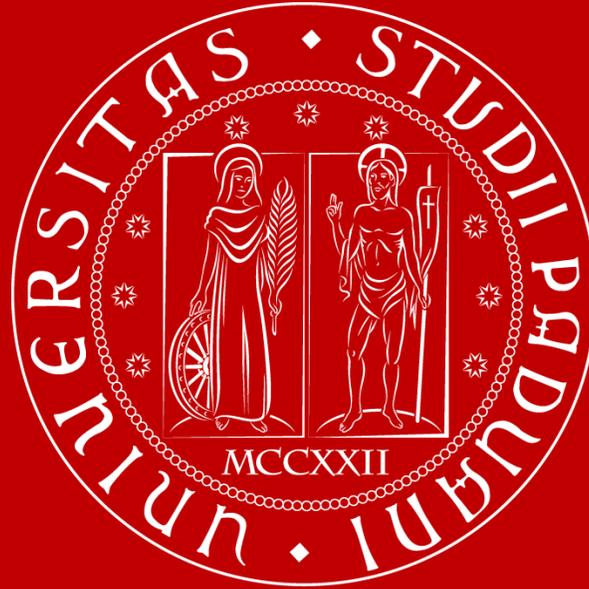


1222-2022
800
ANNI



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

**PHD EDUCATIONAL WEEK ON
TRANSFERABLE SKILLS**

PADOVA , 21-25 June 2021



SISTEMA BIBLIOTECARIO
DI ATENEIO

II.

From Open Access to Open Data: the Open Science framework



SISTEMA BIBLIOTECARIO
DI ATENEIO

Padova University Library System

Open Access to research data

Research data

Definition, categories, raw and primary data, data lifecycle

Data management: main steps

Collect, name, structure, annotate, file formats

Data Management: more tips

Preservation, privacy, GDPR

How to manage Open Data and why

Data Management Plan, Open Licenses, FAIR principles, general advantages

Who asks for Open research data?

European projects, EOSC, UniPd Policy

Where to archive research data

Research Data Unipd, other repositories



Open Science? Echoes of the COVID-19 outbreak ...

THE LANCET
Digital Health



eurostat
Your key to European statistics

News Data

European Commission > Eurostat

COVID-19

Sign In | Register

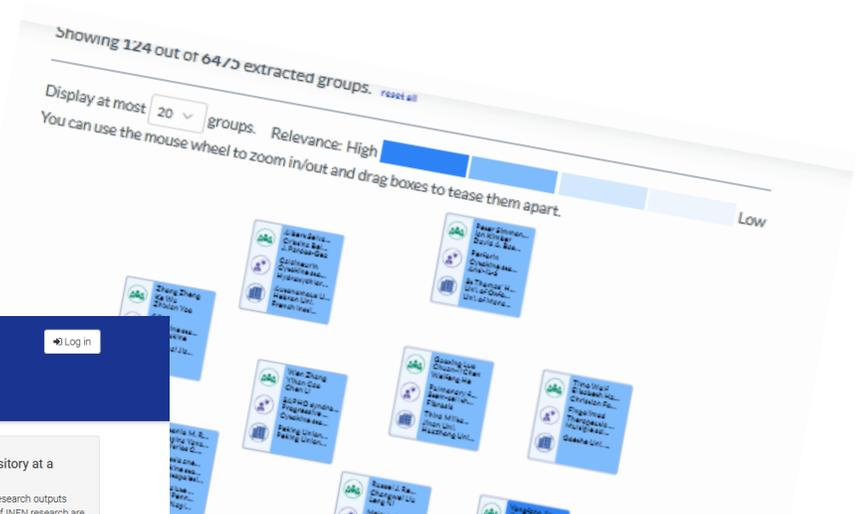
English (en)

Type a keyword, a publication title, a dataset title...

News & Events

Resources

About Eurostat Help



Showing 124 out of 64/3 extracted groups.

Display at most 20 groups. Relevance: High Low

You can use the mouse wheel to zoom in/out and drag boxes to tease them apart.

network of top proteins/genes/cells associate



EUROPEAN OPEN ACCESS REPOSITORY

Impact & Studies

What we do

Order by: Last Modified

Filter by location

64 datasets found

COVID-19 Coronavirus data

COVID-19 Coronavirus surveillance

The dataset provides surveillance COVID-19 Coronavirus.

Settings

Operator AND OR

Countries

Czechia	22
Spain	7
United Kingdom	7
Italy	2
EU institutions	

+ Energy

+ Transport

+ Tourism



INFN OAR

Open Access Repository

Search

Upload Communities

Log in

Latest entries

September 20, 2020 (v167) CovidStat project summary plots

This record contains the daily CovidStat Working Group Physics with the aim of making

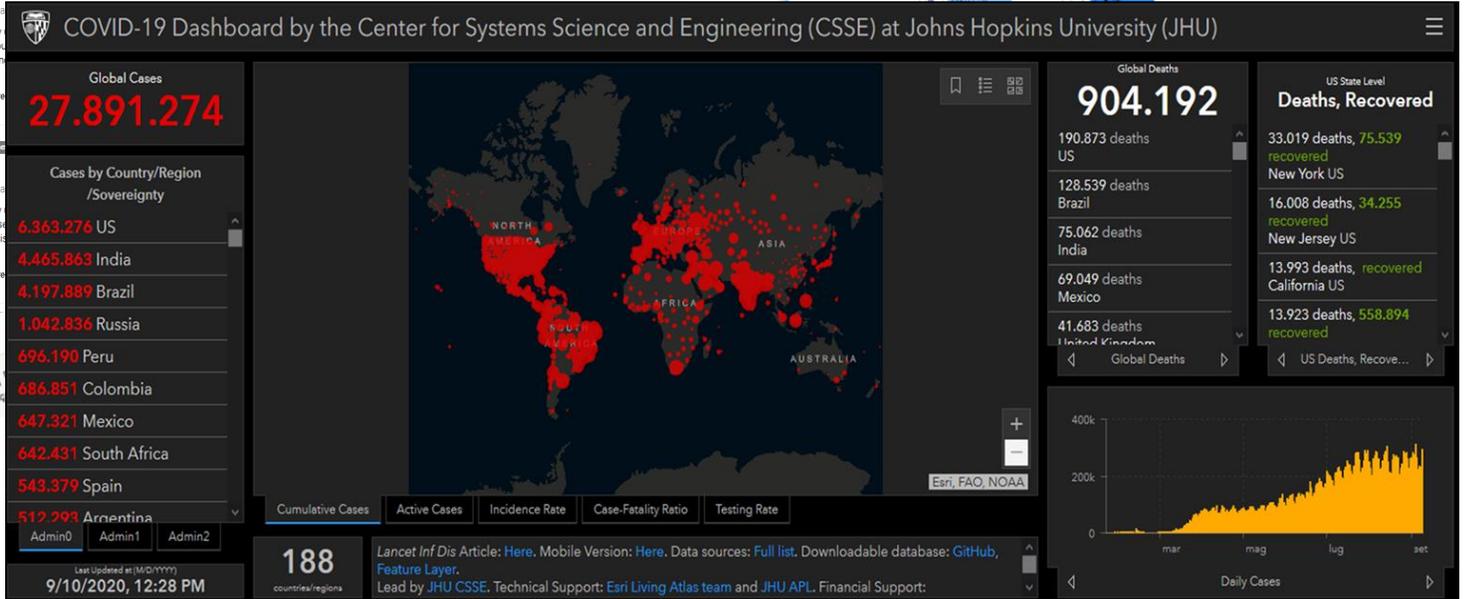
Updated on September 21, 2020 166 more version(s) exist for this record

September 20, 2020 (v166) CovidStat project data

Updated on September 21, 2020 165 more version(s) exist for this record

INFN Open Access Repository at a glance

- Research. Shared. — all research outputs from across all domains of INFN research are



COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)

Global Cases: 27.891.274

Cases by Country/Region /Sovereignty

6.363.276	US
4.465.863	India
4.197.889	Brazil
1.042.836	Russia
696.190	Peru
686.851	Colombia
647.321	Mexico
642.431	South Africa
543.379	Spain
512.293	Argentina

188 countries/regions

Global Deaths: 904.192

US State Level Deaths, Recovered

190.873 deaths	US
128.539 deaths	Brazil
75.062 deaths	India
69.049 deaths	Mexico
41.683 deaths	United Kingdom
33.019 deaths, 75.539 recovered	New York US
16.008 deaths, 34.255 recovered	New Jersey US
13.993 deaths, recovered	California US
13.923 deaths, 558.894 recovered	

Cumulative Cases Active Cases Incidence Rate Case-Fatality Ratio Testing Rate

Daily Cases

Esri, FAO, NOAA

Lancet Inf Dis Article: Here. Mobile Version: Here. Data sources: Full list. Downloadable database: GitHub, Feature Layer.

Lead by JHU CSSE. Technical Support: Esri Living Atlas team and JHU APL. Financial Support:

Genetic Sequence Database Data on SARS-CoV-2 sequences can be found in GenBank® and the NIH Sequence Read Archive (SRA).



NCBI Virus Portal that collects virus sequence data from RefSeq, GenBank and other NCBI (National Library of Medicine) archives

Coronavirus Disease Research Community - COVID-19

Zenodo is a general-purpose open-access repository developed under the European OpenAIRE program and operated by CERN and create



Open Data
European Union

<https://www.covid19dataportal.org/>



GISAID, an Initiative that promotes the international sharing of all influenza virus sequences, related clinical and epidemiological data associated with human viruses

Covid-19 opendata On GitHub, many examples of data reworking



The real-time sharing of research publications, software and data to fight COVID-19 is unprecedented

The World Health Organization opens a process of Data sharing for novel coronavirus (COVID-19)



The Chan Zuckerberg Initiative foundation has decided also to provide access to data related: New Dataset Makes Coronavirus Research Open and Machine Readable - Chan Zuckerberg Initiative

Open Access

“Open access (OA) literature is digital, online, free of charge, and free of most copyright and licensing restrictions”

Peter Suber (2012), Open Access, MIT Press
[https://cyber.harvard.edu/hoap/Open_Access_\(the_book\)](https://cyber.harvard.edu/hoap/Open_Access_(the_book))



Open
Access

OA refers particularly to academic, technical and scientific contents

It is a type of (open) access to digital content

There are several [editorial models](#) and licenses compatible with OA, in constant evolution

Any type of digital content can be openly accessible

Open Data

Open Data are online, free of cost, accessible data that can be used, reused and distributed, provided that the data source is attributed.

Open Data



It is the philosophy of Open Access applied to data

Data are open when **anyone can access, use and share.**

Examples:

- government open data (e.g. open by default according to the [Italian Digital Administration Code](#))
- research data available to citizens

Open Data: accessible, reusable

- Data must be accessible both to users of the relevant scientific community and to ordinary citizens (citizen science)

Accessible data

Open Data

- Data are open if they can be freely consulted, used, modified, extracted and shared by anyone and for any purpose

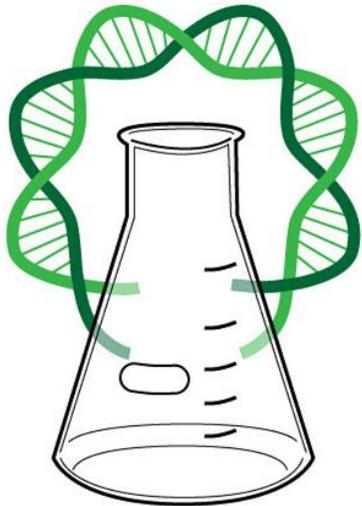
- [Checklist](#): How much open are your data?
- [Codata] [Legal Interoperability of Research Data: Principles and Implementation Guidelines](#)

Tools

Open Science

“Open science is the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society”

FOSTER consortium



Open
Science

Open Data
Open Source in Open Science
Open Methodology
Open Peer Review
Open Access
Open Educational Resources

Open Access to research data

Research data

Data management: main steps

Data Management: more tips

How to manage Open Data and why

Who asks for Open research data?

Where to archive research data



What are research data?

Recorded **information** (regardless of the form or the media in which they may exist) **necessary to support or validate a research project's observations, findings or outputs**



Digital copies
of images



GIS and
spatial
data



Spreadsheets



Digital texts or digital
copies of text



Databases



Audio



Diagrams



Protein or genetic
sequences



Video

BUT ALSO...

- Computer-aided Design (CAD) outputs
- Waveforms
- Computer codes
- Statistics (SPSS, SAS)
- Matlab files
- Artistics products
- Web files
- ...

Raw data, primary data

Raw data have been collected or generated in the course of research, but have not been analysed or manipulated yet.

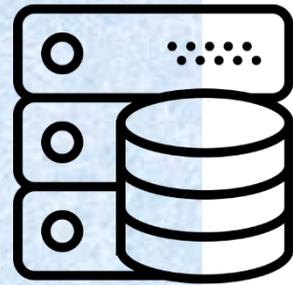
Primary data have been collected in the first person through direct observation, recording, measurement.



General categories of data

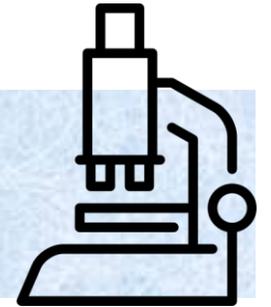
Derived or compiled:

Derived data involves using existing data points to create new data (e.g. compiled databases, text or data mining); reproducible but expensive



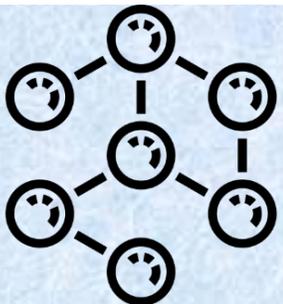
Observational: Observational data are captured through observation of a behavior or activity (e.g. sensor readings, survey instruments); usually irreplaceable and not replicable

Experimental: Experimental data are collected through active intervention by the researcher (e.g. gene sequences, magnetic fields); generally reproducible but expensive

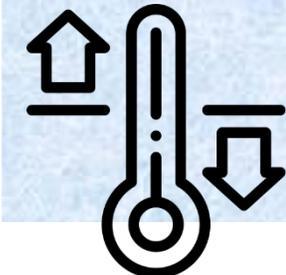


Reference

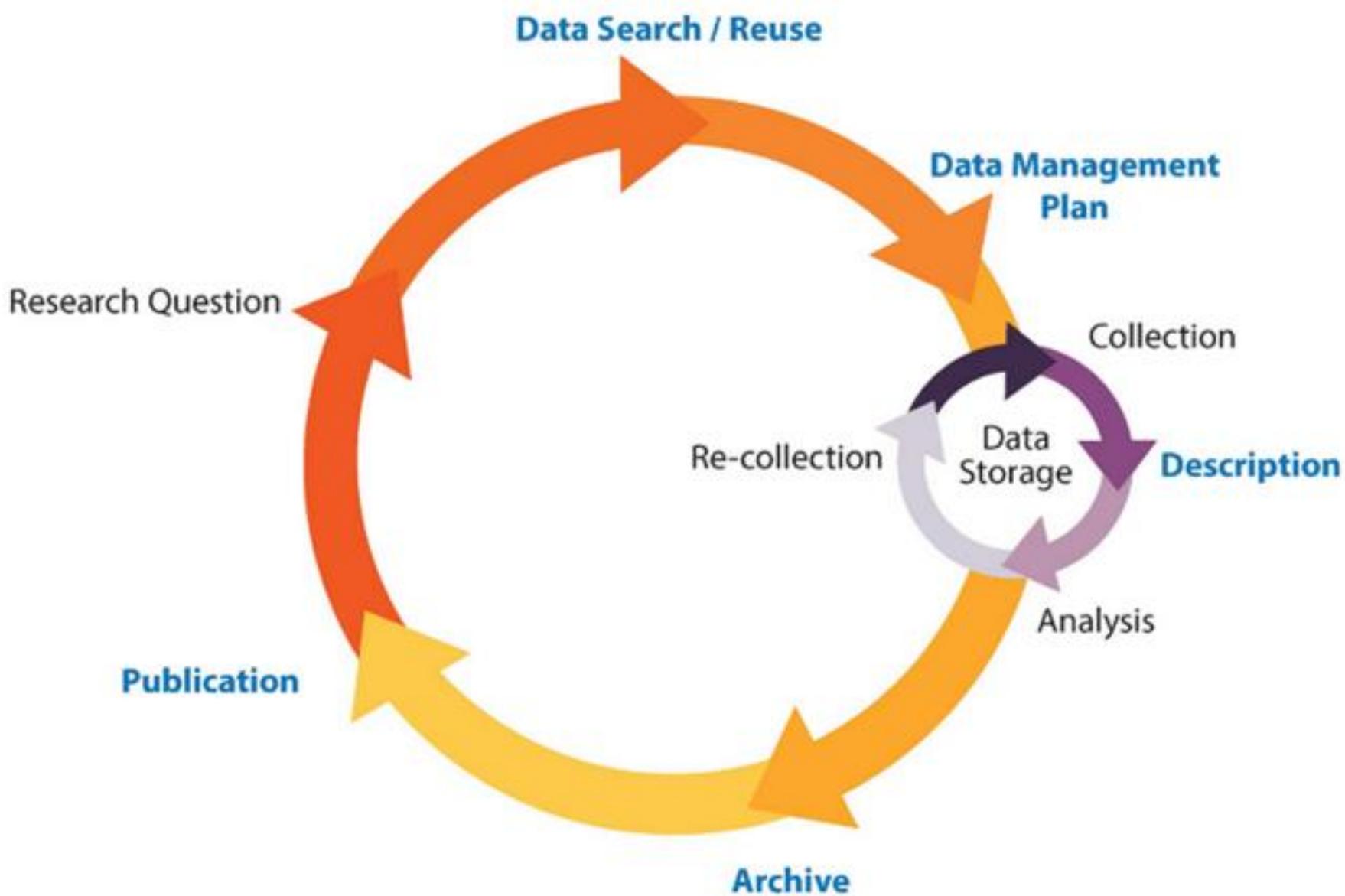
(e.g. gene sequences databases, chemical structures, portals with spatial data)



Simulation : Simulation data are generated by imitating the operation of a real-world process or system over time using computer test models (e.g. climate models); not always replicable

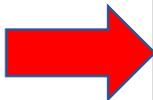


Research data lifecycle

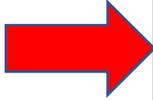


Research data lifecycle

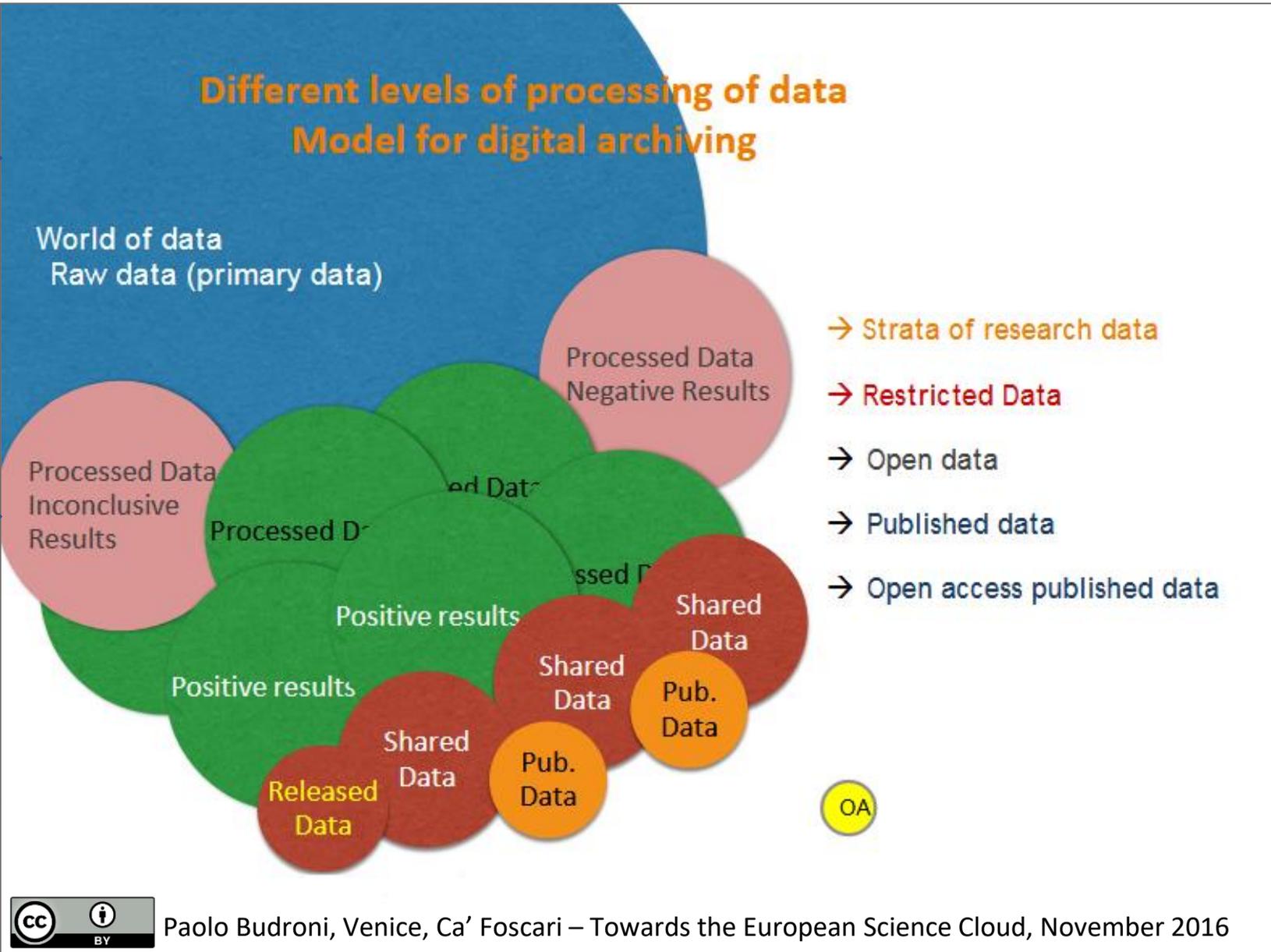
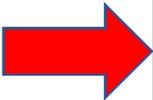
Raw / primary data are collected or generated during the research, but they are not yet analyzed or manipulated.



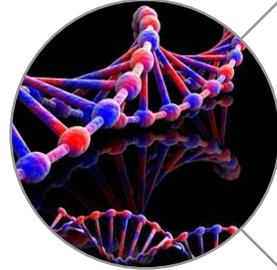
Data is then processed and analyzed, and they can lead to positive, negative or inconclusive results.



Only a very small part of data collected during a research comes to be included in a publication.



Relationship between data & articles & reported findings



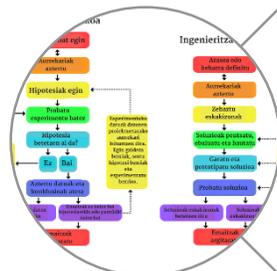
in a paper announcing the sequencing of an entire genome, the sequence would be a central aspect of the paper



data are necessary to support the claims of the paper and essential to enable a knowledgeable peer to reproduce and verify the results



data or databases provide a background to a publication: without them the findings or conclusions could not have been derived



background information might be important to make comparisons with alternative methods or sources of data

Raw Data: example from a journal

CLINICAL CANCER RESEARCH

Home About Articles For Authors Alerts News Search Q

Clinical Trials: Targeted Therapy

Phosphorylated Acetyl-CoA Carboxylase Is Associated with Clinical Benefit with Regorafenib in Relapsed Glioblastoma: REGOMA Trial Biomarker Analysis

Stefano Indraccolo, Gian Luca De Salvo, Roberta Rudà, Alba Ariela Brandes, Ton

DOI: 10.1158/1078-0432.CCR-19-4055

Article **Figures & Data**

ARTICLE FIGURES & DATA

Figures

Additional Files

SUPPLEMENTARY DATA

Figure S1 - Immunohistochemical staining of pACC in three representative GBM samples

Figure S2 - Immunohistochemical staining of three markers (MCT4, pAMPK and pACC) showing their expression in peri-necrotic areas of GBM samples.

Figure S3 - Kaplan-Meier curves of overall survival (top) and progression according to pAMPK status

Table S1 - Digital pathology raw data

Table S2 - MVD values in GBM samples

SUPPLEMENTARY DATA

SUPPLEMENTARY DATA

- Figure S1 - Immunohistochemical staining of pACC in three representative GBM samples.
- Figure S2 - Immunohistochemical staining of three markers (MCT4, pAMPK and pACC) showing their expression in peri-necrotic areas of GBM samples.
- Figure S3 - Kaplan-Meier curves of overall survival (top) and progression according to pAMPK status
- Table S1 - Digital pathology raw data
- Table S2 - MVD values in GBM samples

Figures Tables

Marica Eoli,

2020

Suppl. Table 1. Digital pathology values of biomarkers evaluated by IHC in GBM samples

Sample ID	pACC				Sample ID	pAMPK				Sample ID	MCT4				Sample ID	MCT1			
	0+	1+	2+	3+		0+	1+	2+	3+		0+	1+	2+	3+		0+	1+	2+	3+
133-1	53,42	0,00	46,31	0,27	133-1	78,56	20,21	1,13	0,10	133-1	35,01	17,60	41,55	5,85	133-1	1,70	26,21	49,32	22,77
133-2	99,34	0,00	0,64	0,02	133-2	99,27	0,25	0,23	0,25	133-2	56,05	18,97	24,86	0,12	133-2	5,20	26,58	60,87	7,35
133-3	99,65	0,00	0,31	0,04	133-3	98,99	0,73	0,24	0,04	133-3	95,98	3,55	0,47	0,01	133-3	92,95	3,41	3,64	0,00
133-4	87,20	0,00	12,76	0,00	133-4	98,23	1,51	0,19	0,07	133-4	75,92	10,49	13,55	0,04	133-4	43,55	20,67	35,61	0,17
133-6	95,80	0,00	3,83	0,37	133-6	99,00	0,47	0,34	0,19	133-6	53,73	18,37	27,78	0,11	133-6	11,12	27,84	59,41	1,63
133-7	57,13	0,00	42,49	0,38	133-7	87,37	11,84	0,59	0,20	133-7	1,09	27,68	55,18	16,06	133-7	3,31	26,42	57,47	12,80
133-8	95,55	0,00	4,35	0,09	133-8	97,14	1,79	0,64	0,43	133-8	34,48	24,20	40,80	0,52	133-8	29,79	23,95	44,49	1,77
133-9	99,88	0,00	0,10	0,01	133-9	96,85	1,37	1,14	0,64	133-9	81,48	8,62	9,79	0,10	133-9	20,89	25,87	50,30	2,94
133-10	97,26	0,00	2,71	0,03	133-10	94,08	5,73	0,19	0,00	133-10	9,38	25,56	42,99	22,07	133-10	10,39	32,05	56,04	1,53
133-11	76,76	0,00	23,23	0,01	133-11	95,14	3,88	0,66	0,31	133-11	NV				133-11	19,44	30,31	45,47	4,78
133-12	87,56	0,00	12,42	0,03	133-12	99,10	0,81	0,08	0,02	133-12	45,84	21,60	32,30	0,29	133-12	7,20	28,50	55,05	9,26
133-13	86,48	0,00	13,47	0,05	133-13	98,82	0,90	0,12	0,17	133-13	64,76	14,67	20,27	0,10	133-13	25,35	24,65	48,66	1,34
133-14	92,66	0,00	7,31	0,02	133-14	99,89	0,09	0,01	0,01	133-14	85,10	8,78	6,03	0,25	133-14	61,92	16,05	21,80	0,23
133-15	47,00	0,00	51,61	1,39	133-15	97,98	1,61	0,26	0,14	133-15	36,82	22,55	40,39	0,40	133-15	8,88	28,94	56,87	5,30
133-16	94,74	0,00	5,12	0,14	133-16	82,32	15,76	1,81	0,11	133-16	72,63	11,74	15,24	0,12	133-16	90,55	4,77	4,66	0,01
133-17	100,00	0,00	0,00	0,00	133-17	99,40	0,42	0,12	0,05	133-17	84,06	11,01	4,81	0,42	133-17	93,52	4,17	2,23	0,08
133-18	73,06	0,00	26,82	0,13	133-18	24,47	43,71	27,02	4,81	133-18	66,66	19,16	13,76	0,72	133-18	18,36	26,17	48,86	6,60
133-19	80,22	0,00	19,76	0,02	133-19	95,32	2,29	1,40	0,99	133-19	85,89	9,41	3,99	0,97	133-19	98,58	1,07	0,36	0,00
133-20	80,77	0,00	18,87	0,36	133-20	94,00	4,25	1,50	0,25	133-20	51,73	18,64	28,66	0,03	133-20	89,25	5,14	5,48	0,13
133-21	76,51	0,00	23,37	0,12	133-21	88,67	11,05	0,26	0,02	133-21	74,82	11,74	13,41	1,57	133-21	97,93	1,86	0,20	0,01
133-22	57,26	0,00	41,69	1,05	133-22	96,27	3,14	0,47	0,12	133-22	44,61	19,62	34,21	1,89	133-22	23,64	26,55	43,88	5,93
133-23	98,21	0,00	1,79	0,00	133-23	95,08	4,04	0,30	0,58	133-23	96,82	2,64	0,48	0,20	133-23	87,70	6,59	5,65	0,06
133-24	92,55	0,00	7,43	0,01	133-24	48,04	38,76	11,43	1,76	133-24	91,28	4,14	4,38	0,20	133-24	93,92	3,88	2,07	0,14
133-25	91,71	0,00	8,28	0,01	133-25	79,49	0,00	20,26	0,25	133-25	47,13	21,79	30,67	0,41	133-25	33,05	23,33	40,61	3,01

Increasingly more journals require that raw data are supplied along with the articles.

Alternatively a link may be given to an open data repository or, if the data are not directly accessible, a contact may be given to a researcher who will grant permission for their retrieval.

Data & article reported findings

Original Research | Published: 19 June 2020

The Political Economy of Football: Democracy, Income Inequality, and Men's National Football Performance

[Kin-Man Wan](#) , [Ka-U Ng](#) & [Thung-Hong Lin](#)

Social Indicators Research **151**, 981–1013(2020) | [Cite this article](#)

242 Accesses | **4** Altmetric | [Metrics](#)

Article:
<https://link.springer.com/article/10.1007/s11205-020-02410-y#article-info>

Table 12 Ranking of average \ln FIFA score points by country, 1999–2014 ($FIFA = \ln FIFA$; $Years = Association\ years$)

From: [The Political Economy of Football: Democracy, Income Inequality, and Men's National Football Performance](#)

Rank	Country	Abb.	FIFA	Years	Rank	Country	Abb.	FIFA	Years	Rank	Country	Abb.	FIFA	Years
1	Spain	ESP	7.034	105	23	Chile	CHL	6.574	119	45	Hungary	HUN	6.377	113
2	Brazil	BRA	6.992	100	24	Nigeria	NGA	6.569	69	46	Honduras	HND	6.377	79
3	Germany	DEU	6.970	114	25	Paraguay	PRY	6.558	108	47	Senegal	SEN	6.349	54
4	Argentina	ARG	6.961	121	26	Ukraine	UKR	6.554	23	48	Mali	MLI	6.345	54
5	Netherlands	NLD	6.947	125	27	Japan	JPN	6.539	93	49	Morocco	MAR	6.334	59
6	Italy	ITA	6.898	116	28	Ecuador	ECU	6.523	89	50	Peru	PER	6.296	92
7	Portugal	PRT	6.863	100	29	Belgium	BEL	6.517	119	51	Finland	FIN	6.290	107
8	United Kingdom	GBR	6.863	151	30	Ghana	GHA	6.512	57	52	South Africa	ZAF	6.280	23
9	France	FRA	6.851	95	31	Norway	NOR	6.508	112	53	Austria	AUT	6.265	110
10	Croatia	HRV	6.770	102	32	Ireland	IRL	6.500	93	54	Venezuela, RB	VEN	6.257	88

Table: <https://link.springer.com/article/10.1007/s11205-020-02410-y/tables/12>

Data & article reported findings



Research Article | Published: 12 March 2020

Patterns of trends in niveograph characteristics across the western United States from snow telemetry data

[S. R. Fassnacht](#)  & [J. I. López-Moreno](#)

[Frontiers of Earth Science](#) **14**, 315–325(2020) | [Cite this article](#)

44 Accesses | **1** Citations | [Metrics](#)

Additional information

Data Access

The SNOTEL daily data are available from the National Water and Climate Center of the Natural Resources Conservation Service at <http://www.wcc.nrcs.usda.gov/snow/> (last access 25 January 2020). The spatial data used in Fig. 1 were obtained from the US Geological Survey National Viewer Data set <viewer.nationalmap.gov/advanced-viewer> (last access 21 February 2019). The PRISM data set was obtained from <http://www.prism.oregonstate.edu> (last access 25 January 2020).

Data & article reported findings

Journal of Vegetation Science

Advances in plant community ecology



RESEARCH ARTICLE |  Full Access

Effects of climate change on treeline trees in Sagarmatha (Mt. Everest, Central Himalaya)

Sudip Pandey, Paolo Cherubini, Matthias Saurer, Marco Carrer, Gaii Petit 

First published: 10 July 2020 | <https://doi.org/10.1111/jvs.12921> | Citations: 1

Open Research

DATA AVAILABILITY STATEMENT

Original data used for this study are available at the public repository of the University of Padua (Research Data Unipd) (<https://doi.org/10.25430/researchdata.cab.unipd.it.00000344> ; URI: <http://researchdata.cab.unipd.it/id/eprint/344>).

DOI: 10.1111/jvs.12921 (Publisher)

<http://hdl.handle.net/11577/3345504> (Padua Research Archive/IRIS)

Open Access to research data

Research data

Data management: main steps

Data Management: more tips

How to manage Open Data and why

Who asks for Open research data?

Where to archive research data



Managing research data: 5 steps



Collect
research
data



Name
research
data
rationally



Structure
research
data
hierarchically



Annotate
research
data using
metadata

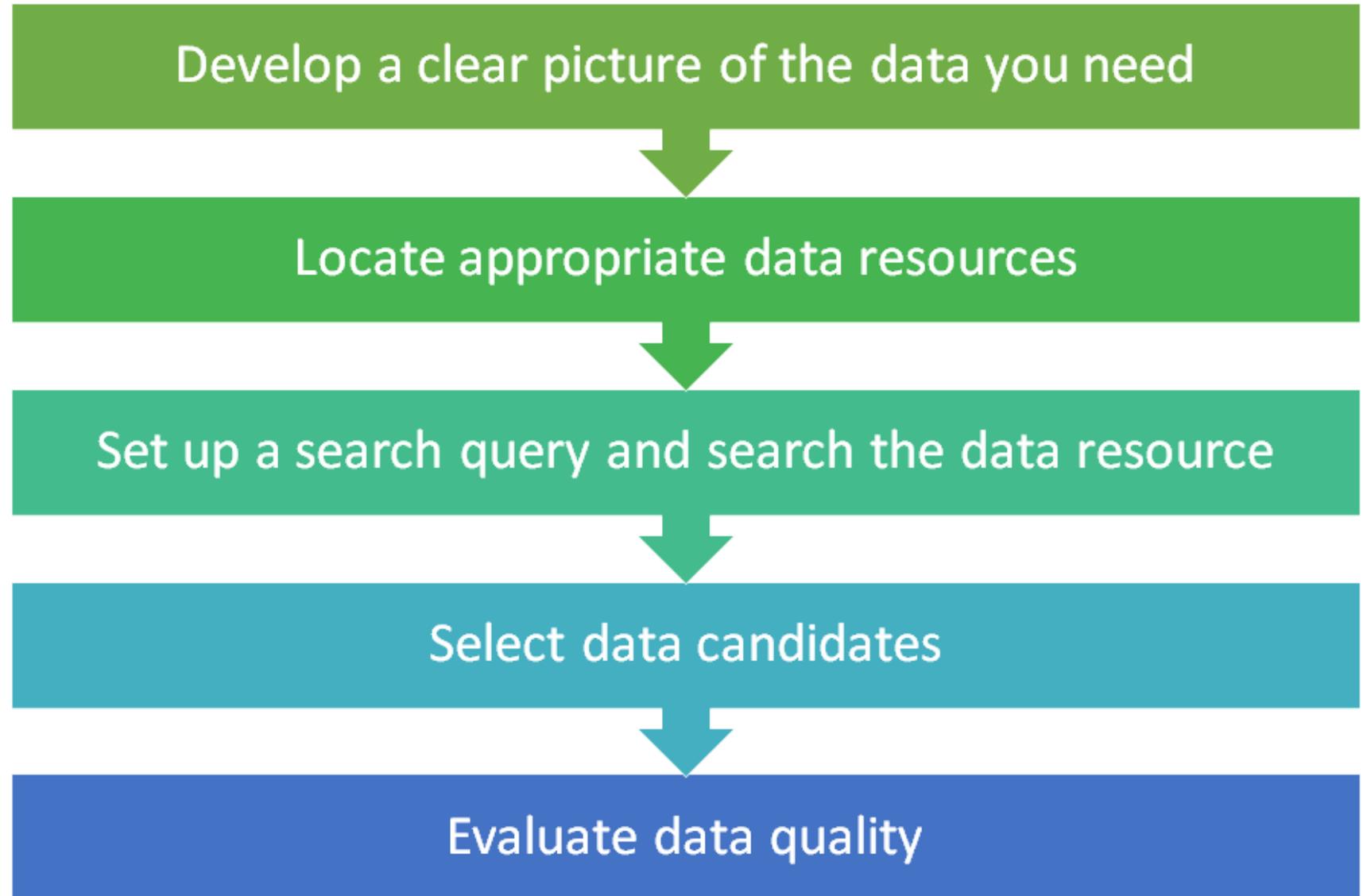


Pay attention
to file
formats



First step: collect research data

CC BY-SA Gaelen Pinnock



Second step: file name strategy

A file name is a principal identifier of a file

- File name should help to identify the content of the file
- Good file names provide useful clues to the status and version of a file, uniquely identify a file and help in classifying and sorting files



File naming strategy should be consistent in time and among different people

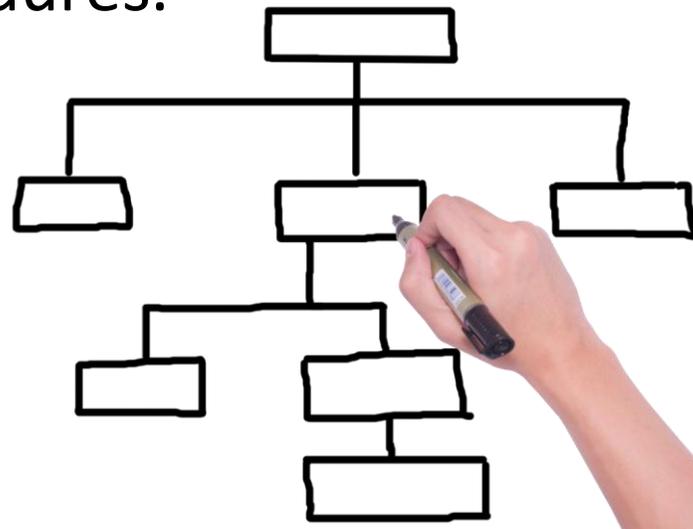
- In both quantitative and qualitative research file naming should be systematic and consistent across all files in the study
- A group of cooperating researchers should follow the same file naming strategy.



Third step: structure research data

Structuring your data files in folders is important for making it easier to locate and organize files and versions.

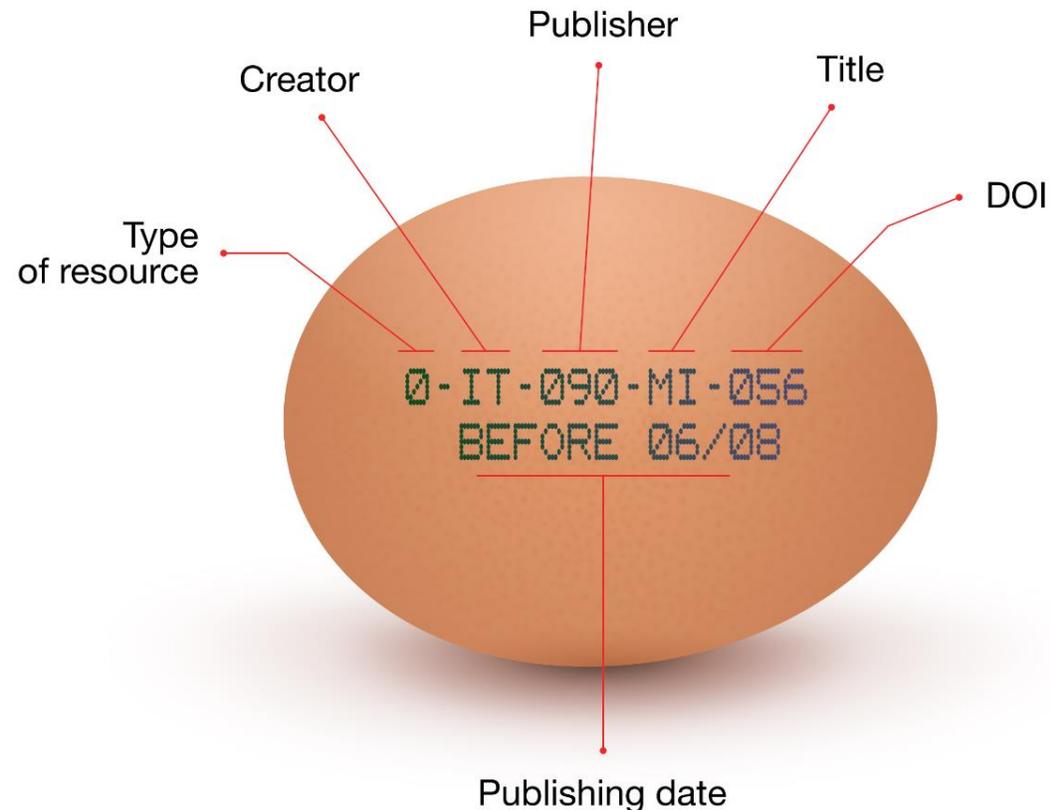
The decision on how to organize your data files depends on the plan and organization of the study. All material relevant to the data should be entered into the data folders, including detailed information on the data collection and data processing procedures.



Fourth step: annotate using metadata

Metadata means "data about data".

It is defined as the data providing information about one or more aspects of the data and it is used to summarize basic information about data, which can make easier to track and work with specific data.



Examples of
metadata
standards

Fifth step: file formats

When preparing to collect research data, you should chose **open, well-documented** and **non-proprietary formats** wherever possible.

The choice of format will vary depending on how you plan to analyze, store and share your data.

It is advisable to **store your data for use in future**, which means to convert them from a current data format to a long-term preservation format. Most software applications offer export or exchange formats that allow a text-formatted file to be created for importing into another program.



Open Access to research data

Research data

Data management: main steps

Data Management: more tips

How to manage Open Data and why

Who asks for Open research data?

Where to archive research data



Storage and preservation

Data **storage** in safe archives adhering to relevant standards.

Preservation actions should ensure that data remains authentic, reliable and usable while maintaining its integrity.

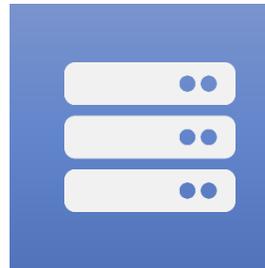


Open, non-proprietary, well documented formats



Regular backup

Multiple and different storage media



Checkup of integrity of files



Copy or migration of files



Track changes in metadata and files (versioning)

Reproducibility Issues

Retraction watch:
<https://retractionwatch.com/2016/09/23/author-asks-to-retract-nearly-20-year-old-paper-over-figure-questions-lack-of-data/>

Author asks to retract nearly 20-year old paper over figure questions, lack of data

The last author of a 1999 paper has asked the journal to retract it less than one month after a user raised questions about images on PubPeer.



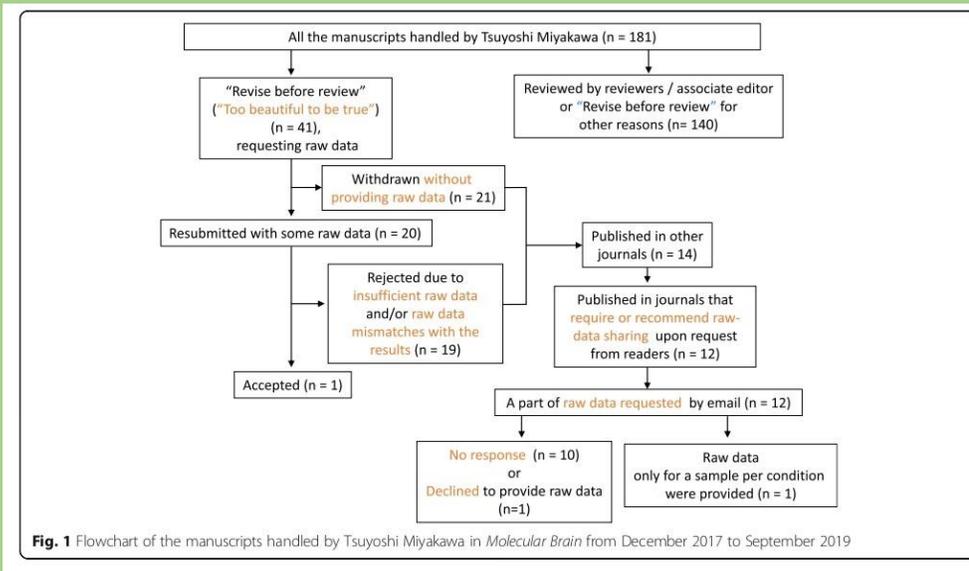
Yesterday, last author Jim Woodgett posted a note on the site saying the author who generated the figures in question could not find the original data, and since he agreed the images appeared “suspicious,” he had contacted the journal to retract the paper.

Here's the note from Woodgett, based at Lunenfeld-Tanenbaum Research Institute at Mount Sinai Hospital in Toronto:

...the person who generated the original data cannot source it and, as a consequence, a request to retract this paper based on the discrepancies in figure 5B and C has been submitted and approved.

The PubPeer exchange is over a pair of figures in the 1999 paper, “Regulation of the protein kinase activity of Shaggy(Zeste-white3) by components of the wingless pathway in Drosophila cells and embryos,” which has been cited 77 times, according to Thomson Reuters Web of Science.

Reproducibility Issues: No raw data no science



REPLICATIONINDEX.COM
Psychologists are not immune to the Dunning-Kruger Effect
 5 Likes | 1 Comment | 1 Condivision

Science
Only 36% of studies replicated!!

IS THERE A REPRODUCTION CRISIS?

7% Don't know
 3% No, there is no crisis

1,576 researchers

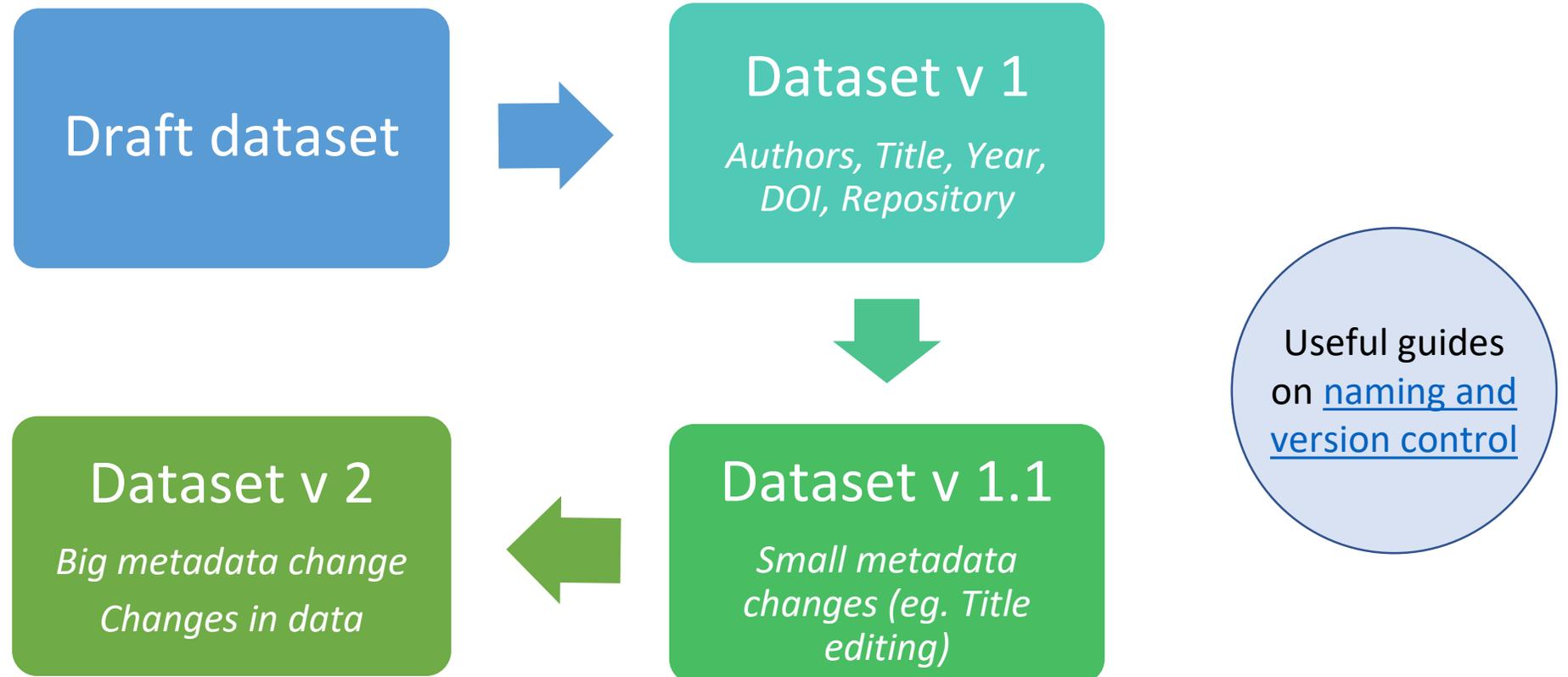
NATURE | NEWS
Over half of psychology studies fail reproducibility test
 Largest replication study to date casts doubt on many published positive results.
 Monya Baker
 27 August 2015
 Rights & Permissions

Miyakawa, T. No raw data, no science: another possible source of the reproducibility crisis. *Mol Brain* 13, 24 (2020). <https://doi.org/10.1186/s13041-020-0552-2>

Organize data: dataset versions

Versioning is important for long-term research data management where metadata and/or files are updated over time.

It is used to **track any metadata or file changes** (e.g., by uploading a new file, changing files structure, adding or editing file metadata...) once a dataset has been published.



Open methodology

= the use of open methodologies throughout the entire research cycle, making it open and available to everyone online at the very moment the research is conducted.

Open Notebooks

- <https://openlabnotebooks.org>
- <https://theopennotebook.com/>
- [OpenLab/Notebook](#) % Foster
- [Code Ocean](#)
- [Protocols.io](#)



Foto di [Ann H](#) da [Pexels](#)

Pre-registration

= the practice of pre-recording experiments



Retrospectively registered trials: the Editors' dilemma

Prospective clinical trial registration aims to address publication and reporting bias. Unfortunately, not all clinical trials are registered before they start. Here we discuss the dilemma faced by editors when receiving submissions reporting a clinical trial that was not registered prospectively, and a new policy for increasing transparency when a trial was registered

<http://blogs.biomedcentral.com/bmcblog/2016/04/15/retrospectively-registered-trials-editors-dilemma/>

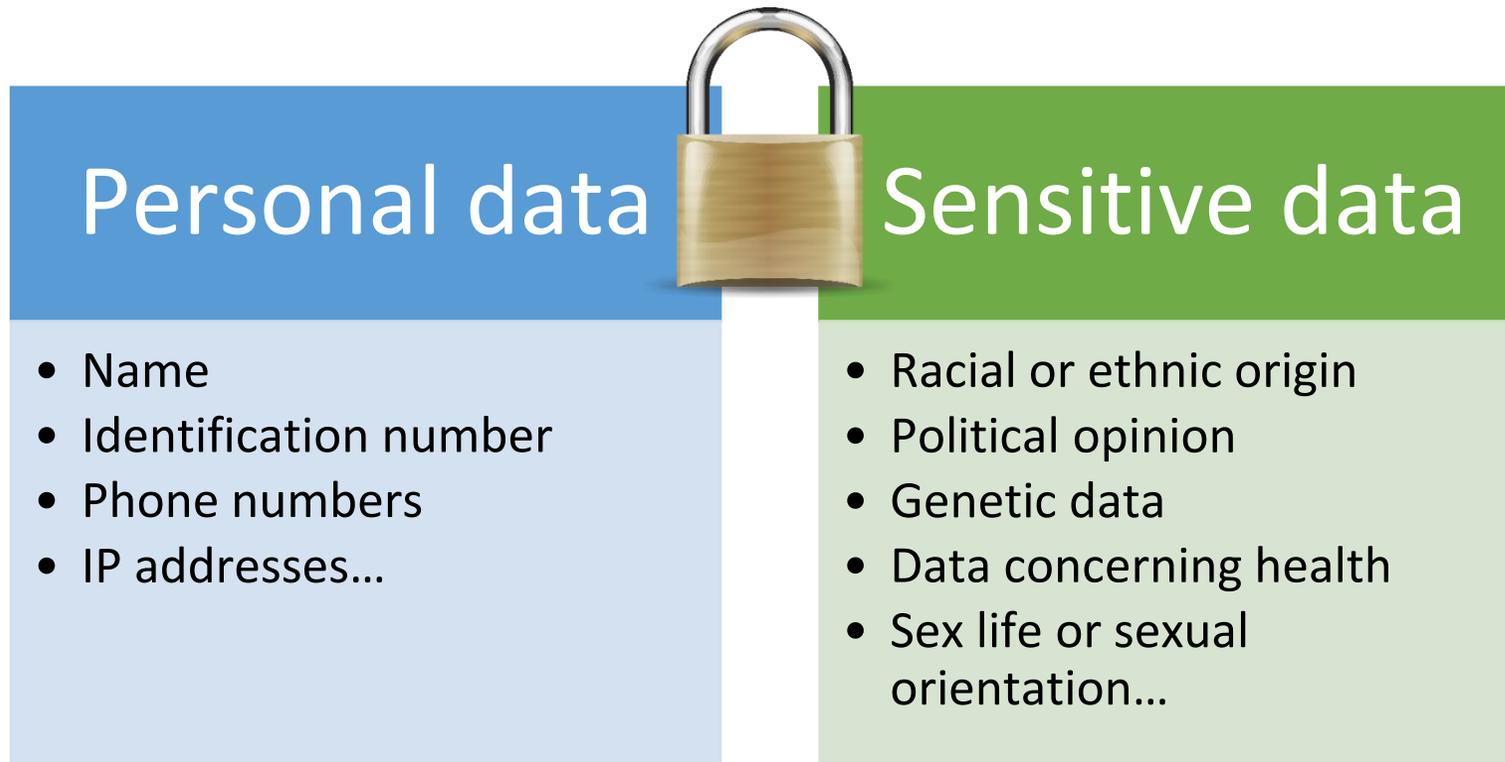
[OSF](#) – Open Science Framework

Privacy, sensitive and personal data



Privacy, sensitive and personal data

Research data may contain information about living, identifiable individuals, or other information that is sensitive, for example about criminal justice or national security. You are responsible for ensuring that your handling of all these data is secure and compliant with laws and regulations.



General Data Protection Regulation

Since 25 May 2018, the [General Data Protection Regulation](#) (GDPR, European Union, 2016) applies to any EU researcher who collects **personal data of living persons**.

So, when processing personal data, researchers should adhere to the following **six principles**:

I. Process lawfully, fair and transparent

II. Keep to the original purpose

III. Minimise data size

IV. Personal data should be accurate and, where necessary kept up to date

V. Remove data which are not used

VI. Ensure data integrity and confidentiality

Privacy, sensitive and personal data

Before you
collect data

- Make a **risk assessment**
- **Choose which data to collect**, ensuring compliance with the minimization principle
- Prepare **informed consent**, with information on: research, data sharing and conservation, subjects involved, rights of the interested party



Privacy, sensitive and personal data

After data
collection

- **Protect IDs** (eg. with pseudonymisation, or retaining information that allows identification in a separate archive)
- **Anonymize** whenever possible
- **Aggregate** data
- **Regulate** access where necessary



GDPR: research exemption

The GDPR contains an exemption which entails that some of the 6 principles are slightly different when you collect and process personal data for research purposes. This is called the “**research exemption**”.

Article 89

*Processing for archiving purposes in the **public interest, scientific or historical research purposes or statistical purposes**, shall be subjected to appropriate safeguards, in accordance with this Regulation, for the rights and freedoms of the data subject. Those safeguards shall ensure that technical and organisational measures are in place in particular in order to ensure respect for the principle of data minimisation. Those measures may include pseudonymisation provided that those purposes can be fulfilled in that manner. ...*

Principle II. and V. become less strict: further processing of research data shall not be considered to be incompatible with the initial purposes (even when they weren't mentioned earlier); also personal data may be stored for longer periods for the same reasons.



Open Access to research data

Research data

Data management: main steps

Data Management: more tips

How to manage Open Data and why

Who asks for Open research data?

Where to archive research data



The five stars of open data

The number of stars increases if data are:

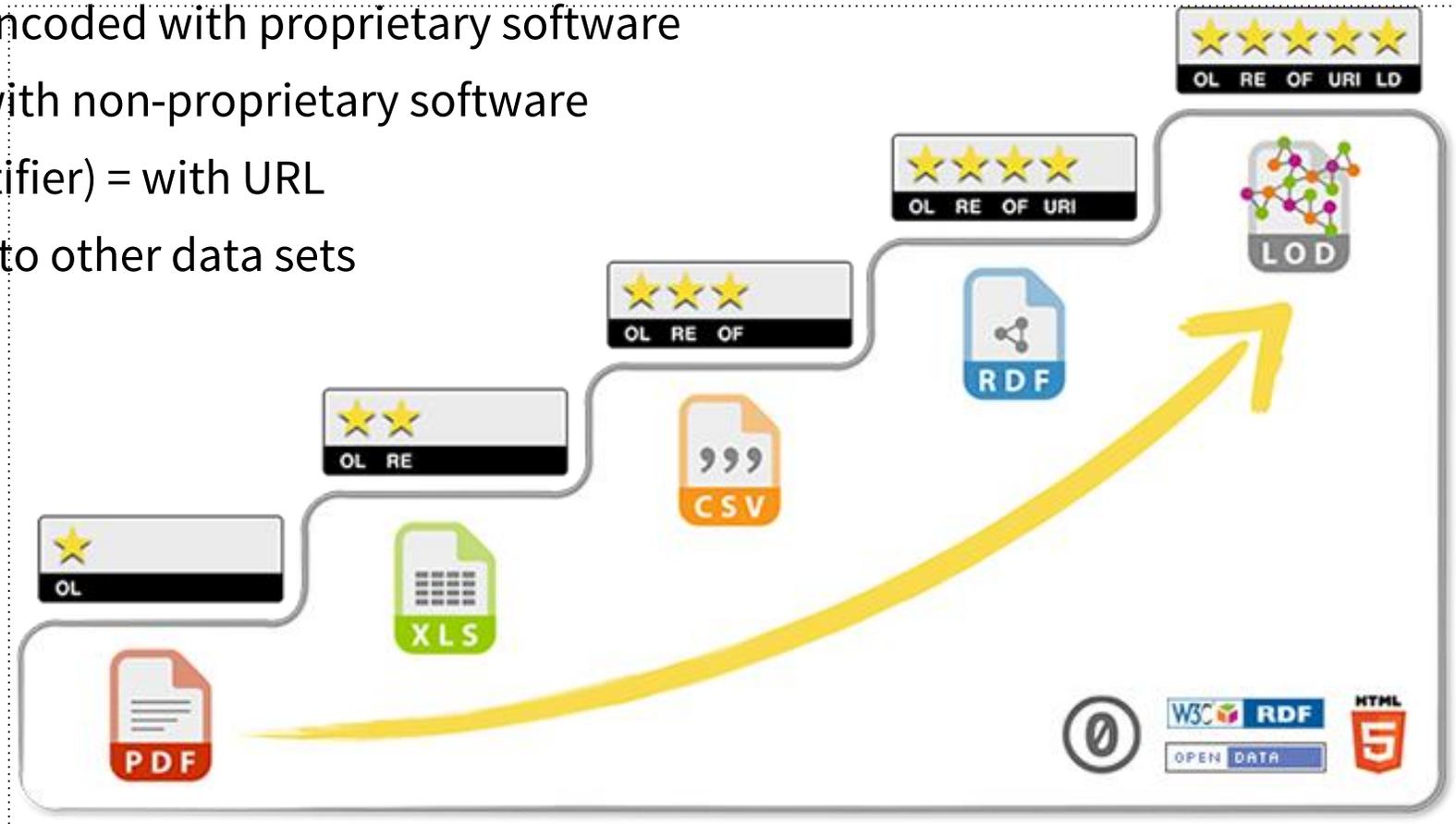
★ OL (OnLine) = distributed with an open license

★★ RE (Readable) = structured data encoded with proprietary software

★★★ OF (Open format) = encoded with non-proprietary software

★★★★ URI (Uniform Resource Identifier) = with URL

★★★★★ LD (Linked Data) = linked to other data sets



DMP = Data Management Plan

RESEARCH DATA - OPEN BY DEFAULT

Projects must have



Provides information on:



the data the research will generate



how to ensure its curation, preservation and sustainability



what parts of that data will be open (and how)

<https://www.openaire.eu/what-is-a-data-management-plan>

Unless otherwise indicated, all materials created by OpenAIRE are licenced under [CC ATTRIBUTION 4.0 INTERNATIONAL LICENSE](#).

DMP = Data Management Plan

1. Data Collection and Documentation



- What kind of data are generated
- How will data be generated
- What metadata are needed

2. Ethics, legal and security Issues



- How will ethical issues be handled
- How are the data accessed
- Are there copyright issues
- Are there sensitive data
- What about intellectual property rights

3. Data Storage and Preservation



- How are the data stored?
- Are there back up systems
- How are data safely preserved

4. Data Sharing and reuse



- How and where will the data be shared?
- How are sensitive data protected
- How can data be accessed

DMP: Guidelines & tools



DCC = Digital Curation Centre

- <http://www.dcc.ac.uk/resources/data-management-plans>
- <http://www.dcc.ac.uk/resources/tools-and-applications>

DMPTool

- <https://blog.dmptool.org/2018/02/27/new-dmptool-launched-today/>



Italian Open Science Support Group

- Italian checklist
- http://bibliotecadigitale.cab.unipd.it/bd/per_chi_publica/documenti-e-materiali/Grigliapianodigestionedatiricerca.pdf

OpenAIRE

- <https://www.openaire.eu/what-is-a-data-management-plan-and-how-do-i-create-one?highlight=WyJob3ciLCJ0byIsImNyZWFOZSIsImRtcCIsImRtcCdzIiwiaG93IH RvIiwiaG93IHRvIGNyZWFOZSIsInRvIGNyZWFOZSJd>



Canadian Association of Research Libraries (CARL)

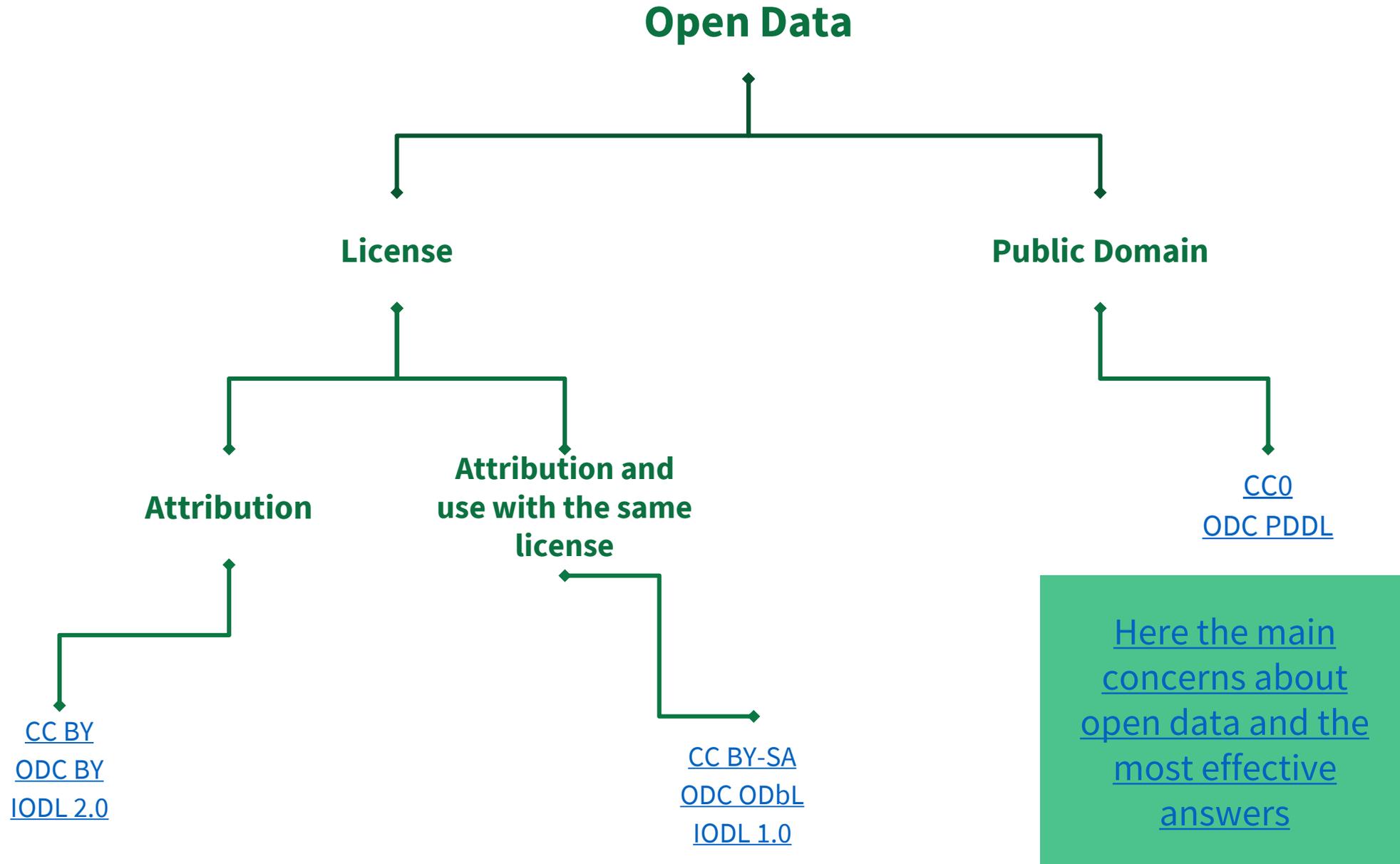
- [Portage](#)

Tools for researchers

- UniPD Ufficio Ricerca Internazionale
- Strumenti per la progettazione e il proposal writing
- <https://elearning.unipd.it/uffici SERVIZI applicazioni/course/view.php?id=112> (with SSO)



Licenses for Open data



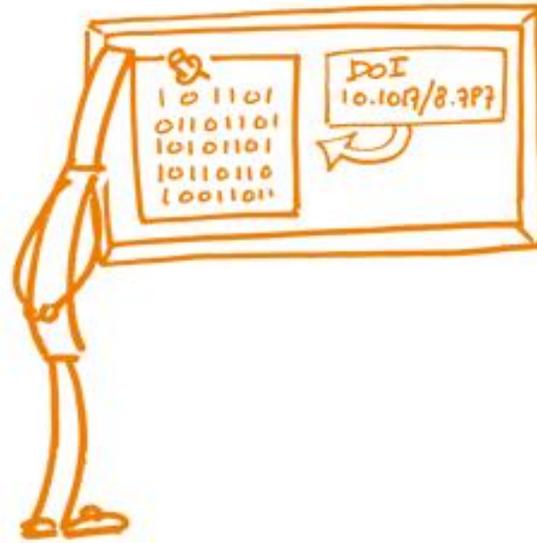
FAIR principles

FAIR DATA PRINCIPLES

AH!



FINDABLE



ACCESSIBLE

HOW DO YOU
OPEN A .XZQ FILE?



INTEROPERABLE



REUSABLE

FAIR principles



Why is it important to manage research data [properly] and make them OPEN?



To allow the continuity of research through the use of secondary data

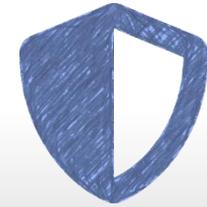


To increase the efficiency of research

To ensure compliance with the requirements set by funders



To support the contents of a paper and improve the peer-review



To guarantee the integrity of research and the validation of the results



To ensure greater dissemination and greater impact

Open Access to research data

Research data

Data management: main steps

Data Management: more tips

How to manage Open Data and why

Who asks for Open research data?

Where to archive research data



UNESCO Recommendation on Open Science

UNESCO Recommendation on Open Science

At the 40th session of UNESCO's General Conference, 193 Member States tasked the Organization with the development of an international standard-setting instrument on Open Science in the form of a UNESCO Recommendation on Open Science to be adopted by Member States in 2021.

The Recommendation is expected to define shared values and principles for Open Science, and identify concrete measures on Open Access and Open Data, with proposals to bring citizens closer to science and commitments to facilitate the production and dissemination of scientific knowledge around the world. The Recommendation will be developed through a regionally balanced, multistakeholder, inclusive and transparent consultation process.

UNESCO Recommendation on Open Science will complement the 2017 Recommendation on Science and Scientific Research. It will also build upon the UNESCO Strategy on Open Access to Scientific Information and Research and the new UNESCO Recommendation on Open Educational Resources.



Funding programs requiring OA



- projects funded with public funds (Horizon 2020 and Horizon Europe, Marie Curie, ERC)
- projects funded by private foundations (e.g. Bill & Melinda Gates Foundation or Wikimedia Foundation)
- projects funded by institutions or research networks that adhere to [cOAlition S](#)



Initiative for the publication in
OA of all research contents
financed with public money.

https://www.coalition-s.org/plan_s_principles/

“Although the Plan S principles
refer to peer-reviewed
scholarly publications,
cOAlition S also strongly
encourages that research data
and other research outputs are
made as open as possible and
as closed as necessary. The
early sharing of research
results through preprints is
also strongly encouraged.”

<https://www.coalition-s.org/guidance-on-the-implementation-of-plan-s/>



Plan S

Making full and immediate Open Access a reality

Funding programs requiring OA: European Union



OPEN  ACCESS

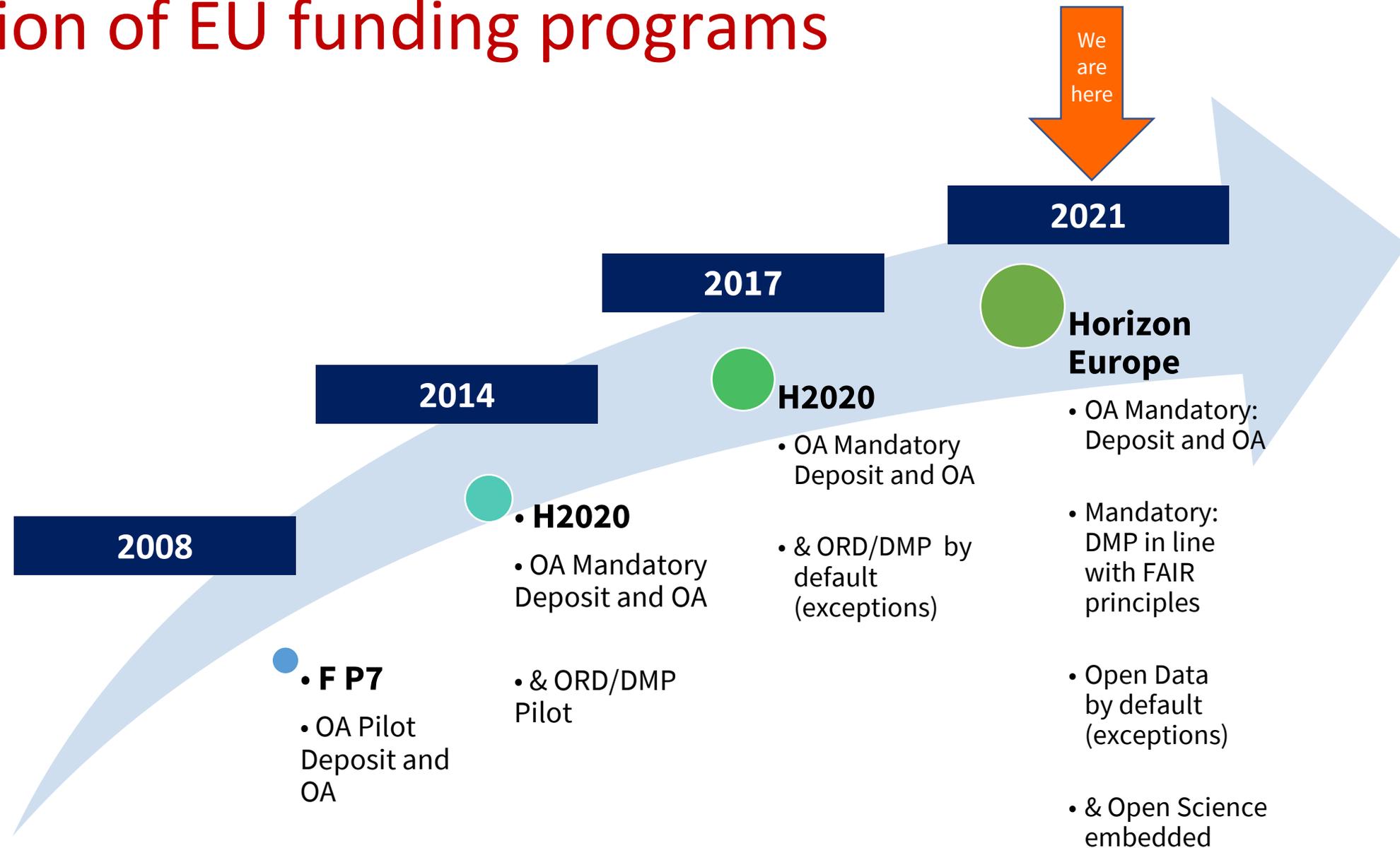
- [COMMISSION RECOMMENDATION \(EU\) 2018/790 of 25 April 2018 on access to and preservation of scientific information](#)
- [DIRECTIVE \(EU\) 2019/1024 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 20 June 2019 on open data and the re-use of public sector information](#)

UE Directive 2019/1024

- Art. 28:

“Open access is understood as the practice of providing online access to research outputs free of charge for the end user and without restrictions on use and re-use beyond the possibility to require acknowledgement of authorship. Open access policies aim in particular to provide researchers and the public at large with access to research data as early as possible in the dissemination process and to facilitate its use and re-use. Open access helps enhance quality, reduce the need for unnecessary duplication of research, speed up scientific progress, combat scientific fraud, and it can overall favour economic growth and innovation. Beside open access, commendable efforts are being made to ensure that data management planning becomes a standard scientific practice and to support the dissemination of research data that are findable, accessible, interoperable and re-usable (the FAIR principle).”

Evolution of EU funding programs



Reprinted from OpenAIRE webinar: Horizon 2020 Open Science Policies and beyond, October 22nd, 2019 by Emilie Hermans (Ghent University)

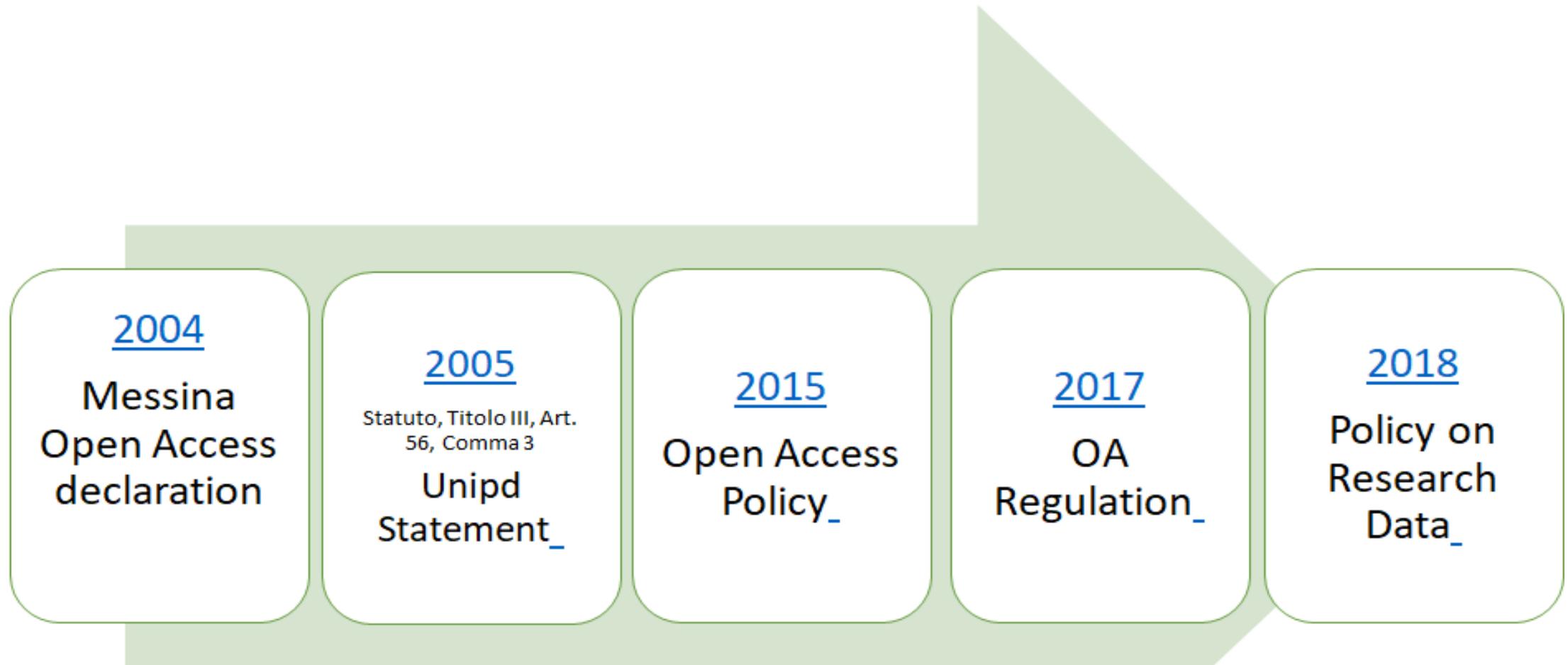
https://www.slideshare.net/OpenAIRE_eu/horizon-2020-open-science-policies-and-beyond-with-emilie-hermans-openaire

Horizon Europe 2021-2027

- Horizon Europe 2021-2027 is the **European Framework Program for Research and Innovation for the period 2021-2027**;
- with a budget of around **100 billion euros**, it is the most ambitious research and innovation program ever;
- based on the success of the Horizon 2020 experience, the new Framework Program will continue to support and promote scientific excellence in Europe fostering, among other forms and models of funding, the **open science policy for a better dissemination of results**.



The long journey towards open science @ University of Padua



Open Data @ UniPD

The 1° December 2018 the [Policy on the management of research data](#) of the University of Padova entered into force.

WHO and WHAT

- “This policy applies to all University research projects limited to the parts for which the University is responsible”
- “Staff people are required to observe it”.

WHERE

- “Research data must be archived into the digital repository of the University of Padova called Research Data Unipd, or into a digital repository that complies with international standards”.

HOW

- Data must be stored correctly, completely, respecting their integrity.
- They must also be accessible, identifiable, traceable, interoperable and, where possible, available for subsequent use (FAIR principles)”.

Research Data Unipd

Research Data Unipd

is a platform for long-term management and archiving of research data and for the access and re-use of data necessary to validate the results of scientific publication

It is already equipped with:

- *Authentication via the University's SSO;
- ***DOI attribution**;
- ***Connection between dataset and articles from the publisher's website or deposited in Padua Research Archive**;
- ***ERC subjects**.

It allows the self-archiving of datasets of any format with **FAIR mode** (Findable, Accessible, Interoperable, Reusable), as recommended by the European Commission.

<http://researchdata.cab.unipd.it>

Welcome to Research Data Unipd

Research Data Unipd is a research data archive. The service aims to facilitate data discovery, data sharing, and reuse as required by funding institutions (eg. European Commission).

Anyone has access to data. The deposit of datasets is reserved to institutional users: they can login with their SSO credentials.

For more information on Research Data Management and Repositories, please refer to the [Research Data Management Service web pages](#) or contact the [Library Help-line](#).

 Atom  RSS 1.0  RSS 2.0

[Latest Additions](#)

View items added to the repository in the past 90 days.

[Search Repository](#)

Search the repository using a full range of fields. Use the search field at the top of the page for a quick search.

[Browse Repository](#)

Browse the items in the repository by [Year](#), [Subject](#), [Department](#) and [Authors](#).

[About this Repository](#)

More information about this site.

Research Data Unipd supports [OAI 2.0](#) with a base URL of <http://researchdata.cab.unipd.it/cgi/oai2>

About the Repository

About Research Data Unipd

Research Data Unipd supports research produced by members of the University of Padova. The service aims to facilitate data discovery, data sharing, and reuse as required by funding institutions (eg. European Commission).

Quality

Datasets published in the Archive have a set of metadata that ensure that data are described and discoverable. Before publication, dataset records are checked by Editors for presence of appropriate metadata.

Metadata Policy

All published metadata are released under a CC0 licence.

Re-using data

We encourage Researchers to use licences on their datasets to promote reuse of the research data. The licence to be preferred is Creative Commons Attribution 4.0, but several others are used. Any re-use must acknowledge the Creators in an appropriate manner, ideally through a citation similar to that provided with the record.

Recommended formats and data files

[Formats and data files.](#)

Submission policy

[Submission policy concerning depositors, quality & copyright.](#)

Data deposit agreement

[Agreement to terms and conditions.](#)

Data deposit agreement

When you deposit data in the Research Data Unipd Archive, you will need to agree to the conditions below. This is done by clicking the "Deposit" button in the archive, before depositing the item.

This agreement confirms that you, the depositor, have the right to submit the dataset to the repository.

This agreement ensures that the archive administrators have the right to carry out activities necessary to facilitate the long-term preservation and sharing of datasets.

By submitting your dataset for deposit, you grant a non-exclusive licence to the University of Padova to archive, publish and disseminate any material within the dataset. The licence is non-exclusive, and therefore does not prevent you exercising any rights you might have to publish and distribute any of the dataset, in its present or future versions, elsewhere.

A dataset

A dataset for hand-eye calibration evaluation

Koide, Kenji and Menegatti, Emanuele (2019) *A dataset for hand-eye calibration evaluation*. [Data Collection]

Related publications: [https://ieeexplore.ieee.org/abstract/doc... \(Publisher\)](https://ieeexplore.ieee.org/abstract/doc...)

Collection description

Description: This dataset aims to assess the accuracy of hand-eye calibration methods (i.e., estimation of the transformation between a robot end effector frame and a camera mounted on it). It contains two sets of images and corresponding robot hand poses. The first one (calib_test) contains images of a calibration pattern to estimate the hand-eye transformation. The second one (spirit_reconst) contains images of a pattern to be 3D reconstructed and manually annotated 2D feature points on the images. By performing multi-view 3D reconstruction on the second set and checking the flatness of the reconstructed points, the calibration accuracy can be assessed. The dimension of the calibration pattern in this dataset is 32 mm. Paper: Kenji Koide and Emanuele Menegatti, General Hand-Eye Calibration based on Reprojection Error Minimization, IEEE Robotics and Automation Letters/ICRA2019

Keywords: Hand-eye calibration

Subjects: [Physical Sciences and Engineering > Computer Science and Informatics: Informatics and information systems, computer science, scientific computing, intelligent systems > Computer graphics, computer vision, multi media, computer games](#)

Department: [Departments > Dipartimento di Ingegneria dell'Informazione \(DEI\)](#)

Depositing User: Kenji Koide

Date Deposited: 29 Apr 2019 11:49

Last Modified: 25 Jun 2019 12:24

DOI: 10.25430/researchdata.cab.unipd.it.00000122

URI: <http://researchdata.cab.unipd.it/id/eprint/122>

[+ Additional details](#)

Available Files

Data

[+ st_handeye_eval.tar.gz](#)

Cite As

Select Formatting Style:

Begin typing (e.g. Chicago or IEEE.) or use the drop down menu.

Select Language and Country:

Begin typing (e.g. en-GB for English, Great Britain) or use the drop down menu.

Export As

Additional details and info on files

Additional details

Creators/Authors:	Creators Zane, Antonella	Email antonella.zane@unipd.it	ORCID  orcid.org/0000-0001-7218-6068
Type of data:	Text		
Contributors:	Contribution Editor	Name Chavarria Arnau, Alexandra	Email UNSPECIFIED
	Editor	Brogio, Gianpietro	UNSPECIFIED
Collection period:	From 1999	To 2000	
Geographic coverage:	Italia - Veneto		
Data collection method:	Utilizzata sonda elettronica (EMPA), microscopio a Trasmissione elettronica (TEM), diffrazione RX su polveri, analisi petrografica al microscopio polarizzatore.		
Statement on legal, ethical and access issues:	La ricerca non ha prodotto dati sensibili né altri tipi di dati con rilevanza etica.		
Data processing and preparation activities:	Campioni di roccia provenienti da cave di pietra ollare delle Alpi centro-occidentali; frammenti di reperti archeologici provenienti da recipienti in pietra ollare rinvenuti in Veneto.		

Available Files

Data

 [Monselice_ollare ...
ci_Zane2017.PNG](#)

 [Monselice_ollare ...
io_Zane2017.PNG](#)

Visible to: Anyone

Content type: Data

Description: microscopia

Metadata Revision: 3

Mime-Type: image/png

License: Creative Commons: Attribution 4.0

File size: 381kB

Read me

 [Monselice_readme_file.txt](#)

Visible to: Anyone

Content type: ["content_type_name_readme" not defined]

Metadata Revision: 3

Mime-Type: text/plain

License: Creative Commons: Attribution 4.0

File size: 922B

International registers of Open Access repositories

Research Data Unipd is indexed to re3data.org e OpenDOAR

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Research Data Unipd

Subject(s)

Humanities and Social Sciences Life Sciences Natural Sciences Engineering Sciences

Content type(s)

Databases Images Structured graphics Audiovisual data Scientific and statistical data formats Raw data
Plain text Structured text Archived data Software applications Source code Standard office documents

Country

Research Data Unipd is a
data discovery, data sharing
set of metadata that ensu

Research Data UNIPD

Repository Information

Repository Name	Research Data UNIPD [English]
Repository Type	Institutional
Description	This site is a research data archive from the University of Padua. The interface is available in English.
Repository URL	http://researchdata.cab.unipd.it/
OAI-PMH URL	http://researchdata.cab.unipd.it/cgi/oai2
Software Name	EPrints
Languages	English Italian
Content Types	Datasets
Subjects	Multidisciplinary

OpenDOAR

Interdisciplinary repositories: Zenodo



<https://zenodo.org/>

For self-archiving of publications and data, open to all researchers in the world.

Managed by CERN for OpenAIRE (EU)

Search results are stored securely in the same cloud infrastructure as CERN's LHC search data

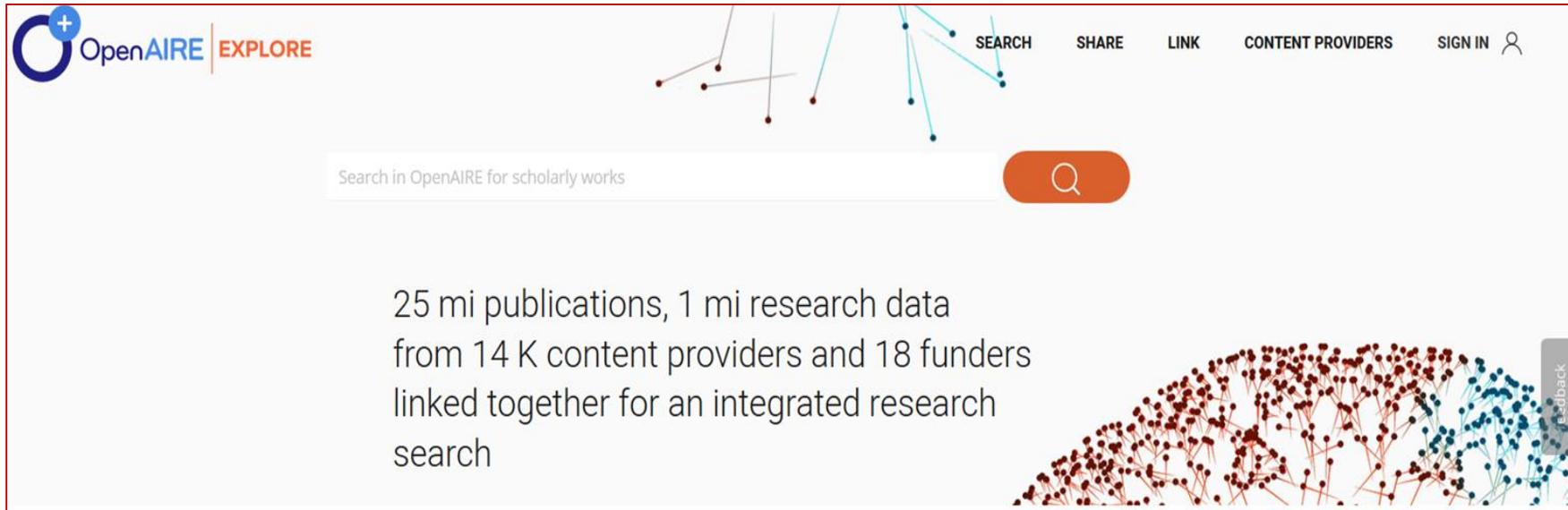
Assignment of a digital object identifier (DOI)

Possibility of identifying any subsidies, as they are integrated into the reporting lines

Possibility of assigning flexible guarantees of use, since not everything is under Creative Commons

Make your content count with OpenAIRE

Research Data Unipd is also registered as **Content Provider of OpenAire** (Open Access Infrastructure for Research in Europe), an infrastructure financed by the European Commission with the aim of collecting and disseminating the results of research (publications and data) financed with public funds.



The screenshot shows the OpenAIRE Explore website interface. At the top left is the OpenAIRE logo with a plus sign and the word "EXPLORE". To the right are navigation links: "SEARCH", "SHARE", "LINK", "CONTENT PROVIDERS", and "SIGN IN" with a user icon. Below the navigation is a search bar with the placeholder text "Search in OpenAIRE for scholarly works" and a magnifying glass icon. The main content area features the text: "25 mi publications, 1 mi research data from 14 K content providers and 18 funders linked together for an integrated research search". The background of the interface is decorated with abstract network diagrams consisting of nodes and connecting lines, with a "Feedback" button visible on the right side.

Library System support services

In the section “[About publishing](#)” of the Library System web portal, researchers will find information on Open Access, on publishing, and on the management of data.

About publishing



Filed under: [digital repositories](#), [open access](#), [self archiving](#), [OAI](#), [license agreement](#), [publication](#), [open archives](#), [publication standards](#), [Impact Factor](#)



Research repositories

Get your articles viewed more often



Open Access

Increase the impact of your research



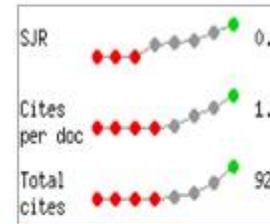
Research Data Unipd

A safe place for your research



Research data management

Manage your data in the best possible way



Measure your impact

Evaluate the scientific impact of your research



Publishing support

Open Access publishing? It's easier with us!

Library System support services

Authors can submit specific requests using the Library System [Help Service](#), choosing the following address:

- 11 Supporto Open Access (Supporto Ricerca)



Kyle James <https://www.flickr.com/photos/jameskm03/2711755476>



Before and after publishing articles and data,
improve your knowledge with:

Scholarly Communication and principles of Open Science

a Training Course on SBA Moodle,
composed of five modules.

It aims to introduce early-career researchers
to scientific communication and to the
principles of Open Science (Open Access,
Open Data, Open Licences).



OA Support Group of the UniPd Library System



This work is distributed under a Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)