, stating that the article published online in November 1990 has been retracted by agreement between the journal's Editor-in-Chief, Dr Andrew Klein and John Wiley and Sons Ltd, following an investigation by Jacob-Liebig-University into the research conduct of J. Boldt.

Association of Anaesthetists

Search

PUBLICATIONS

Anaesthesia
Peri-operative medicine, critical care and pain

THIS ARTICLE HAS BEEN RETRACTED [Free Access](#)

Retracted: Volume therapy with hypertonic saline hydroxyethyl starch solution in cardiac surgery

[Retraction\(s\) for this article](#)

J. BOLDT, D. KLING, C. HEROLD, F. DAPPER, G. HEMPELMANN

Anaesthesia 2021, 76, 563 doi:10.1111/anae.15378

Retraction

Retraction: Boldt J, Kling D, Herold C, Dapper F, Hempelmann G. Volume therapy with hypertonic saline hydroxyethyl starch solution in cardiac surgery. *Anaesthesia*. The above article from the *Anaesthesia*, published online November 1990 in Wiley Online Library (<https://associationofanaesthetists-publications.onlinelibrary.wiley.com/doi/abs/10.1111/j.1365-2044.1990.tb14621.x?sid=nlm%3Apubmed>), and in Volume 45, pp. 928–34, has been retracted by agreement between the journal's Editor-in-Chief, Dr Andrew Klein and John Wiley and Sons Ltd. This retraction follows an investigation by Jacob-Liebig-University into the research conduct of J. Boldt while teaching at the university and working at the Hospital of the City of Ludwigshafen. They found that data included in his student's MD theses were significantly different from the data in the published research papers where J. Boldt was the responsible author. Findings in the investigation include changed mean values and standard deviations providing considerably more positive results. Therefore the Jacob-Liebig-University recommends that journal editors retract all papers where J. Boldt is the responsible author.

Data can be invented

REPORT

Coping with Chaos: How Disordered Contexts Promote Stereotyping and Discrimination

Diederik A. Stapel^{1,*}, Siegwart Lindenberg^{1,2,*}
* See all authors and affiliations

Science 08 Apr 2011;
Vol. 332, Issue 6026, pp. 251-253
DOI: 10.1126/science.1201068

Article Figures & Data Info & Metrics eLetters PDF

**This article has been retracted. Please see:
[Is retracted by - December 02, 2011](#)**

Abstract

Being the victim of discrimination can have serious negative health- and quality-of-life-related consequences. Yet, could being discriminated against depend on such seemingly trivial matters as garbage on the streets? In this study, we show, in two field experiments, that disordered contexts (such as litter or a broken-up sidewalk and an abandoned bicycle) indeed

- 58 articles published by Diederik Stapel were withdrawn because they were based on invented data.
- His papers had been published in scientific journals considered prestigious (very high IFs!).
- Following reports from three doctoral students, the Dutch university for which he worked had started an investigation. Stapel then admitted that he had fabricated the data on numerous occasions.
- If he had shared his data before, he probably wouldn't have been able to fabricate fakes for so long.
- This case led the Netherlands become one of the pioneer countries in Open Science policy and practices

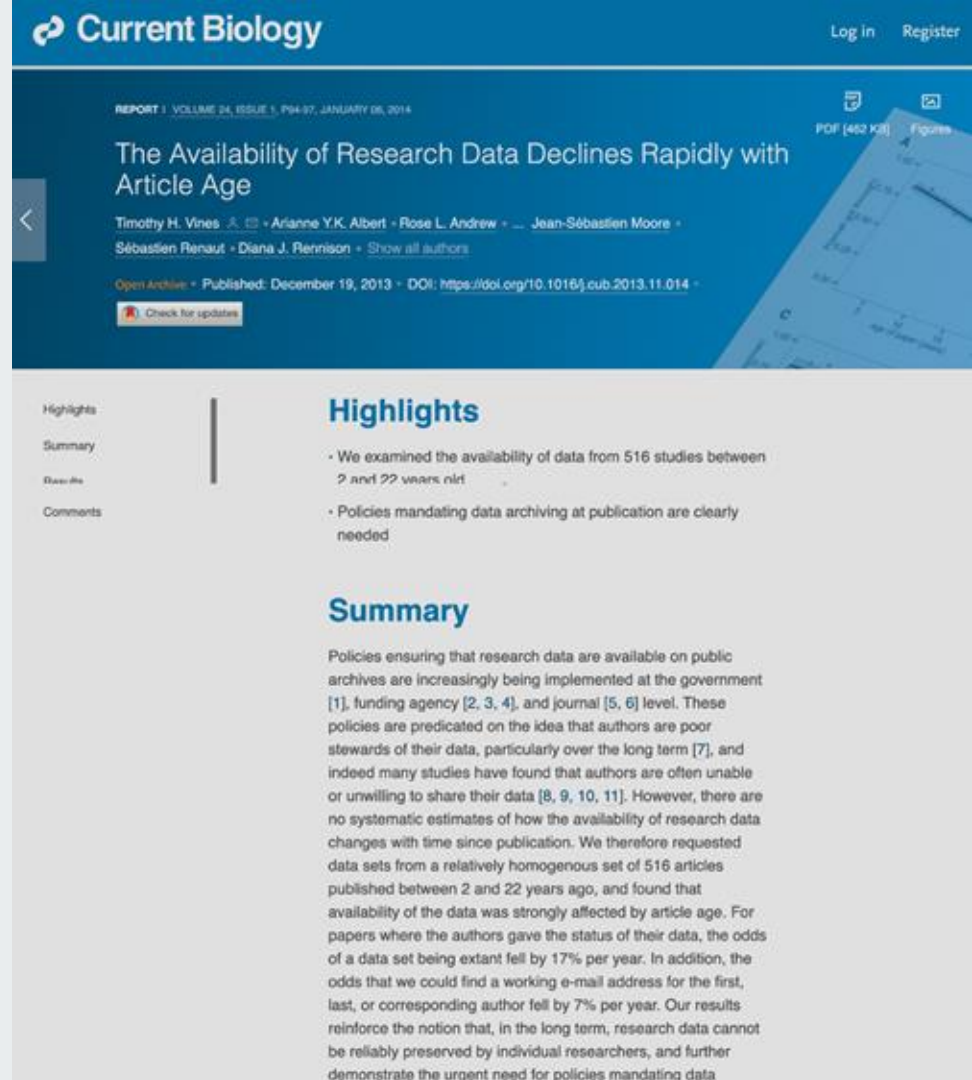
Reproducibility crisis

80% of data are unavailable after 20 years

In their parents' attic, in boxes in the garage, or stored on now-defunct floppy disks — these are just some of the inaccessible places in which scientists have admitted to keeping their old research data.

Such practices mean that data are being lost to science at a rapid rate.

[https://www.cell.com/current-biology/fulltext/S0960-9822\(13\)01400-0](https://www.cell.com/current-biology/fulltext/S0960-9822(13)01400-0)
<https://www.nature.com/articles/nature.2013.14416>



The screenshot shows the top half of a Current Biology article page. The header is blue with the journal name 'Current Biology' and links for 'Log in' and 'Register'. Below the header, the article title 'The Availability of Research Data Declines Rapidly with Article Age' is prominently displayed. The authors listed are Timothy H. Vines, Arianne Y.K. Albert, Rose L. Andrew, Jean-Sébastien Moore, Sébastien Renaut, and Diana J. Rennison. The article is dated January 08, 2014, and has a DOI of 10.1016/j.cub.2013.11.014. A 'Check for updates' button is visible. On the left, a sidebar contains links for 'Highlights', 'Summary', 'Abstract', and 'Comments'. The main content area on the right features a 'Highlights' section with two bullet points: 'We examined the availability of data from 516 studies between 2 and 22 years old' and 'Policies mandating data archiving at publication are clearly needed'. Below this is a 'Summary' section which begins with the text: 'Policies ensuring that research data are available on public archives are increasingly being implemented at the government [1], funding agency [2, 3, 4], and journal [5, 6] level. These policies are predicated on the idea that authors are poor stewards of their data, particularly over the long term [7], and indeed many studies have found that authors are often unable or unwilling to share their data [8, 9, 10, 11]. However, there are no systematic estimates of how the availability of research data changes with time since publication. We therefore requested data sets from a relatively homogenous set of 516 articles published between 2 and 22 years ago, and found that availability of the data was strongly affected by article age. For papers where the authors gave the status of their data, the odds of a data set being extant fell by 17% per year. In addition, the odds that we could find a working e-mail address for the first, last, or corresponding author fell by 7% per year. Our results reinforce the notion that, in the long term, research data cannot be reliably preserved by individual researchers, and further demonstrate the urgent need for policies mandating data

Why should you care?



If you manage it, you probably will not **loose** it



Organising your data will make your work more **efficient**



Some data is **unique and not reproducible** (meteorology, observation from the field) so you should take care of it



By correctly managing your data, you can improve **research integrity**



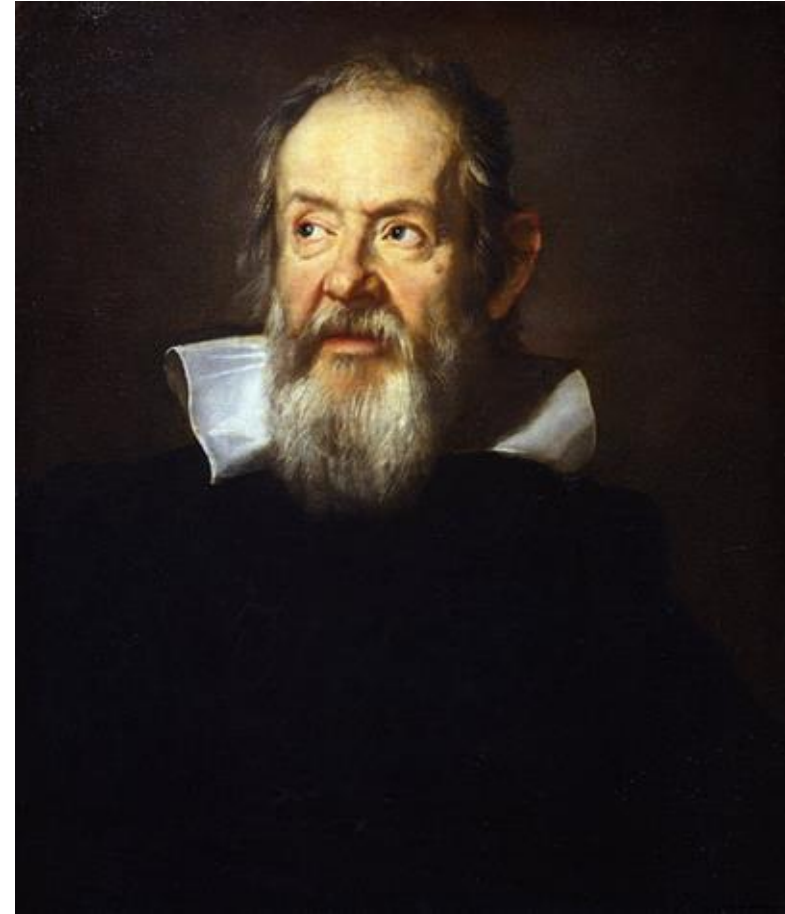
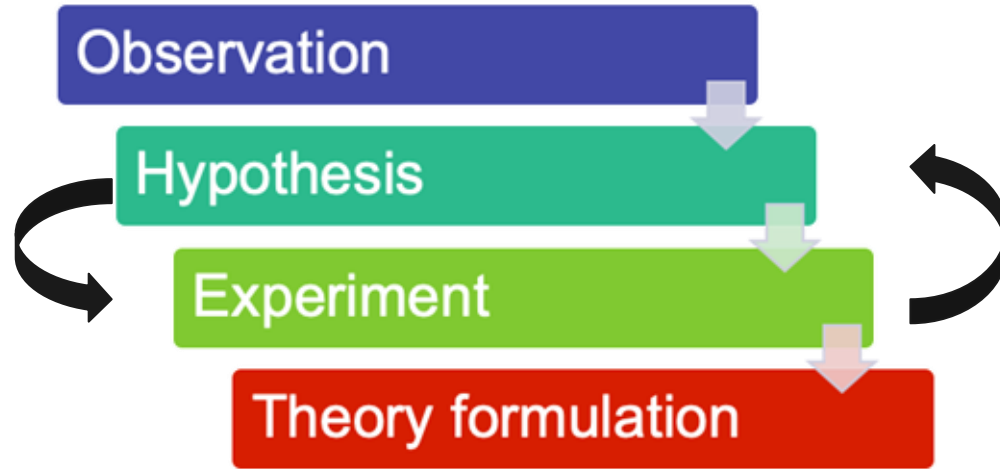
By managing your data, you enable **validation and control**



Someone else could use it in the future to **advance scientific progress**



The scientific method: data is key



RDM benefits

For researchers

- More visibility and citations
- Opportunity for collaboration
- Career recognition
- Decrease of non-compliance risks (legal, ethical, institutional and funders' policies)


For science

- Facilitates data finding and reuse
- Enables new research and new insights on the data
- Protection of valuable data
- Supports research integrity and reproducibility

For society

- Efficient use of public resources
- Better quality research can benefit to better decision-making
- Opportunities for citizen science
- Increased transparency and trust in science

What is Research Data Management?



Identify your research data

- Go to www.menti.com and insert code:
4758 1389

OR

- use your mobile/tablet to scan the QR code





Mentimeter Answers: Identify your research data 1/2

- Dati linguistici
- clinical longitudinal data
- Dati sulle trasferte del personale
- Dataset linguistici
- Timesheet del personale
- Dataset
- clinical data
- Microdati relativi a indagini campionarie dell'Istat

- Focus group
- Diversi dataset su tematiche legate alle scienze sociali
- timesheet, dati relativi ai costi sostenuti, survey
- Interviste
- Interviste utente
- Corpus
- Log di sistema
- Dati di monitoraggio superficiale e sub-

superficiali per la caratterizzazione dell'evoluzione di fenomeni di frana, nonché schede standardizzate, descrittive di un fenomeno di frana e danni associati

- Dati derivanti da questionari per identificare le necessità formative di un gruppo molto grande di stakeholder
- Risorse linguistiche, risorse archiviate di musei, gallerie, archivi storici e materiali culturali digitali.



Mentimeter Answers: Identify your research data 2/2

- Censimento di un fondo fotografico.. biblioteca digitale di un archivio inedito
- Video e foto
- Risposte a questionari
- Microscopy data
- Contributi da Citizen Science, sia qualitativi sia quantitativi
- Interviste
- food science data such as proximate

composition

- Localizzazione manufatti, interviste orali, elementi di cartografia storica (strade, manufatti, edifici etc) digitalizzati
- Dato testuali sia strutturati che non strutturati, metadati, file in diversi formati quali tiff ecc
- Metodologia/protocollo, dati documentali e strumentali, descrizione strumenti usati, composizione competenzaiale del team

- Dati ambientali, uso territorio, risorse ambientali, indicatori di funzionamento ecosistemico, dati socioeconomici
- Dati numerici di valutazione di sperimentazioni di tool AI o LLM in specifici task
- Dati su grandezze fisiche provenienti da strumenti che acquisiscono in continuo



What is data?

Data or it didn't happen!

Facts, observations or experiences on which an argument or theory is constructed or tested.

**Data are/contain information!
(in a variety of forms and formats)**

Types of research data

There is a huge variety of data types. Research data can be classified in different ways, for example based on their:

Content: numerical, textual, audiovisual, multimedia...

Format: spreadsheets, databases, images, maps, audio files, (un)structured text...

Mode of data collection: experimental, observational, simulation, derived/compiled from other sources

Digital (born-digital or digitized) or non-digital nature (e.g. paper surveys, notes...)

Primary (generated by the researcher for a particular research purpose or project) or secondary nature (originally created by someone else for another purpose)

Raw or processed nature

<https://www.ugent.be/en/research/datamanagement/why/rdm-explained.htm>



Image by [Gerd Altmann](#) from [Pixabay](#)

Research Data Management

Actions and practices to ensure that research data are:

Secure



Sustainable




(Re)usable



Some of the following slides are inspired to the Ghent University guide on RDM:

<https://www.ugent.be/en/research/datamanagement>



RDM means taking proper care of data.
Not only during the research project
but also in the longer term, and even
before!

Research data management is simply the effective handling of information that is created in the course of research.

How and why you should manage your research data: a guide for researchers
An introduction to engaging with research data management processes.
Caroline Ingram, JISC Guides

Data are first-class research objects

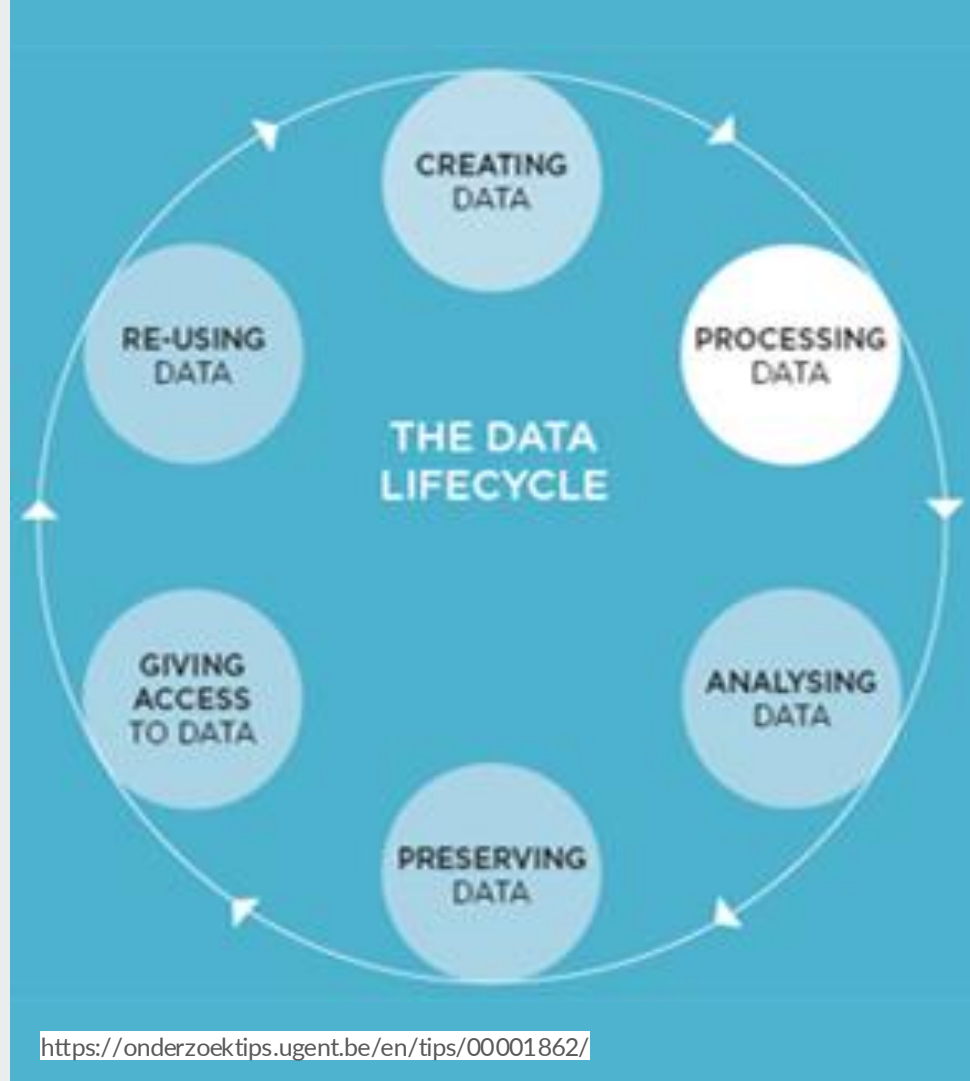
Check
Validation
Follow-ups
New research questions
Teaching

...



RDM practices encompass activities in every stage of the work with data

Before, during and after the research project. Choices made in one stage influence the next one.



Before starting data collection

- Design research
 - Plan and cost data management
 - Plan compliance with policies and regulations
(e.g. plan informed consent for data sharing)
 - Identify existing data sources
 - Plan data collection & processing
Look for protocols, standard procedures, templates
 - Design data security measures



Planning

Collecting data

- Data collection

Experiments, observations, surveys, modelling, compilation from other sources, acquisition of third party data

- Metadata generation
- Documenting data acquisition
- Data quality assurance

During the research project

Processing & analyzing data

- Entering, digitizing, transcribing, deriving, anonymizing data
- Data transformation, harmonization and integration
- Quality checking, validation, cleaning
- Analysis and interpretation
- Describing and documenting
- Storing, organizing, version control

During the research project

After the research project

- Data selection and appraisal
- Migrating to open, sustainable formats
 - Archiving data and documentation in appropriate location

Preserving data

After the research project

- Choose discipline specific or generic repository
 - Provide documentation and discovery metadata
- Choose appropriate license and access control
 - Publish data paper



Sharing
data

Definitions

Digital Object

In the context of this course: refers to any research result in its digital form, which can be uploaded into a repository (and possibly openly shared).

Examples: articles (pdf), datasets, software, images, videos, reports, posters or conference presentations, lectures (ppt), etc

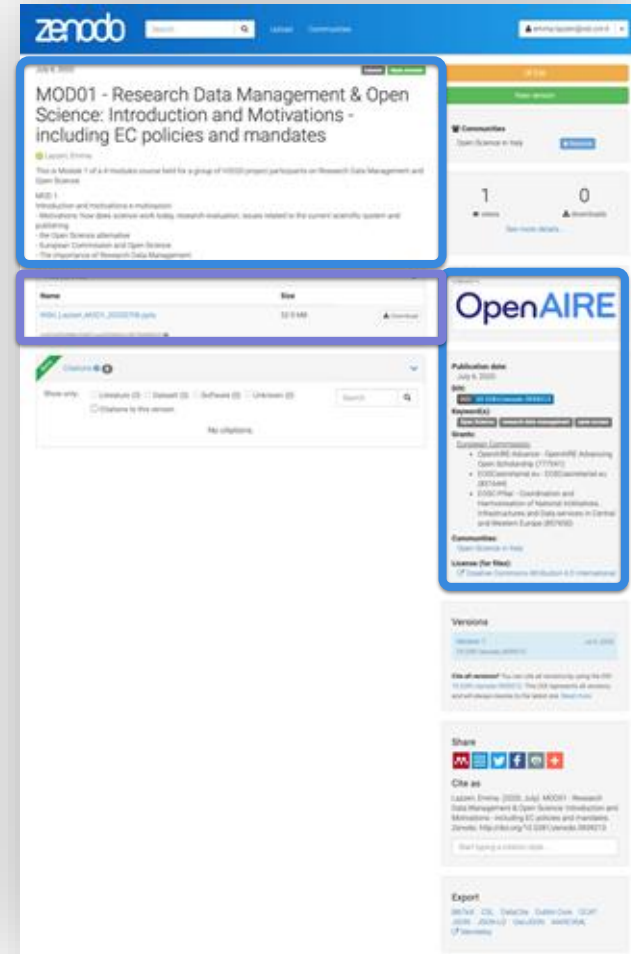
...

Metadata

Data about the digital object you are depositing

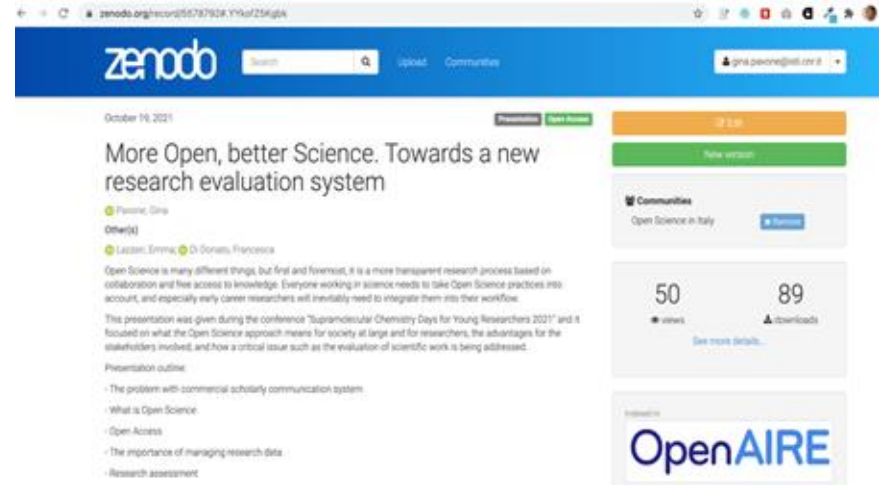
Payload

The digital object (or digital objects) that you are uploading for deposit (and that you can share). Also includes accompanying or description file (readme file, etc ...)



Repository

An **open-access repository** or **open archive** is a digital platform that holds research output and provides free, immediate and permanent access to research results for anyone to use, download and distribute. To facilitate **open access** such repositories must be **interoperable** according to the **Open Archives Initiative Protocol for Metadata Harvesting** (OAI-PMH). Open-access repositories, such as an **institutional repository** or **disciplinary repository**, provide free access to research for users outside the institutional community and are one of the recommended ways to achieve the **open access** vision described in the **Budapest Open Access Initiative** definition of open access. This is sometimes referred to as the **self-archiving** or "green" route to open access.





Persistent identifiers

- A **persistent identifier** (PI or PID) is a long-lasting reference to a document, file, web page, or other object.
- The term persistent identifier is usually used in the context of **digital objects** that are accessible over the Internet.
- Typically, such an identifier is not only persistent but **actionable**: you can plug it into a web browser and be taken to the identified source.
- It is like the barcode used on products...



PID exemples: DOI e ORCID

- In computing, a **digital object identifier (DOI)** is a persistent identifier or handle used to identify objects uniquely, standardized by the International Organization for Standardization (ISO).
- A DOI aims to be **resolvable**, usually to some form of access to the information object to which the DOI refers.
- This is achieved by **binding the DOI to metadata** about the object, such as a URL, **indicating where** the object can be found
- a DOI differs from identifiers such as ISBNs and ISRCs which aim only to identify their referents uniquely

The **Open Researcher and Contributor ID (ORCID)**

is a nonproprietary alphanumeric code to uniquely identify scientific and other academic authors and contributors. Do you have one? You should!!!

The screenshot displays a sample ORCID iD profile for Francesca Di Donato. The profile includes the following sections:

- Header:** Name (Francesca Di Donato), ORCID iD (https://orcid.org/0000-0003-0144-8934), and a 'Print view' link.
- Also known as:** @ederinita (twitter).
- Keywords:** Open science, Scholarly communication, Open Web Annotation, Digital Humanities.
- Email:** didonato@netseven.it, f.didonato@gmail.com.
- Biography:** Pisa (Italy), 07.05.1973.
- Employment and Qualifications:** A list of roles and positions, including TRIPLE eu project leader, Europe Ambassador of CO-OPERAS Go-FAIR Implementation Network, WG2 Leader of CO-OPERAS Go-FAIR Implementation Network, Member of European Research Task Force, Certificate: Essentials 4 Data Support, RDNI-DANS, External Expert Evaluator for the European Commission, Member of the board of directors of AISA - Associazione Italiana Scienza Aperta, Employee at Net7 srl, Pisa, European Association Network Councilor, National Abilitation - Associate Professor of Political Philosophy (14/A1), Research fellow at the University of Pisa and at the Scuola Normale Superiore di Pisa, PhD in «History of modern and contemporary Philosophy», Scuola Superiore di Studi universitari e postuniversitari S. Anna di Pisa (100/100 cum laude), and Degree in Political Science, Università di Pisa (110/110 cum laude).
- Open Access:** A note stating that all publications and presentations are released in Open Access and that the user uses deposit/publish many copies around the Web, as they are convinced that a lot of copies keep stuff safe. If a broken link is found, they hope to find a copy of it somewhere else.
- Summary:** A dark box at the bottom showing counts for various categories: Employment (9), Education and qualifications (3), Membership and service (1), Funding (7), and Works (47 of 47).
- Footer:** Record last modified Nov 25, 2020 1:40:35 PM and an 'Aiuto' (Help) button.



Do not harm

- Managing research data is usually an integral part of the research process, **so you probably already do it**. You only have to **reflect on** and to **improve your strategy**.
- Most of the activities should be familiar:
 - **naming files** so you can find them quickly;
 - keeping track of different **versions**, and deleting those not needed;
 - **backing up** valuable data and outputs;
 - controlling who has **access** to your data.

Fifty shades of no (to data sharing)

Too expensive

There's no business case

There's no commercial value

It's private

It's secret

It's our data

We have invested a lot of money in this

Link enough data and one will arrive at sensitive private information

It's not data, it's information

It will never work

We don't know how to do this

We don't have the right people to do this

We need the money

It's not ours, and we don't know who's data it is

No idea what the quality of the data is

We don't know where to find it

It's not our job

It isn't in the right format

I am not authorised

Who is going to use this anyway

People are going to misuse it

Image damage for the minister

We are not ready for this

Image loss for Government

The data file is too big

Not enough bandwidth

This is a first step, we will see what we can do later

We can't find it

We have no access

It is out of date / too old

We have it only on paper

We don't know if it's legal

Management says no

We never did this before

No value in it

No time / no resources

We will open up (but adapt 90%)

It's incorrect

Commercially sensitive

It is dangerous when linked

People are going to make the wrong conclusions

This is going to start a wrong discussion

We can't say whether we have it or we don't

People may download and cache the data and it will be out of date when they reuse it

We don't collect it regularly

Too many people will want to download it, which will cause our servers to fail

People would get upset

It's very sensitive information

We are not ready for this

Tell us who is going to use it and we will make it open

Our website cannot hold files this large

I do not want to know how to share my data.
I want to know how I can re-use other's data.


It's not ours and we don't have authorization from the data owner

We've already published the data (but it's unfindable/ unusable)

We know the data is wrong, and people will tell us where it is wrong, then we'd waste resources inputting the corrections people send us

Our IT suppliers will charge us a fortune to do an ad hoc data extract

We have to be careful with existing contracts



Tell me your excuses not to share and manage your data properly

- Go to www.menti.com and insert code:
4758 1389

OR


- use your mobile/tablet to scan the QR code





Mentimeter Answers: Tell me your excuses not to share and manage your data properly 1/2

- Sensitive data waiting to be published
- Alcune tipologie di dati non possono essere condivise senza accordo espliciti e formali tra le parti
- Sensitive data
- Dati poco importanti e non utili
- richiede parecchio tempo
- I dati sono protetti da un accordo di riservatezza con le parti coinvolte.
- dati sensibili
- problemi con gli editori di libri cartacei basati su quei dati
- Casi in cui l'etica professionale prevale sulla trasparenza, cioè relativi a soggetti fragili, identificabili seppure i dati sono anonimi
- La maggior parte dei dati è cartacea (trascrizioni, fotocopie), ordinata per tematiche. I dati su tabella sono al momento sensibili e soggetti ad elaborazione (ricerca in atto)
- Quelle che ho sentito più di frequente:
 - troppa burocrazia, troppo tempo sprecato, tanto non controlla nessuno, non serve veramente alla ricerca
- sono poco importanti sono pochi si tratta di dati qualitativi
- I dati sono imperfetti o non completi
- Perché i dati che tratto sono pochi e sono dati personali....
- Timore di non riuscire a renderli anonimi



Mentimeter Answers: Tell me your excuses not to share and manage your data properly 2/2

- Timore che siano manipolati
- Alcuni dati sensibili non possono essere condivisi (e.g., dati video in cui sia chiaramente identificabile l'identità di un partecipante minore).
- Il PI non vuole cedere il suo tesoro ad altri ricercatori
- Dati sensibili, specie in ambito medico
- Reperire i dati assorbe risorse scarse. La loro condivisione è possibile solo con precise politiche di accesso che evitino comportamenti opportunistici
- richiede tempo che non ho
- Dati sensibili
- mancanza di tempo
- abitudine errata di gestire i dati, fretta
- Timore che vengano usati da altri per pubblicarli
- Lo farò dopo ... e poi non lo fanno ...
- Non so come fare né se si possa con dati clinici che spesso sono di proprietà di altre strutture
- Difficoltà di trasformazione di analogici in dati digitali aperti in una logica FAIR.
- Meglio non farlo ci sono troppi copioni
- Non ci sono scuse. Va bene proteggerli solo in fase di ricerca e/o elaborazione, poi vanno condivisi, ovviamente nella forma più appropriata

HAVE
A
BREAK





Know your data!

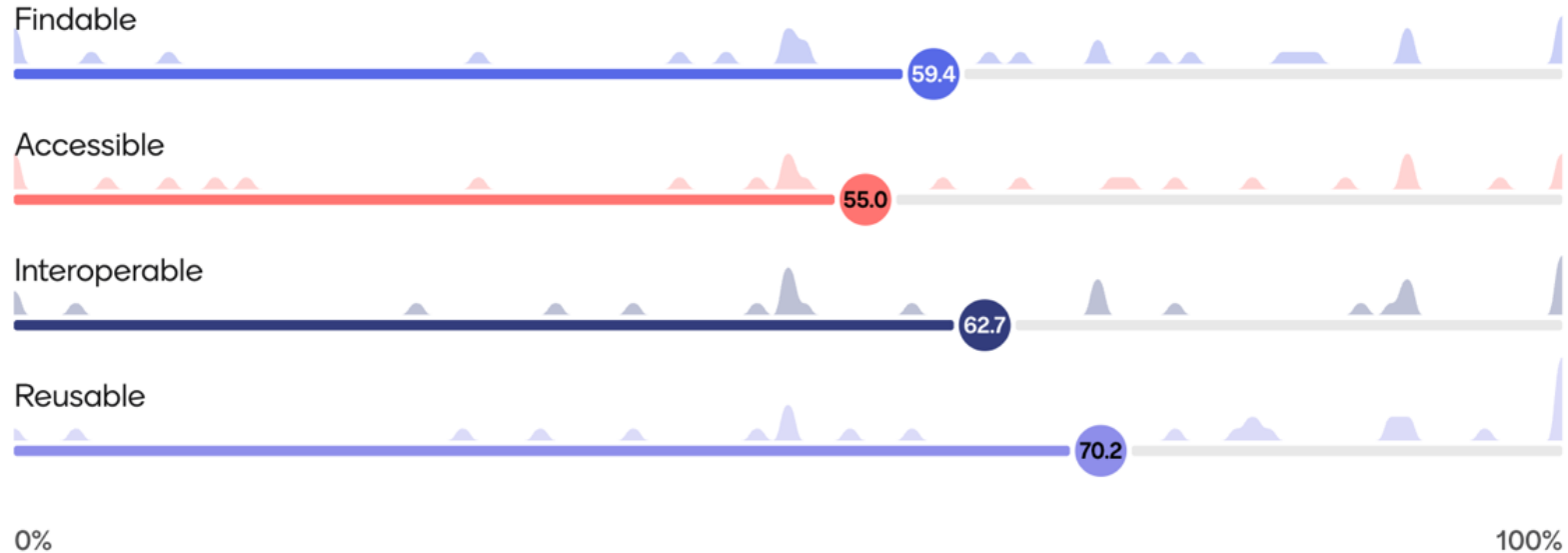
- Go to www.menti.com and insert code:
4758 1389

OR

- use your mobile/tablet to scan the QR code



Know your data: how much is your data...



FAIR principles

The FAIR Principles



Findable

Other can find your data



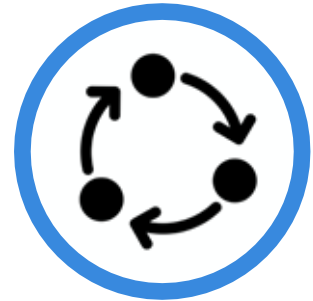
Accessible

Your data is accessible to others



Interoperable

Your data can be integrated with other data and/or they can be easily used and read by machines.



Reusable

Your data can be reused by others in new research

The FAIR Principles



- FAIR indicate a list of principles that can help you in making your data ready for Open Science
- They are **principles**, not standards!
- They were designed to enable optimal use of research data and methods
- A group of different experts designed the **FAIR principles** between 2014 and 2016
- They identified a set of 15 principles

By applying the FAIR principles

- You will produce high quality data
- You will maximise the impact of your research
- You will improve the recognition within and behind your research community
- You will be compliant with the European Commission requirements

No one size fits all

The application of the FAIR principles strongly depends on the specific discipline and on the way the single researcher works





Good practices to make your data FAIR



FAIRification basics

Documentation

Gives the context to make your data understandable by others

Metadata

Make your data easy to find

Data formats

Make your data simple to combine to other data and machine readable.

Access to data

It means to decide who will have access to your data and how

Persistent identifiers

Persistent links to data that allows other to find and cite (give credit to) your data.

Licenses

Are used to tell others how they can reuse your data.



Documentation

- Specifies the context that led to the creation/collection of your data to make them understandable
- At the beginning of a new (project) activity, you need to clearly define with your colleagues the strategy to structure and document your data.
- Document every detail of data collection/generation:
 - Methods
 - Tools
 - Software
 - Processes (who worked with the data? What did he/she did with the data? What are the relation to other data and/or publications?)
 - Metadata

Metadata

- Data describing data
- Very important for:
 - Access
 - Comprehension
 - Process
- Use your discipline specific standards: you will spend less time curating and interpreting data and more time to actually make science!

innovation in metadata design, implementation & best practice

Dublin Core™ Metadata Initiative

Home Specifications News Community Learning About Contact

quick search...

Home

The Dublin Core™ Metadata Initiative supports innovation in metadata design and best practices. DCM is supported by its members and is a project of AIS&T.

Stewardship



For more than twenty years, the DCM community has developed and curated Dublin Core Specifications. More recently, DCM has become recognised as a trusted steward of metadata vocabularies, concept schemes and other metadata artefacts, and has taken responsibility for other community-created specifications. DCM remains committed to this important work, and is actively developing more efficient and sustainable approaches to the stewardship of these standards, through the work of the [DCM Usage Board](#).

Community



DCM is defined by its community which is responsible for the innovative developments and evolving good practices which DCM shares with the world. Much of DCM's work is organised in working and interest groups. DCM's community is and has always been international, with active participants from around the world. The primary community event is the [DCM Annual Conference](#). DCM also organises regular [webinars](#), given by members of the community wishing to share their expertise with like-minded peers. Finally, DCM [collaborates](#) with a number of other organisations.

Learning



DCM supports teachers and learners of modern metadata technologies and practices. An updated [Metadata Basics](#) page highlights current trends in descriptive metadata in the style of Dublin Core, which aims at interoperability through using globally shared vocabularies, constrained in application-specific profiles, based on principles of Linked Data. Interest learners can also explore a [glossary page](#), a [Linked Data Coreference Index](#) that enumerates relevant skills to be learned, a [guide for users of DCM metadata terms](#), occasional [webinars](#) and tutorials at [DCM annual conferences](#).

Development



DCM has a long history of fostering and supporting technical development and innovation through the activities of its community, often in partnership with other organisations. Following on from the development of the ubiquitous [DCM Metadata Terms](#), the community has in more recent years focused on the concept of the [metadata application profile](#), developing supporting frameworks and conceptual models such as the [Singapore Framework](#). Most recently, the [Application Profiles Interest Group](#) has formed to address the next stage of development in this space.



News

DCM 2020 Call for Proposals

Following on from the success of DCM 2019 in Seoul (see [Proceedings](#)), we are pleased to announce the call for proposals in the DCM 2020 International Conference on Metadata, Ottawa, Canada, 14-17 September 2020. We are grateful to Carleton University for offering to host us this year. This year's conference will mark the 25th anniversary of the original Dublin Core™ workshop. We will both reflect on two and a half decades of innovations while looking ahead to future developments.

[Read more...](#)

ISO 15836 Part 2 is published based on a revision of DCM Metadata Terms

View our privacy policy | [DCM 20202020](#) | [Contact](#) | [DCM 20202020](#) | [DCM 20202020](#) | [DCM 20202020](#) | [DCM 20202020](#)

Findable

- The first step in (re)using data is to find them.
- Metadata and data should be easy to find for both humans and computers.
- Machine-readable metadata are essential for automatic discovery of datasets and services, so this is an essential component of the **FAIRification process**.



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



How can you assign a Persistent Identifier to your digital object?

Persistent identifiers need to be assigned by an entity that can ensure the persistency of the link to the object.

Smart Tip: [Zenodo](#) assigns DOIs to digital objects that do not already have one

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

Once the user finds the required data, she/he needs to know how can they be accessed, possibly including authentication and authorisation.

- **How do you give access to your data?**

Through a Repository

- **How do I choose the right repository?**

Directory of Open access repositories:

www.openoar.org

Registry of Research Data Repository

<https://www.re3data.org/>

Types of repositories

Who does curate/deposit in the repository?

Thematic or disciplinary repositories

Designed for specific contents, curated by specific communities: ArXiv, bioarXiv, PMC...

http://oad.simmons.edu/oadwiki/Disciplinary_repositories

Institutional or national repositories

Maintained and curated by single institutions/countries. Typically only authors based in the specific institution/country can deposit, everyone can access

What are the repository contents?

Literature Repositories

Reserved to text deposit (articles, reports, books, ...). Metadata reflect the repository contents.

<https://v2.sherpa.ac.uk/openoair/>

Data repositories

Designed to deposit data. They often are disciplinary and have specific metadata to describe the type of data they preserve.

<https://www.re3data.org/>

Catch-all repositories

All research products can be deposited (data, literature, presentations, poster, images, software, ...). Example: [Zenodo](#)



Interoperable



Data usually needs to be integrated with other data.

In addition, data needs to interoperate with applications or workflows for analysis, storage, and processing.

Use community standard or best practice!

- 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

Related/alternate identifiers

recommended ▾

Specify identifiers of related publications and datasets. Supported identifiers include: DOI, Handle, ARK, PURL, ISSN, ISBN, PubMed ID, PubMed Central ID, ADS Bibliographic Code, arXiv, Life Science Identifiers (LSID), EAN-13, ISTC, URNs and URLs.

■ Related identifiers

e.g. 10.1234/foobar.56789

N/A

Optional. Resource type of the related identifier.

+ Add another related identifier

- ✓
- cites this upload
- is cited by this upload**
- is supplemented by this upload
- is a supplement to this upload
- is referenced by this upload
- references this upload
- is previous version of this upload
- is new version of this upload
- continues this upload
- is continued by this upload
- has this upload as part
- is part of this upload
- reviews this upload
- is reviewed this upload
- documents this upload
- is documented by this upload
- is compiled/created by this upload
- compiled/created this upload
- is the source this upload is derived from
- has this upload as its source
- is identical to this upload
- is an alternate identifier of this upload

✓ N/A

Publication

- Annotation collection
- Book
- Book section
- Conference paper
- Data management plan
- Journal article
- Other
- Patent
- Preprint
- Project deliverable
- Project milestone
- Proposal
- Report
- Software documentation
- Taxonomic treatment
- Technical note
- Thesis
- Working paper

Dataset

Image

Reusable

- The ultimate goal of FAIR is to optimise the reuse of data.
- To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings
- Data should be linked to other resources and outcomes that clarify their context
- Key elements for reuse:

Licenses

Tell others how they can reuse your data!

Provenance

Where is your data coming from? What is the source of your data?



- R1. (Meta)data are richly described with a plurality of accurate and relevant attributes
- R2. (Meta)data are released with a clear and accessible data usage license
- R3. (Meta)data are associated with detailed provenance
- R4. (Meta)data meet domain-relevant community standards

How FAIR is your research data?

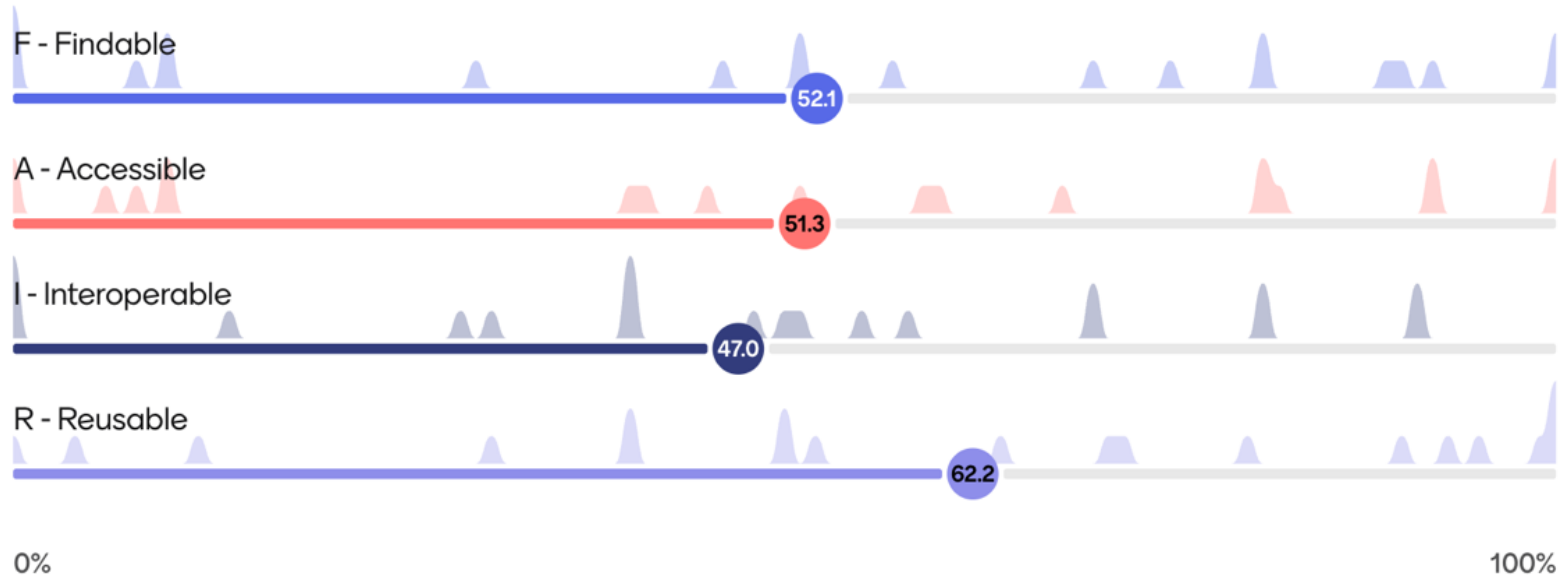
- Go to www.menti.com and insert code:
4758 1389

OR

- use your mobile/tablet to scan the QR code



How FAIR is your data?



Open Data

Copyright and legal aspects

Data is not yours



Data is **not** intellectual work, it is fact and information



Copyright protection covers expressions and not ideas, procedures, operating methods or mathematical concepts as such.



Protection is on databases and not on data. Data are protected only and especially when they are collected and organized in a database.

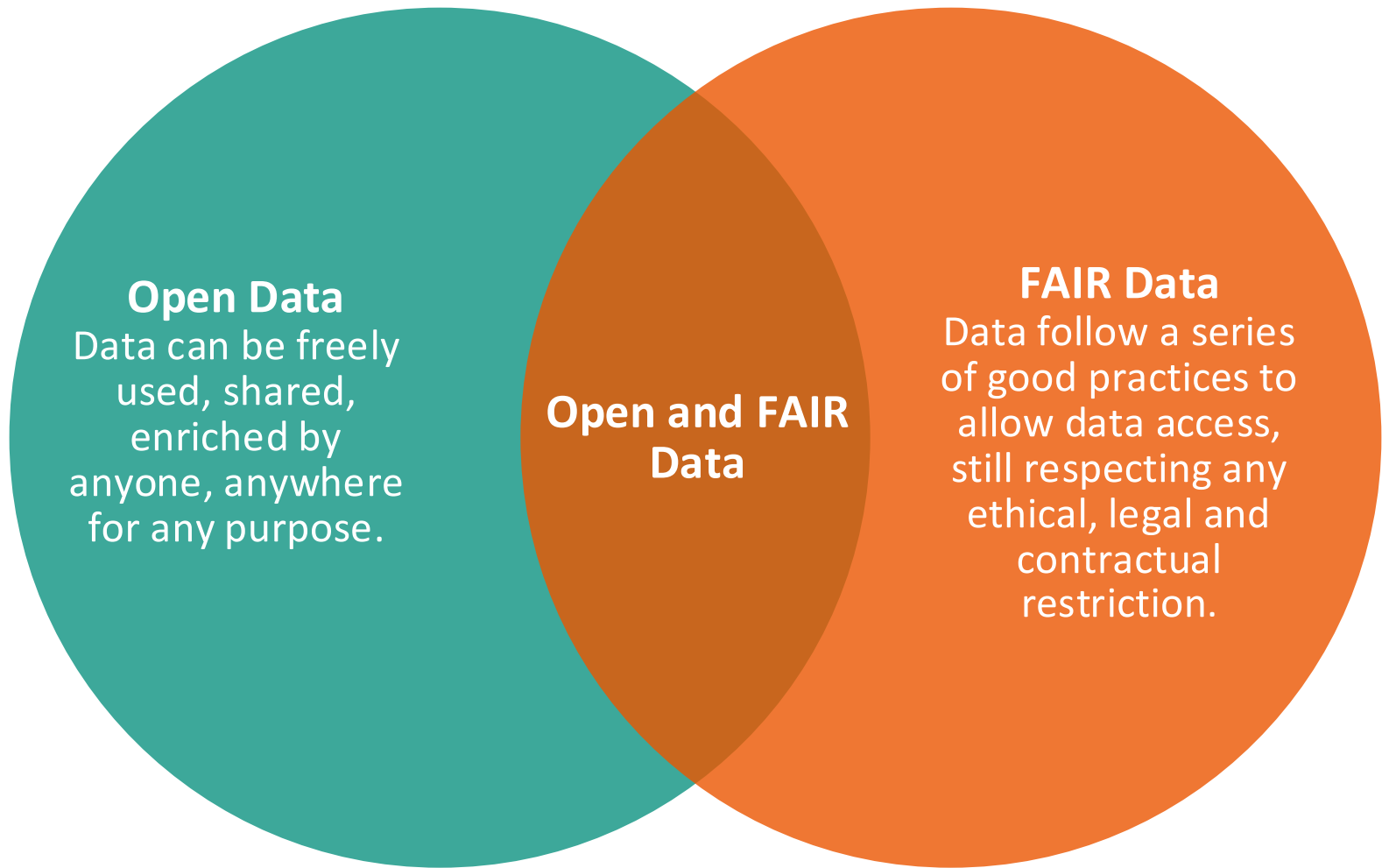


The sui generis property right (only in Europe) covers not only the reproduction and dissemination of the database, but also the extraction and reuse of substantial parts of the database.

Open Data
is simply
a particular kind of
FAIR Data



Photo by [Serhat Beyazkaya](#) on [Unsplash](#)







Why do we need a distinction?

Photo by Possessed Photography on Unsplash



Your research data could:

-  Contain personal information (privacy e GDPR)
-  Fall under copyright (in the case of a database with creative structure)
-  Fall under the Sui Generis right (database obtained thanks to a substantial investment)
-  Be protected by patent or industrial secret

Data sharing needs to respect the specific law.

Data needs to be protected against non authorised access.

How can you adhere to FAIR principles if your data cannot be opened?



Photo by Jon Tyson on Unsplash

Create and share a description of your data

- This way other researchers may **ask for permission to access** your data for reuse purposes, by giving a specific aim and following the rules defined by the law.
- **Restrict access** to the record payload (attachment, files,...)

An important difference



Deposit: upload a digital object (data, articles, ...) on a platform that allows to correctly describe the object through metadata and that implements long-term preservation.



Give access: once the object has been deposited, the authors can choose the type of access that can be granted (open, restricted, closed, embargoed,...) and assigns a licence to reuse the contents (Creative Commons)



Access Right: Open Access

Access right *

- ☒ Open Access
- ☐ Embargoed Access
- ☐ Restricted Access
- ☐ Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

License *

Creative Commons Attribution 4.0 International

Required. Selected license applies to all of your files displayed on the top of the form. If you want to upload some of your files under different licenses, please do so in separate uploads. If you cannot find the license you're looking for, include a relevant LICENSE file in your record and choose one of the *Other* licenses available (*Other (Open)*, *Other (Attribution)*, etc.). The supported licenses in the list are harvested from opendefinition.org and spdx.org. If you think that a license is missing from the list, please [contact us](#).

This should be the default access right

Always assign a licence for reuse

Access Right: Embargoed Access

Access right *

- ☐ Open Access
- ☒ Embargoed Access
- ☐ Restricted Access
- ☐ Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

📅 Embargo date

Required only for Embargoed Access uploads. Format: YYYY-MM-DD. The date your upload will be made publicly available in case it is under an embargo period from your publisher.

● License *

Creative Commons Attribution 4.0 International

Required. Selected license applies to all of your files displayed on the top of the form. If you want to upload some of your files under different licenses, please do so in separate uploads. If you cannot find the license you're looking for, include a relevant LICENSE file in your record and choose one of the *Other* licenses available (*Other (Open)*, *Other (Attribution)*, etc.). The supported licenses in the list are harvested from opendefinition.org and spdx.org. If you think that a license is missing from the list, please [contact us](#).

Use it when you have a **valid reason** to delay access

Always assign a licence for reuse

Note: metadata is always accessible to everyone


Access Right: Restricted Access

Access right *

- ☐ Open Access
- ☐ Embargoed Access
- ☒ Restricted Access
- ☐ Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

Conditions *



Use it when you have a **valid reason** to restrict the access

Always specify conditions under which you grant access (who, how, why can get access to your payload)

Note: metadata is always accessible to everyone

Specify the conditions under which you grant users access to the files in your upload. User requesting access will be asked to justify how they fulfil the conditions. Based on the justification, you decide who to grant/deny access. You are not allowed to charge users for granting access to data hosted on Zenodo.

Access Right: Closed Access

Access right *

- ☐ Open Access
- ☐ Embargoed Access
- ☐ Restricted Access
- ☒ Closed Access

Required. Open access uploads have considerably higher visibility on Zenodo.

Are you really sure you need closed access?
consider restricted or embargoed access instead!



How do I know if my research data is
protected?



Some consideration on data protection

- Copyright is a property right in certain types of original literary, artistic and scientific works.
- Copyright does not protect ideas.
- Confidentiality protects confidential information. This might be imposed by a contract or if the information is marked confidential. Use of confidential information might give rise to a claim for compensation if confidentiality is breached.
- Data Subject Rights arise in information that identifies individuals and are recognised by data protection laws in the EU.
- Patents are registered rights in novel inventions of products or processes.
- Some research data may not benefit from any legal protection, although moral and ethical considerations may apply.

Data is not yours!

- Data is **not** intellectual work, no copyright applies!
- **Copyright** protection covers expressions and not ideas, procedures, operating methods or mathematical concepts as such.
- **Protection is on databases and not on data.** The data are protected only and especially when they are collected and organized in a database.
- **The sui generis database right (only in Europe)** covers not only the reproduction and dissemination of the database, but also the extraction and reuse of substantial parts of the database.



11 Dicembre 2019

Data governance: un dato non appartiene a nessuno... a meno che sia personale

di [Simone Aliprandi](#)

Quando un nostro dato è personale? Come è giusto tutelarla? La risposta deve comprendere due punti di vista, quello della proprietà intellettuale e quello della privacy.

CONDIVIDI

L'autore

[Simone Aliprandi](#)
Simone Aliprandi ha un dottorato di ricerca in Sociologia dell'Informazione ed è un avvocato che si occupa di consulenza, ricerca e formazione nel campo del diritto d'autore e più in generale del diritto dell'ECU. È responsabile del progetto europeo di ricerca, il network dei network, legge e collabora come docente con alcuni istituti universitari, ha pubblicato articoli a libri sul mondo delle tecnologie open e della cultura libera, rilasciando tutto le sue opere con licenze di tipo copyleft.

[Site & blog](#)

Non c'è solo la privacy, quando si parla di dati e di diritto

Si sente spesso parlare di tutela del dato o risolutezza del dato, soprattutto in questi ultimi due/tre anni in cui temi come [big data](#) e open data sono diventati di pubblico dominio e in cui l'entrata in vigore del GDPR ([il nuovo regolamento europeo sui dati personali](#)) ha portato un'ondata - per certi versi ridondante - di corsi di formazione, consulenze, articoli sul tema della protezione dei dati.

Mi occupo di consulenza e formazione proprio in quest'ambito e mi rendo conto che spesso tra gli utenti non c'è piena consapevolezza di come il diritto considera e tratta i dati. Non soprattutto che alle parole tutela del dato o risolutezza del dato **la gente pensa automaticamente all'ambito della privacy**, della tutela del dato personale.

Per inquadrare il tema correttamente e in modo completo, è tuttavia necessario tenere in debita considerazione anche il punto di vista della cosiddetta proprietà intellettuale, punto di vista che a me sta particolarmente a cuore. Anche perché quando acquistiamo, stiamo, diffondiamo dei dati **non è detto che siano dati personali** e dunque non sempre le norme sulla privacy (GDPR e simili) entrano in gioco. Cerchiamo di capire meglio la questione.

Nessun copyright su idee e dati

Inanzitutto: **non esiste alcun diritto di proprietà intellettuale sul dato in sé**. I dati nudi e crudi e le informazioni che da essi si deducono non sono oggetto di alcun tipo di proprietà intellettuale. Questo in virtù di uno dei principi cardine del diritto d'autore secondo cui il diritto tutela non l'informazione, bensì la specifica forma espressiva con cui l'informazione è presentata. Basti leggere il testo dell'articolo 9, numero 2 dell'[Accordo TRIPS](#).



Data and law protection

- **Raw data** are not protected by copyright
- **Database** is defined as a collection of independent works, data or other materials arranged in a systematic or methodical way
- **Copyright** protects the **structure, selection or arrangement** of the database contents, not the data
- **Sui generis database right**: protects the substantial effort in obtaining data (not creating). Note: the **right owner** is often the institution.

Authors and rights owners

Are you the author of the data you collected?

Yes, in case you can prove it (deposit with clear date, DOI, ... use a data repository!)

Do you own any rights on the raw data you collected?

No, data is facts/information and none can own rights on it!





Licenses

Tell other what they
can do with your data

Creative Commons

Not all of us are legal experts capable of writing proper licenses.

Creative Commons and Public Domain create legal certainty for everyone, who wants to use works, that are licensed respectively.

It is important to follow and understand the different meanings of the licenses and follow the rules for using them.

CREATIVE COMMONS LICENSES		COPY & PUBLISH	ATTRIBUTION REQUIRED	COMMERCIAL USE	MODIFY & ADAPT	CHANGE LICENSE
	PUBLIC DOMAIN	✓	✗	✓	✓	✓
	CC BY	✓	✓	✓	✓	✓
	CC BY-SA	✓	✓	✓	✓	✗
	CC BY-ND	✓	✓	✗	✗	✓
	CC BY-NC	✓	✓	✗	✓	✓
	CC BY-NC-SA	✓	✓	✗	✓	✗
	CC BY-NC-ND	✓	✓	✗	✗	✓
You can redistribute (copy, publish, display, communicate, etc.)		You have to attribute the original work		You can use the work commercially		You can modify and adapt the original work
				You can choose license type for your adaptations of the work.		

[CC-Infographic](#) by [Foter](#) under [CC-BY-SA 3.0](#) license.



Types of CC Licenses

Public Domain

Works are not covered by copyright

CC-0 (no rights reserved)

Allows creators to give up their copyright and put their works into the worldwide public domain

CC-BY (Attribution)

This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator



Types of CC Licenses

CC-BY-SA (Attribution – ShareAlike)

This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. The license allows for commercial use. If you remix, adapt, or build upon the material, you must license the modified material under identical terms.

CC-BY-ND (Attribution – NonDerivative)

This license allows reusers to copy and distribute the material in any medium or format in unadapted form only, and only so long as attribution is given to the creator. The license allows for commercial use.

CC-BY-NC (Attribution – NonCommercial)

This license allows reusers to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator.

Licensing your Research Data: Creative Commons

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.

<https://doi.org/10.5281/zenodo.840651>

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.

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](#).

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.

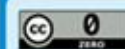
You should keep in mind that there are many situations in which data is *not* protected as a matter of law. Such data can include facts, names, numbers - things that are considered 'non-original' and part of the public domain thus not subject to copyright protections. Similarly, your database (which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

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

Licensing your Research Data: Creative Commons

- Use a CC0 or public domain, then ask for credit
- Provide a citation that researchers using your data can simply copy and paste to give you credit for your work
- Remember that it's bad science not to cite the source
- CC0 does not mean academic unpoliteness

<https://doi.org/10.5281/zenodo.840651>

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 SORS 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 rights 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.

use a CC BY licence, then you just accept it and 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 from selling your dataset to others for a profit. However, it might also prohibit someone using the dataset in their research if they intend eventually to publish that research. This is the case in most academic journals as commercial businesses that charge some sort of fee for access to their content, hence, such use could qualify as 'commercial'. Consequently, using a non-commercial licence prevents researchers from using data in work destined for publication. This subsequently affects the dissemination, retention, and impact of your dataset.

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

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.


**Any comments or
questions?**



- Go to www.menti.com and insert code:
4758 1389

OR

- use your mobile/tablet to scan the QR code

thank you!

emma.lazzeri@cnr.it

This presentation is released under a CC-BY 4.0 Licence
To cite this presentation, please copy and paste

Lazzeri, E. (2026, febbraio 23). Research Data Management and FAIR Principles. Zenodo. <https://doi.org/10.5281/zenodo.18745984>

DOI: **10.5281/zenodo.18745984**

References



- ❑ Components of Open Science, UNESCO 2020, https://en.unesco.org/sites/default/files/open_science_brochure_en.pdf
- ❑ UNESCO Recommendation on Open Science, 2021, <https://doi.org/10.54677/MNMH8546>
- ❑ UCL Research Data Policy, <https://www.ucl.ac.uk/library/research-support/research-data-management>
- ❑ C. Ingram, How and why you should manage your research data: a guide for researchers. An introduction to engaging with research data management processes, JISC Guides, <https://www.jisc.ac.uk/guides/how-and-why-you-should-manage-your-research-data>
- ❑ [B. Kramer, J. Bosman, Vienna Open Science workshop, 17-09-2020, https://docs.google.com/presentation/d/1iw7zN_VSQmzTI81GnVOltrjuvI2_1mOeXXIH3EFx4A8/edit#slide=id.p15
- ❑ K. den Heijer, CTG Data Management training day 1, 10.5281/zenodo.4048591
- ❑ Fifty Shades of No, January 2015, https://philarcher.org/diary/2015/50shadesofno/?fbclid=IwAR0gztaEKljyn89n1azf4dFrYami_Ohb-QSRKm_XpIRyzvLRRB929uCsckg
- ❑ Ghent University guide on Research Data Management: <https://www.ugent.be/en/research/datamanagement>

References

- ❑ Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 3, 160018 (2016). <https://doi.org/10.1038/sdata.2016.18>
- ❑ Martínez-Lavanchy, P.M., Hüser, F.J., Buss, M.C.H., Andersen, J.J., Begtrup, J.W. (2019). 'FAIR Principles'. In: Holmstrand, K.F., den Boer, S.P.A., Vlachos, E., Martínez-Lavanchy, P.M., Hansen, K.K. (Eds.), Research Data Management (eLearning course). <https://doi.org/10.11581/dtu:00000049>

References

- ❑ Simone Aliprandi, Data governance: un dato non appartiene a nessuno... a meno che sia personale, 11 Dicembre 2019, https://www.apogeeonline.com/articoli/data-governance-un-dato-non-appartiene-a-nessuno-a-meno-che-sia-personale-simone-aliprandi/?fbclid=IwAR2LaXV6b6m8aaJ7SeFKVAidbTIB9C4_-mQPirLxs3m5DxAs0jG8gkn8ZKg
- ❑ OpenAIRE Guidelines on data protection: <https://www.openaire.eu/how-do-i-know-if-my-research-data-is-protected>
- ❑ OpenAIRE Guideline: How do I license my research data? <https://www.openaire.eu/how-do-i-license-my-research-data>
- ❑ OpenAIRE Guideline: Can I reuse someone else's research data? <https://www.openaire.eu/can-i-reuse-someone-else-research-data>
- ❑ CC-Infographic by Foter under CC-BY-SA 3.0 license.
- ❑ Creative Commons UK. (2017). Fact Sheet on Creative Commons & Open Science. Zenodo. <https://doi.org/10.5281/zenodo.840652>

References

- ❑ DMP Tool: https://dmptool.org/general_guidance#metadata-data-documentation
- ❑ How to Guides DCC: <https://www.dcc.ac.uk/guidance/how-guides>
- ❑ Five steps to decide what data keep: <https://www.dcc.ac.uk/guidance/how-guides/five-steps-decide-what-data-keep>
- ❑ Adapt your Data Management Plan. A list of Data Management Questions based on the Expert Tour Guide on Data Management: https://www.cessda.eu/content/download/4302/48656/file/TTT_DO_DMPExpertGuide_v1.2.pdf
- ❑ Practical Guide to the International Alignment of Research Data Management - Extended Edition, 2021, <https://doi.org/10.5281/zenodo.4915861>
- ❑ OpenAIRE guide on cost in research data management, <https://www.openaire.eu/how-to-comply-to-h2020-mandates-rdm-costs>
- ❑ Making Your Code Citable, <https://guides.github.com/activities/citable-code/>
- ❑ Setting up an Organised Folder Structure for Research Projects, http://nikola.me/folder_structure.html
- ❑ TILS Document Naming Convention, http://www.data.cam.ac.uk/files/gdl_tilsdocnaming_v1_20090612.pdf
- ❑ FAIRsharing, <https://fairsharing.org/standards/>

References

- ❑ RDA, Metadata Standards Directory Working Group, <http://rd-alliance.github.io/metadata-directory>
- ❑ Ghent University web guide on Research data management: <https://www.ugent.be/en/research/datamanagement>
- ❑ Reading University web guide on RDM: <https://www.reading.ac.uk/research-services/research-data-management>
- ❑ Francesca Di Donato, & Emma Lazzeri. (2021, October 18). Data Management. Zenodo.
<https://doi.org/10.5281/zenodo.5593104>
- ❑ OpenAIRE infographic on RDM costs: <https://www.openaire.eu/rdm-researcher-costs-infographic/view-document>
- ❑ Cambridge University, Data Management Guide: <https://www.data.cam.ac.uk/data-management-guide>

Useful links

- **Zenodo - CERN-OpenAIRE OA repository - catch all**

www.zenodo.org

- **Choose a license - Creative Commons**

<https://creativecommons.org/choose/?lang=en>

<https://chooser-beta.creativecommons.org/>

- **DMP examples by subject - LIBER**

<https://libereurope.eu/dmpcatalogue/>

- **Tools to create your DMP**

<https://www.openaire.eu/argos/>

<https://dmponline.dcc.ac.uk/>

<https://argos.openaire.eu/splash/>

- **Re3Data**

<https://www.re3data.org/>

- **Metadata standard Directory - Research Data Alliance**

<https://rd-alliance.github.io/metadata-directory/>

Figure List

1. Photo by [Serhat Beyazkaya](#) on [Unsplash](#)
2. Photo by [Jon Tyson](#) on [Unsplash](#)
3. Photo by [Possessed Photography](#) on [Unsplash](#)
4. Photo by [Jason Dent](#) on [Unsplash](#)
5. Photo by [Peter Herrmann](#) on [Unsplash](#)
6. Photo by [Maria Ziegler](#) on [Unsplash](#)
7. Photo by [Tsvetoslav Hristov](#) on [Unsplash](#)
8. Photo by [Quino Al](#) on [Unsplash](#)