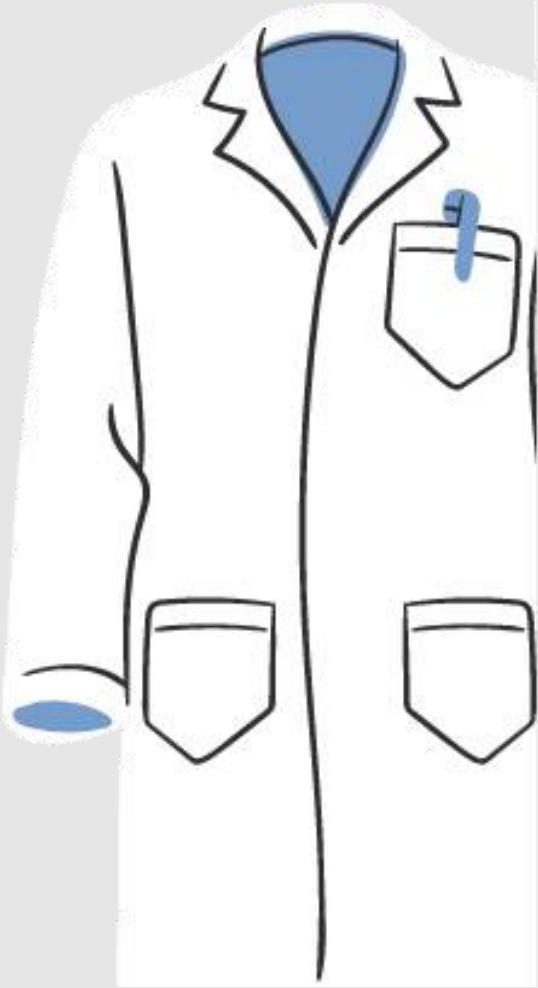
ON LIBRARY TOPICS

PhD Course in Clinical and experimental Oncology and Immunology

2020/2021



DIGITAL LIBRARY & V. PINALI MEDICAL LIBRARY

ON LIBRARY TOPICS

5. OPEN SCIENCE

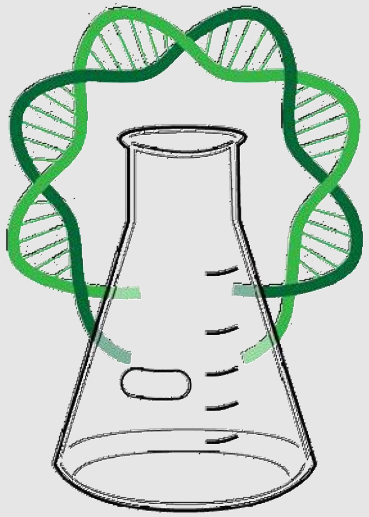
6. RESEARCH DATA

Michela Zorzi michela.zorzi@unipd.it

Open Science

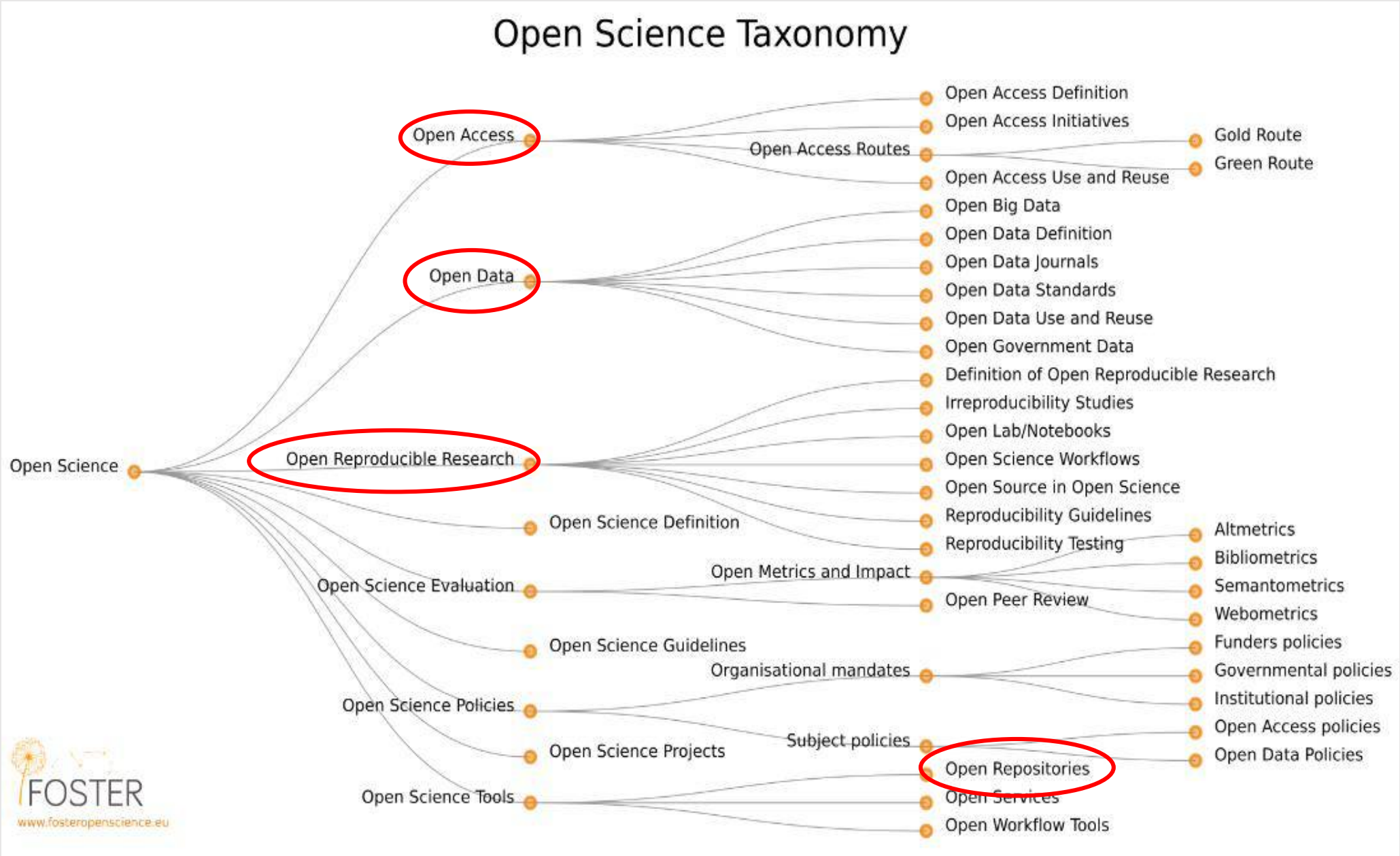
"Open science is the movement to make scientific research, data and dissemination accessible to all levels of an inquiring society"

FOSTER consortium



Open Science

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



FOSTER consortium – Open Science taxonomy – related article

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/

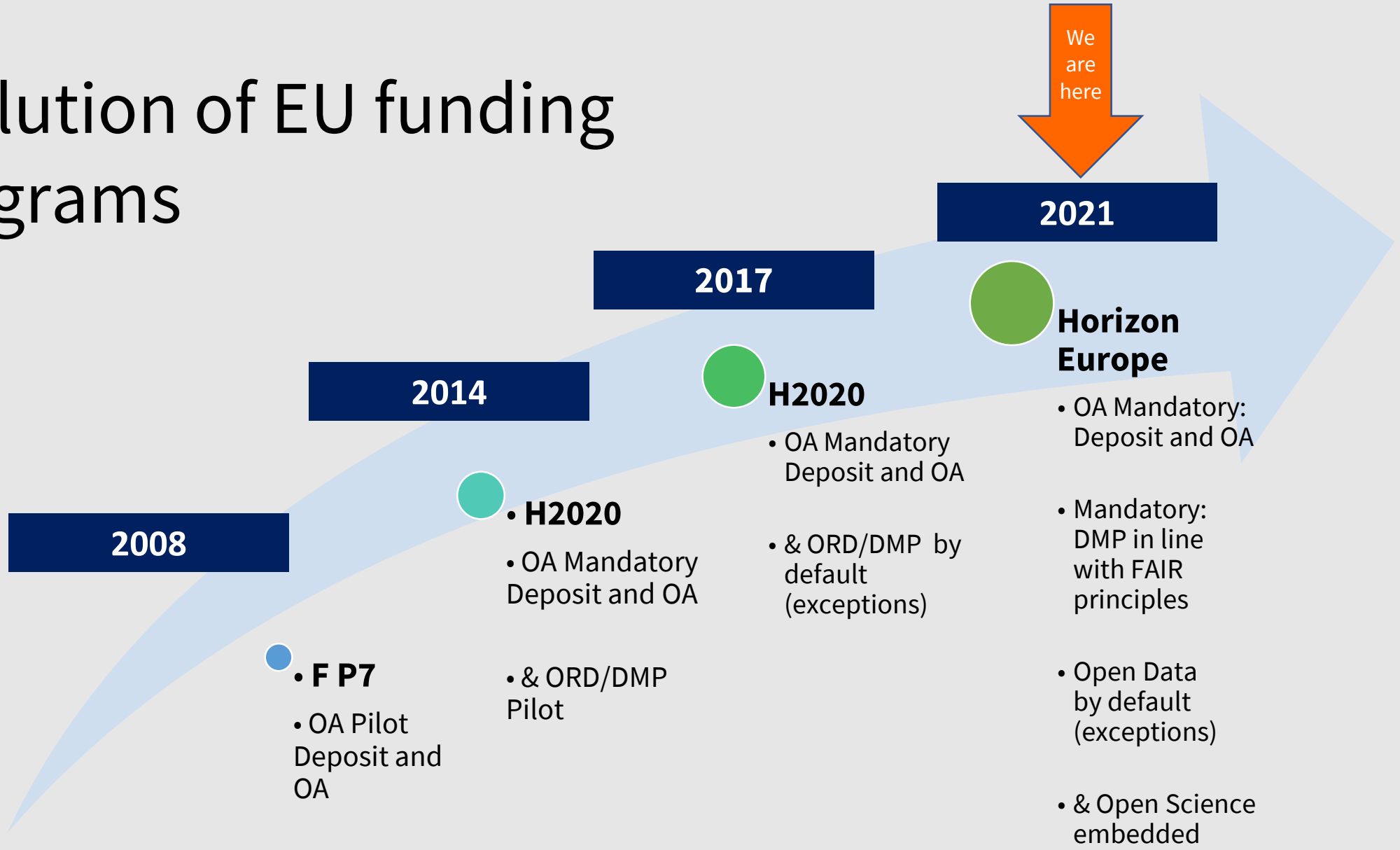
“Although the Plan S principles refer to peer-reviewed scholarly publications, cOAlition S also strongly encourages that research data and other research outputs are made as open as possible and as closed as necessary. The early sharing of research results through preprints is also strongly encouraged.”

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

European Union

- 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

Evolution of EU funding programs



Reprised 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

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

GIS and spatial data

Video

Digital copies of images

Spreadsheets

Digital texts or digital copies of text

Audio

Graphics

Protein or genetic sequences

Databases

BUT ALSO…

- **Computer Aided Design (CAD)**
- **Waveforms**
- **Computer codes**
- **Statistics (SPSS, SAS)**
- **File Matlab**
- **Artistics products**
- **Web files**
- **…**

Raw data, primary data

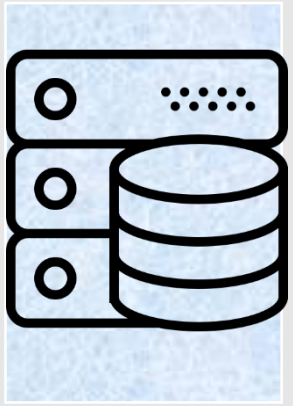
Raw data have been collected or generated in the course of research, but have not been analysed or manipulated yet.

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

https://www.pexels.com/it-it/foto/acqua-ambiente-concentrarsi-crescita-531428/

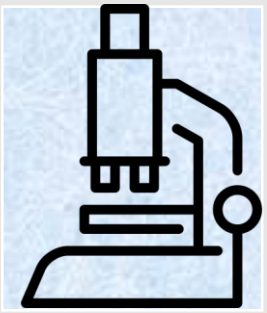
General categories of data

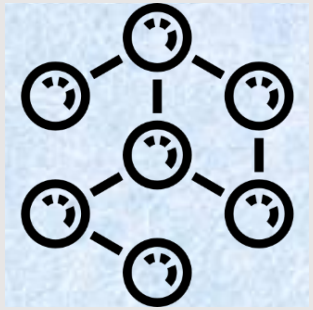
Derived or compiled: Derived data involves using existing data points to create new data (e.g. compiled databases, text or data mining); reproducible but expensive



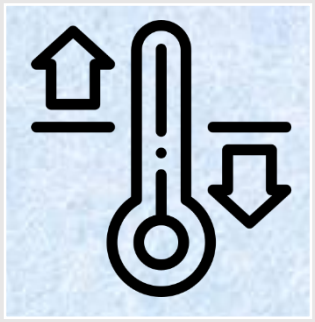
Observational: Observational data are captured through observation of a behavior or activity (e.g. sensor readings, survey instruments); usually irreplaceable and not replicable

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

https://libguides.macalester.edu/c.php?g=527786/&p=3608643

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.

It is the philosophy of Open Access applied to data

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

Examples:

- institutional/government open data (e.g. open by default according to the Italian Digital Administration Code)

- research data available to citizens

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/

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

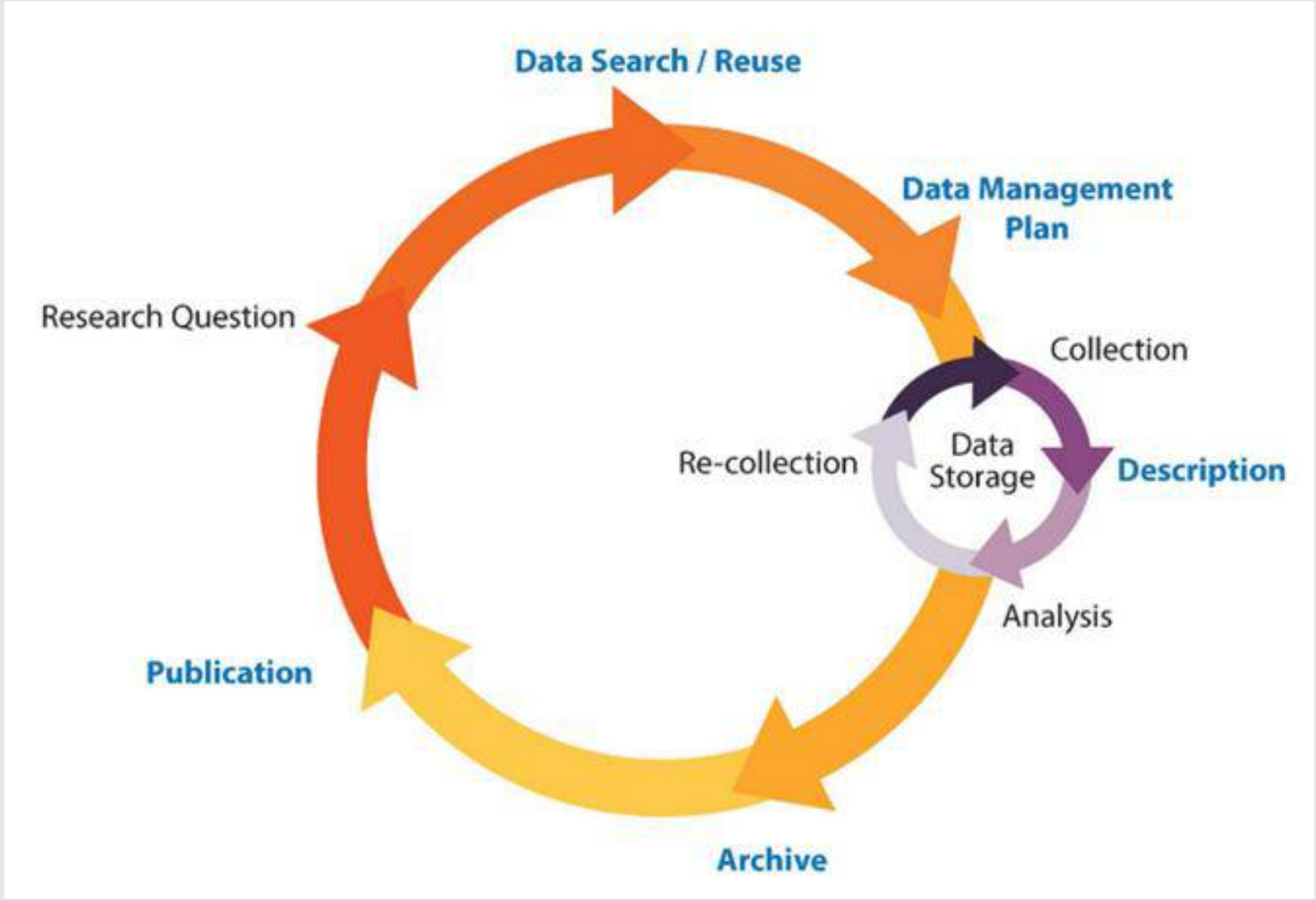
GISAID, an Initiative that promotes the international sharing of all influenza virus sequences, related clinical and epidemiological data associated with human viruses GISAID - Next hCoV-19 App

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

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

Research data lifecycle



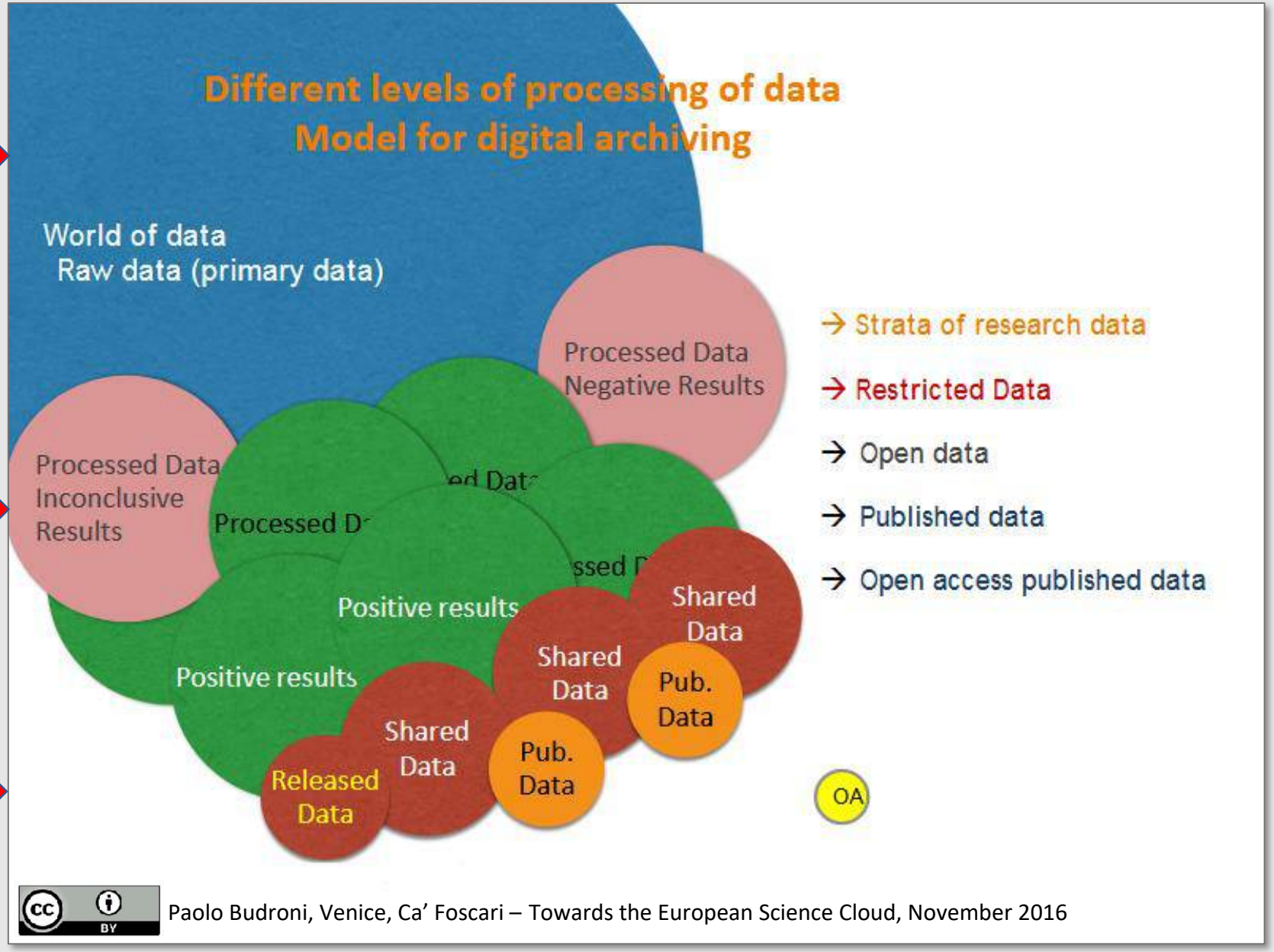
Adapted original source: The University of California, Santa Cruz, Data Management LibGuide, *Research Data Management Lifecycle*, diagram, viewed 15th September 2015 <http://guides.library.ucsc.edu/datamanagement>

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.



Paolo Budroni, Venice, Ca' Foscari – Towards the European Science Cloud, November 2016

Data & article reported findings

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

In other cases, the data are *integral* to the findings being reported, that is, necessary to support the major claims of the paper and essential to enable a knowledgeable peer to reproduce and verify the results.

In other cases, the data or a database provides *background* to a publication—that is, not integral to the findings or conclusions being presented, but without them the findings or conclusions could not have been derived.

Background information would not be essential for reproducing, verifying, or building on the claims in the paper; it might be considered as background, for instance, because obvious alternative methods or sources of data could be substituted.

Data & article reported findings

Springer Link

Original Research | Published: 19 June 2020

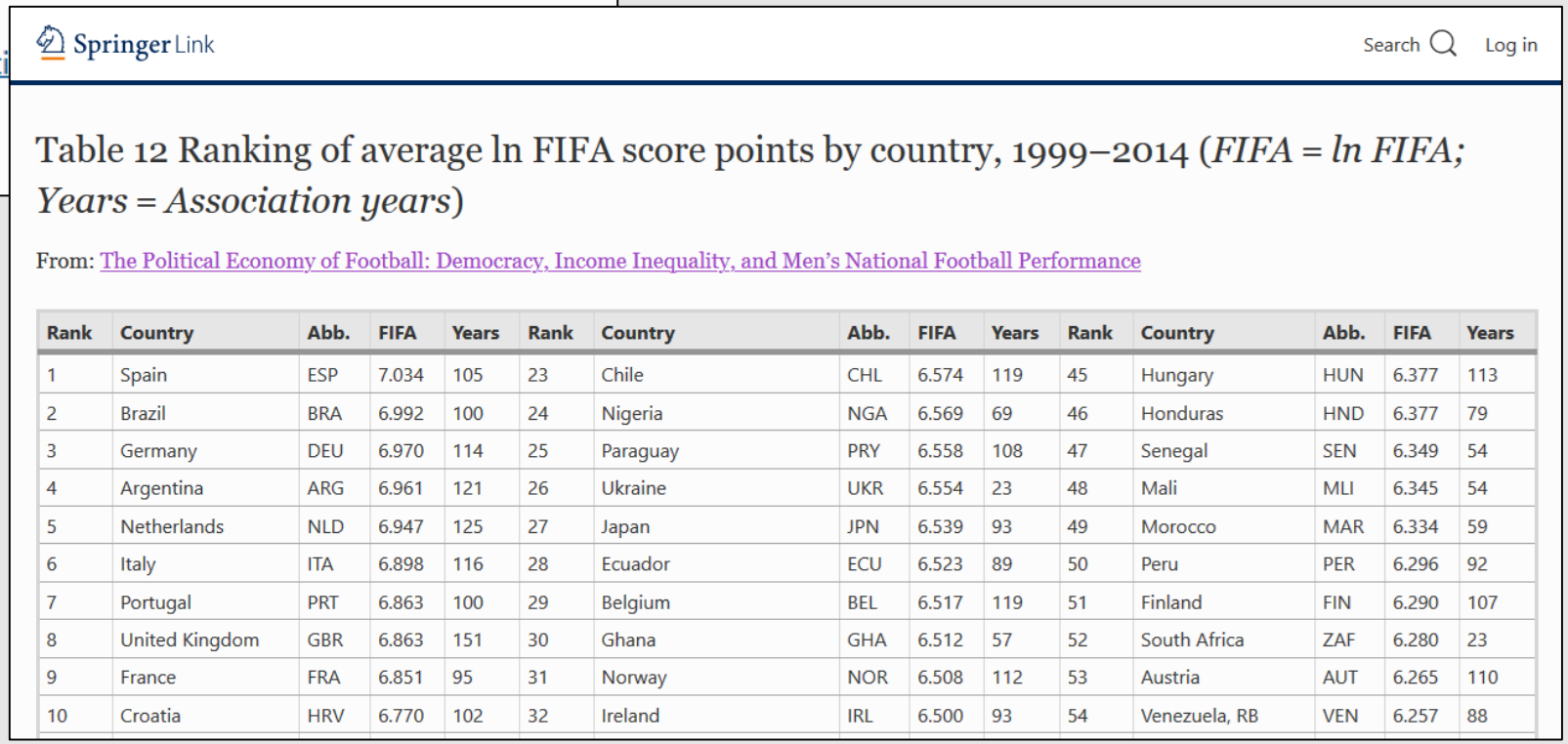
The Political Economy of Football: Democracy, Income Inequality, and Men's National Football Performance

Kin-Man Wan, Ka-U Ng & Thung-Hong Lin

Social Indicators Research **151**, 981–1013(2020) | Cite this arti

242 Accesses | **4** Altmetric | Metrics

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



Springer Link Search Log in

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

Springer Link

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

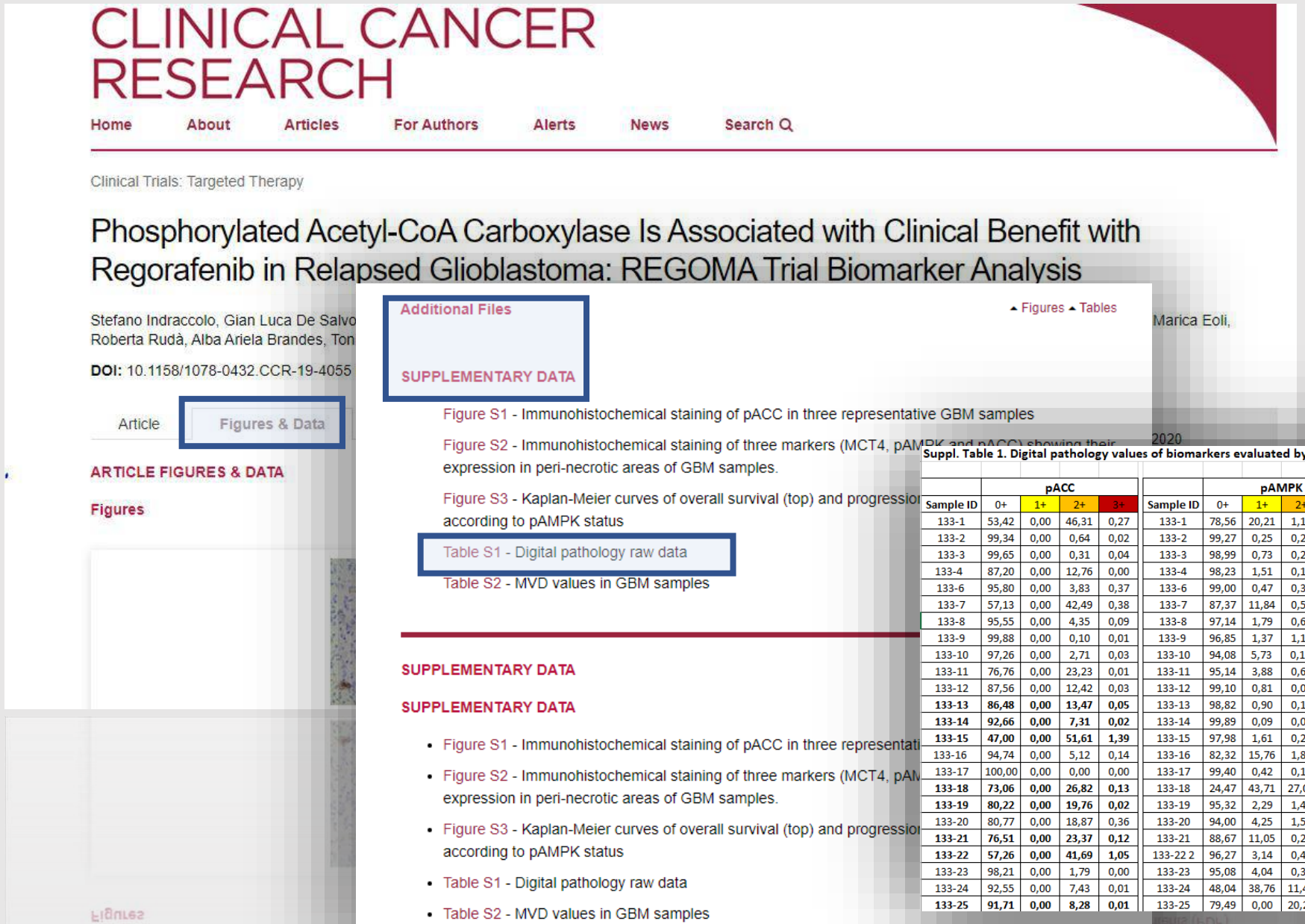
The SNOTEL daily data are available from the National Water and Climate Center of the Natural Resources Conservation Service at <http://www.wcc.nrcs.usda.gov/snow/> (last access 25 January 2020). The spatial data used in Fig. 1 were obtained from the US Geological Survey National Viewer Data set <*viewer.nationalmap.gov/advanced-viewer*>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

https://link.springer.com/article/10.1007/s11707-020-0813-5#article-info

Raw Data

If required by the publisher:

- link to an institutional open data repository
- if data are not directly accessible, link through contact information to a person who can grant the permission to their retrieval.



CLINICAL CANCER RESEARCH

Home About Articles For Authors Alerts News Search

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... Marica Eoli,

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

Article | Figures & Data

ARTICLE FIGURES & DATA

Figures

Additional Files

SUPPLEMENTARY DATA

Figures Tables

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

Figure S2 - Immunohistochemical staining of three markers (MCT4, pAMPK and pACC) showing their ... 2020 expression in peri-necrotic areas of GBM samples.

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

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

	pACC				pAMPK				MCT4				MCT1			
Sample ID	0+	1+	2+	3+	Sample ID	0+	1+	2+	3+	Sample ID	0+	1+	2+	3+	Sample ID	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 2	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

Data & article reported findings

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)

Data Management: main steps

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

Develop a clear picture of the data you need

- What is your theory
- What is your research question
- What is your theme/domain

Locate appropriate data resources

- Set up - and adjust - a search strategy to find suitable data for your research purposes
- Where looking for information: there are different types and modes of access to data
- Choose a safe place where storing your data (and learn how much it costs)

Set up a search query and search the data resource

- Understand that data repositories are important sources for discovering data
- If you decide to use data already stored in a database, learn how it works

Select data candidates

- Establish if all the data you selected are relevant for your research

Evaluate data quality

- Ask yourself questions on the description of your data
- Evaluate the quality and usefulness of data also for secondary analysis

https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/7.-Discover/The-process-of-data-discovery

Second step: file name strategy

A file name is the 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

- 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.

https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure

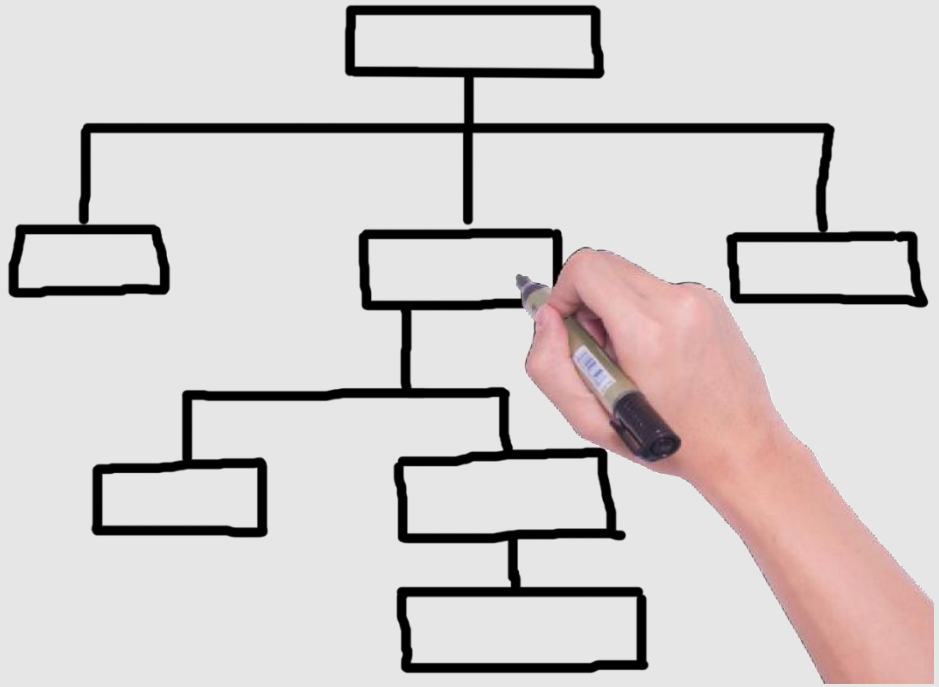
File name strategy …

From Propaganda Live, Friday 27/03/2020, La7

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.

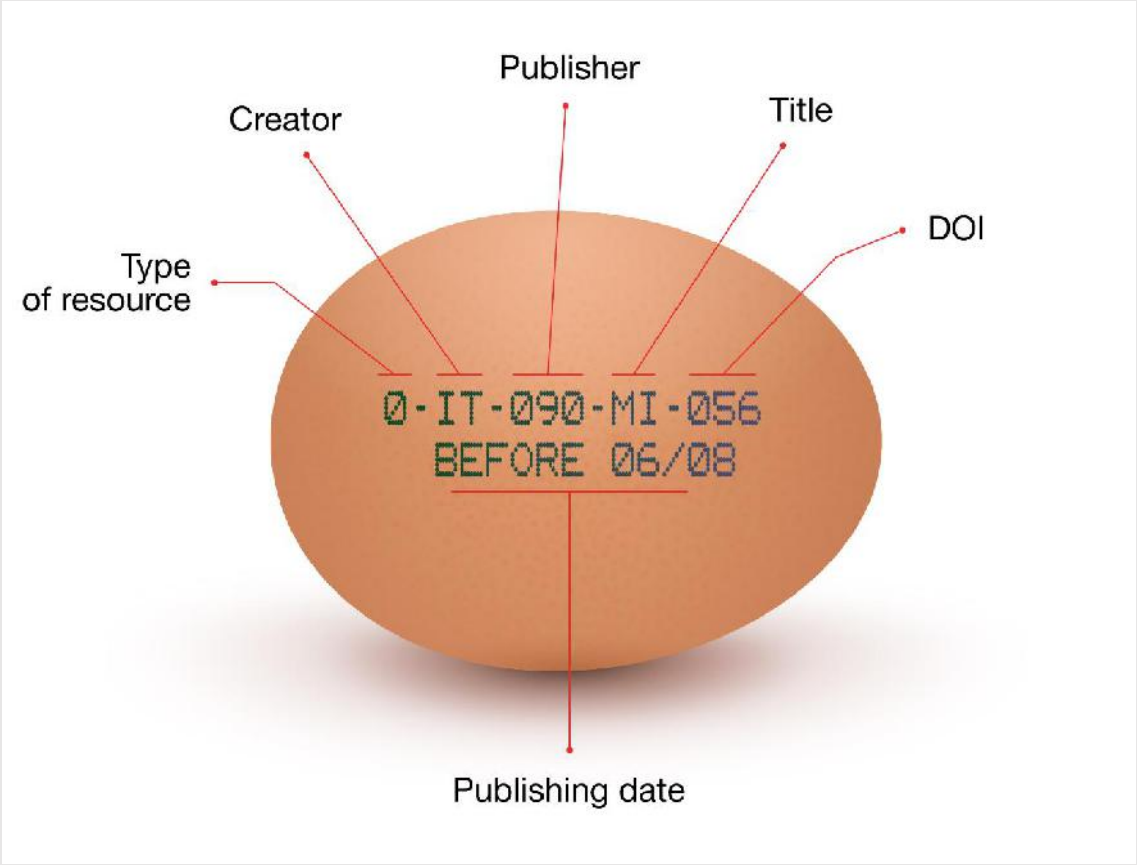


https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/2.-Organise-Document/File-naming-and-folder-structure

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

http://www.ucl.ac.uk/library/research-support/research-data/best-practices/guides/creating

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.

Guides on formats

Recommended formats

The five stars of open data

the number of stars increases if data are:

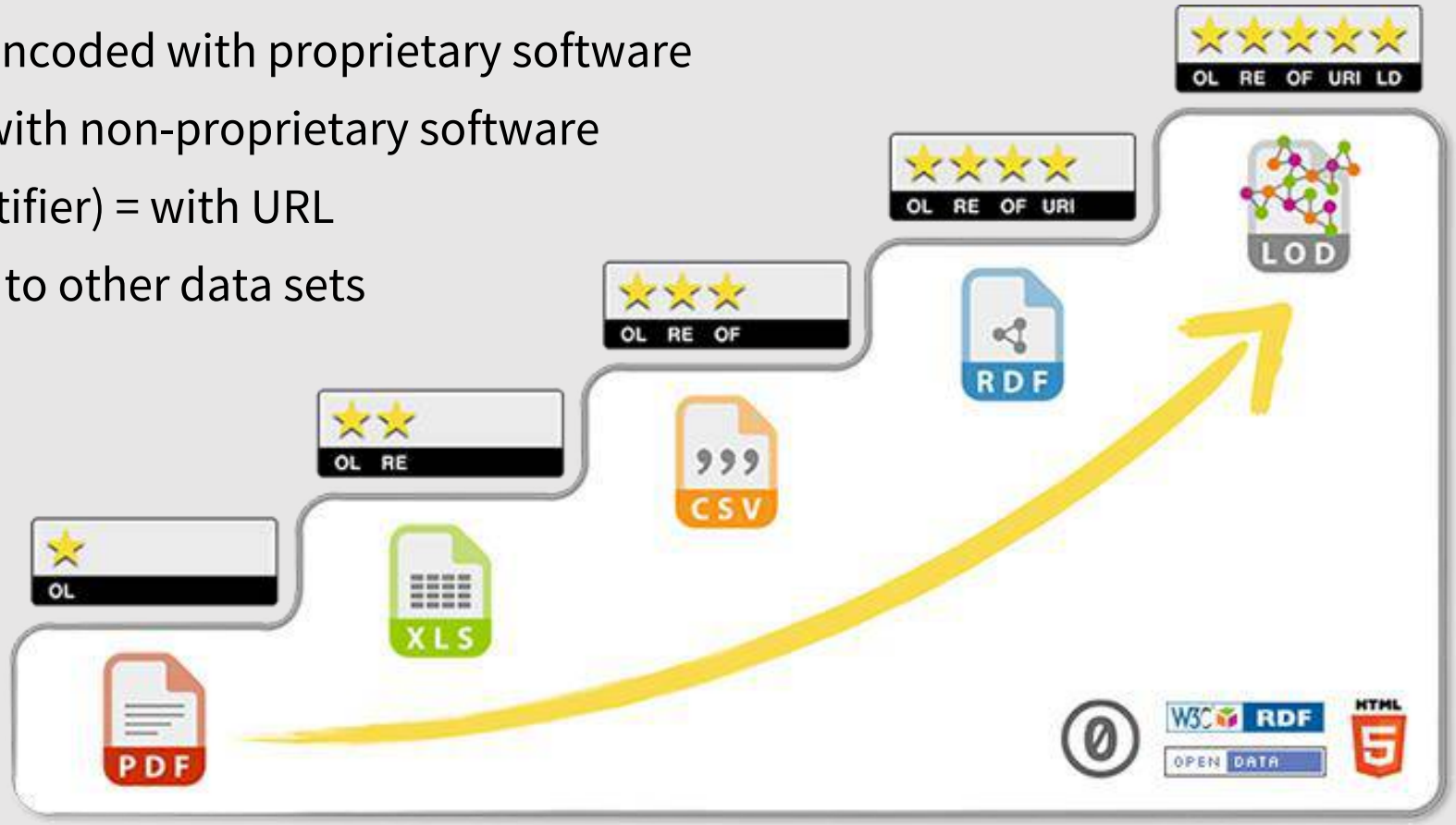
★ OL (OnLine) = distributed with an open license

★★ 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

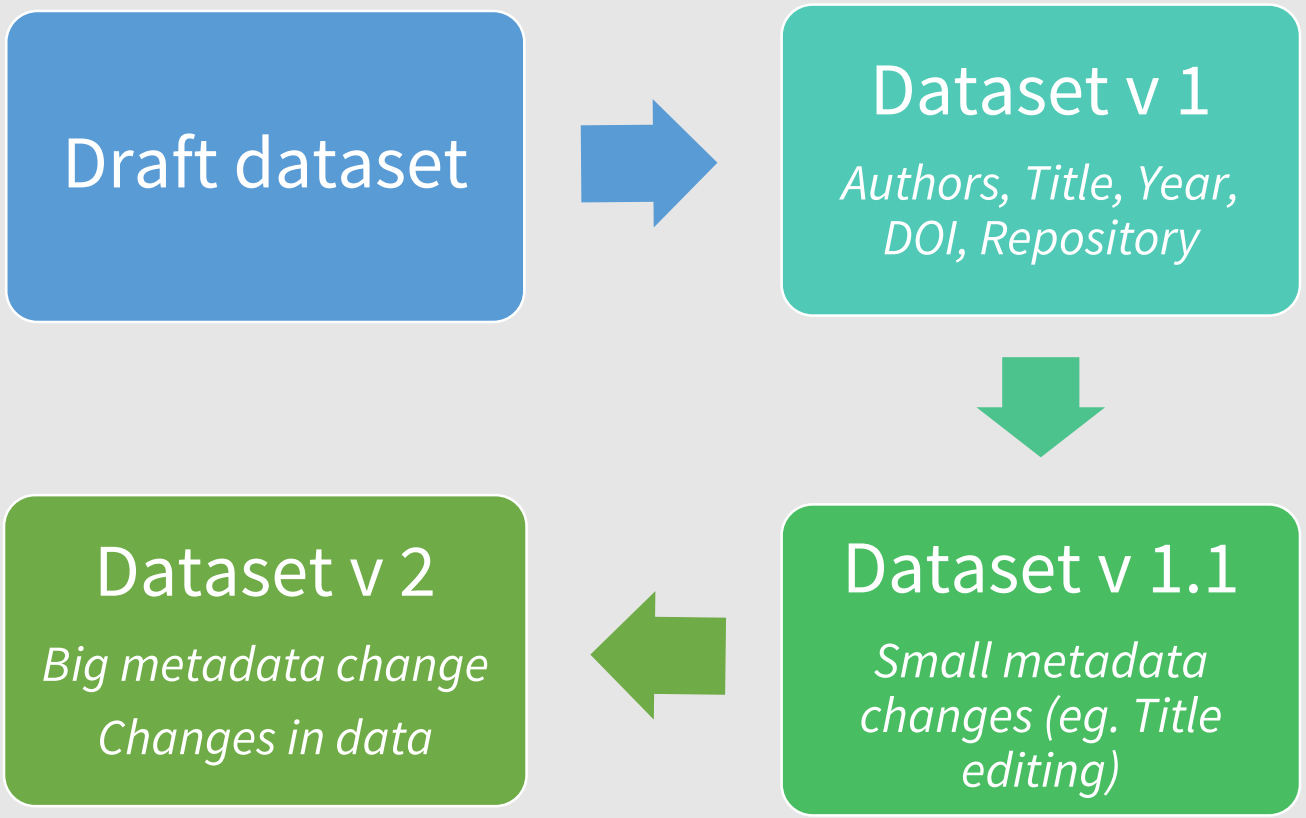


Data Management: more tips

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.



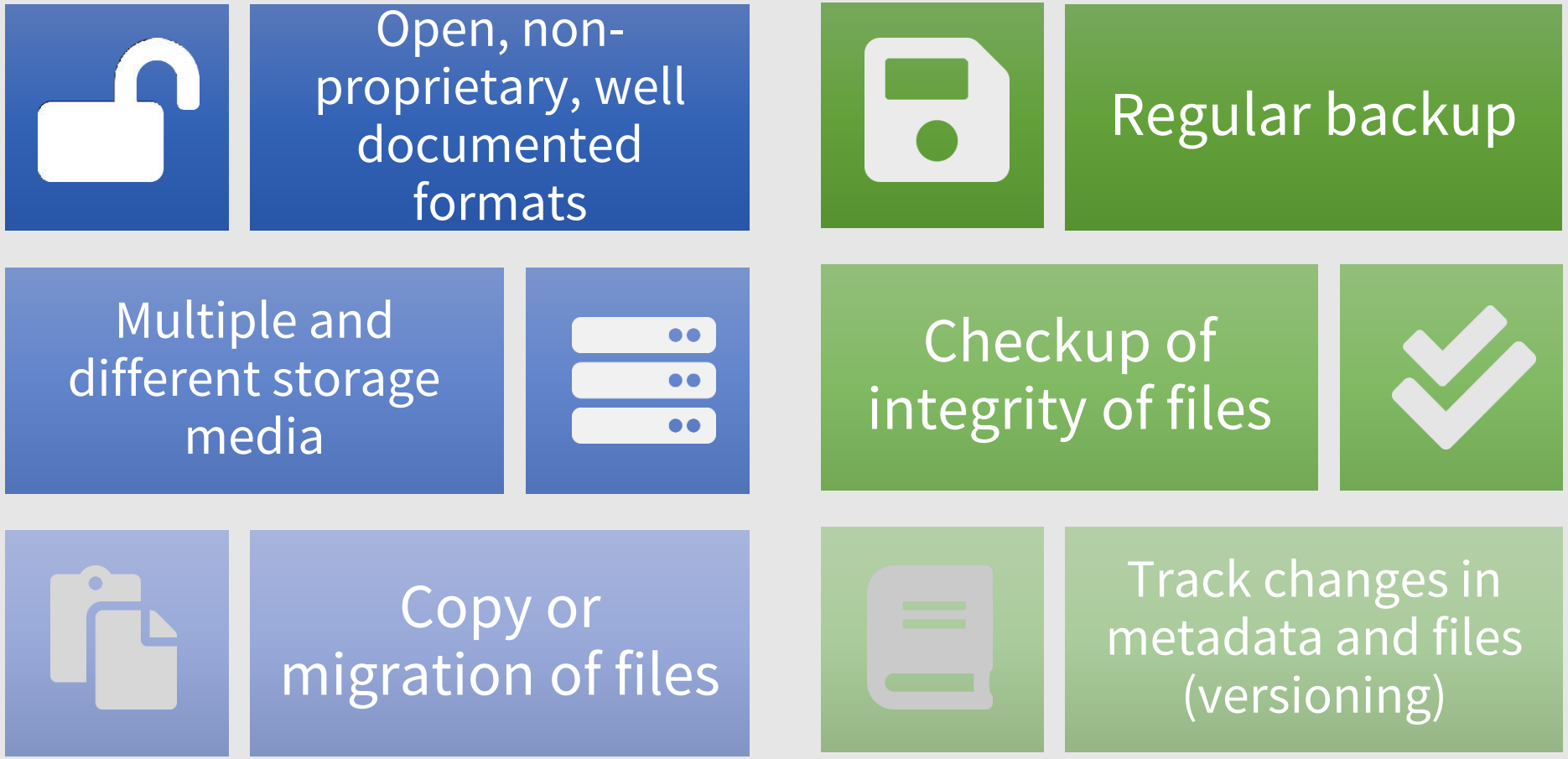
Useful guides on naming and version control

http://guides.dataverse.org/en/latest/user/dataset-management.html

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



Checklist for storage and preservation

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.

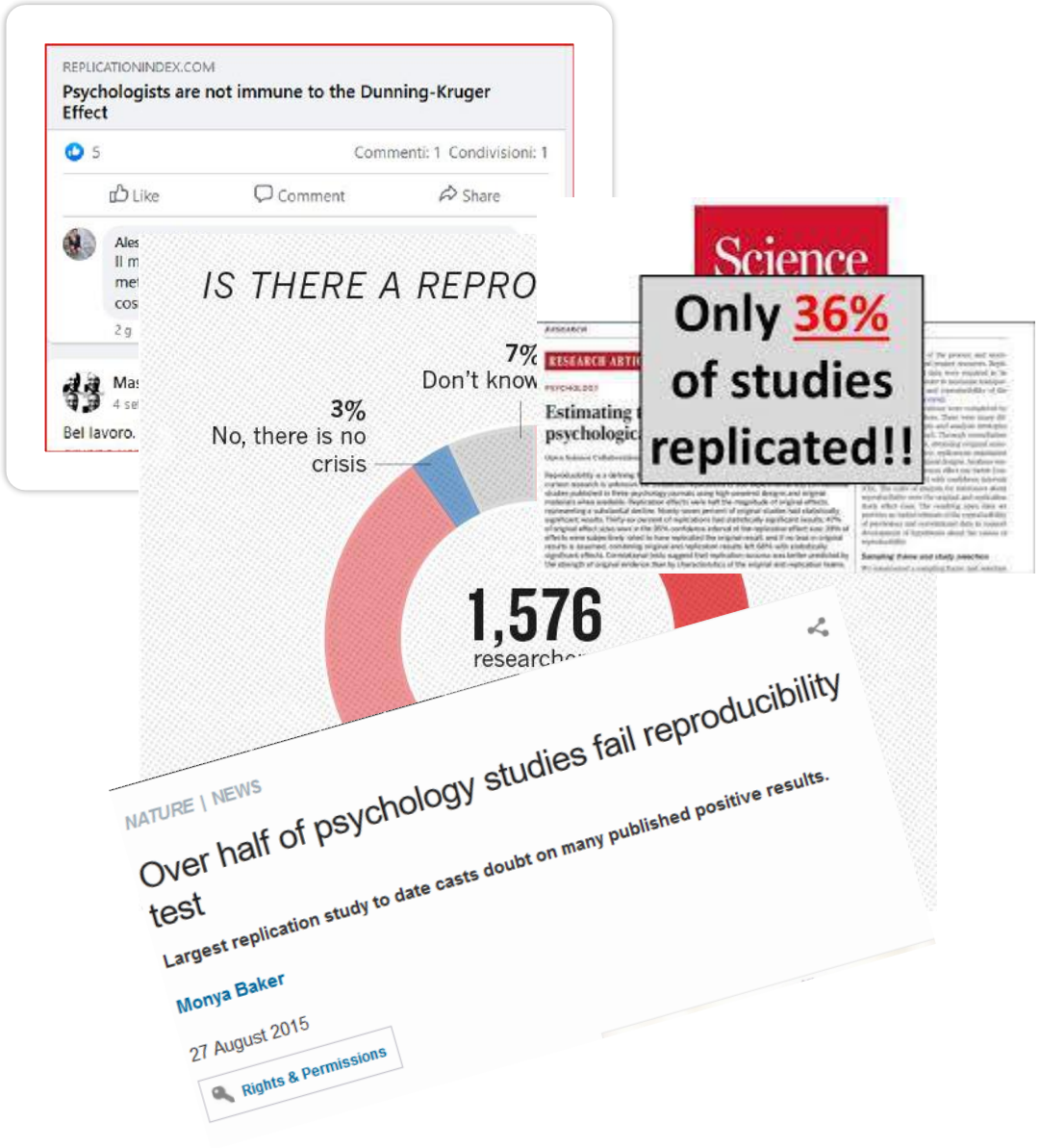
Here's the note from Woodgett, based at Lunenfeld-Tanenbaum Research Institute at Mount Sinai Hospital in Toronto:

...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: Psychology

Rete Italiana Open Science
https://www.facebook.com/groups/172297443522463/



Open Methodology

= the use of open methodologies throughout the entire research cycle

Foto di Ann H da Pexels

Open Notebooks

- https://openlabnotebooks.org
- https://theopennotebook.com/
- OpenLab/Notebook % Foster

- Code Ocean
- Protocols.io

Pre-registration

= the practice of pre-recording experiments

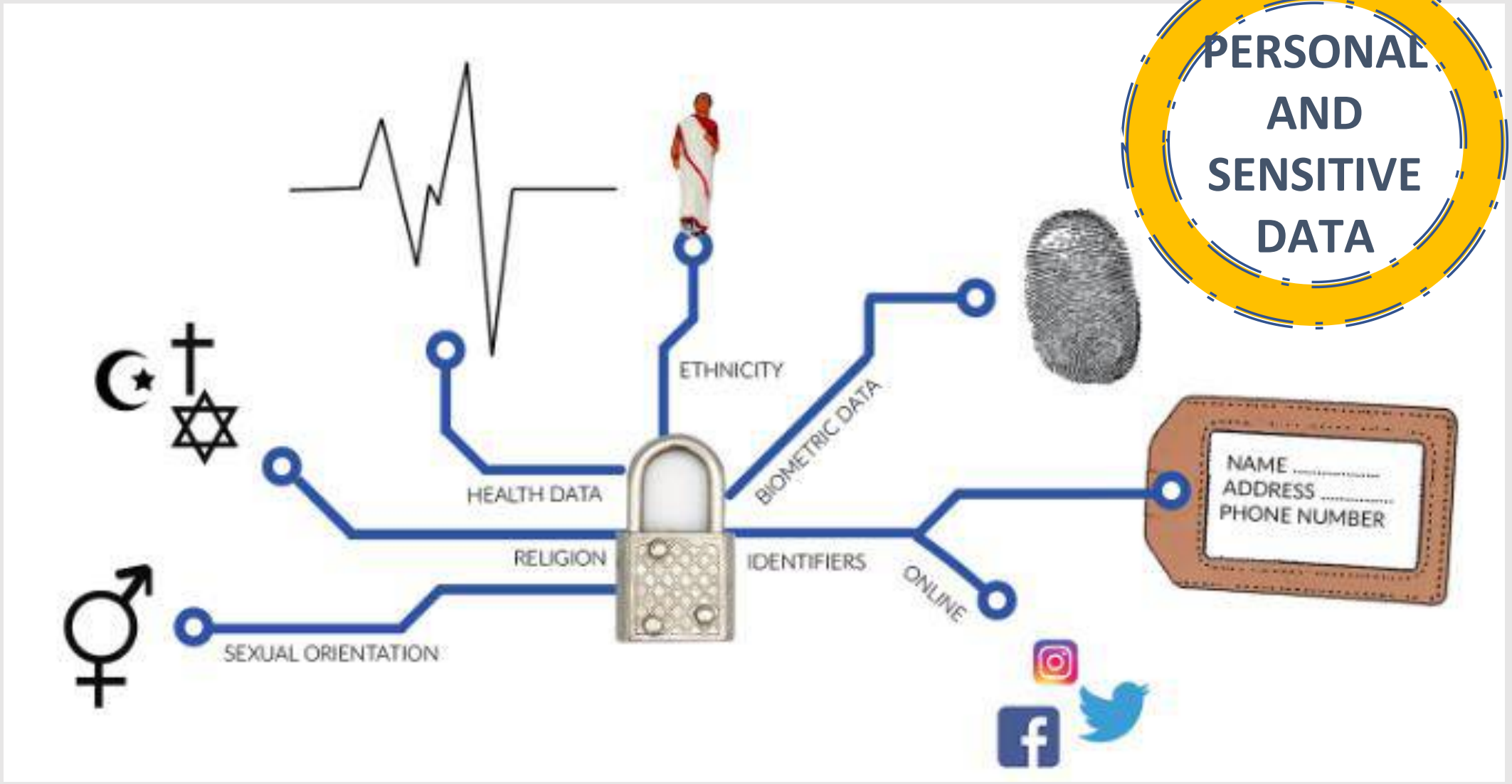
Retrospectively registered trials: the Editors' dilemma

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

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

Information about research integrity and research ethic

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

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.

When processing personal data, researchers should adhere to the following six principles:

I. Process lawfully, fair and transparent

II. Keep to the original purpose

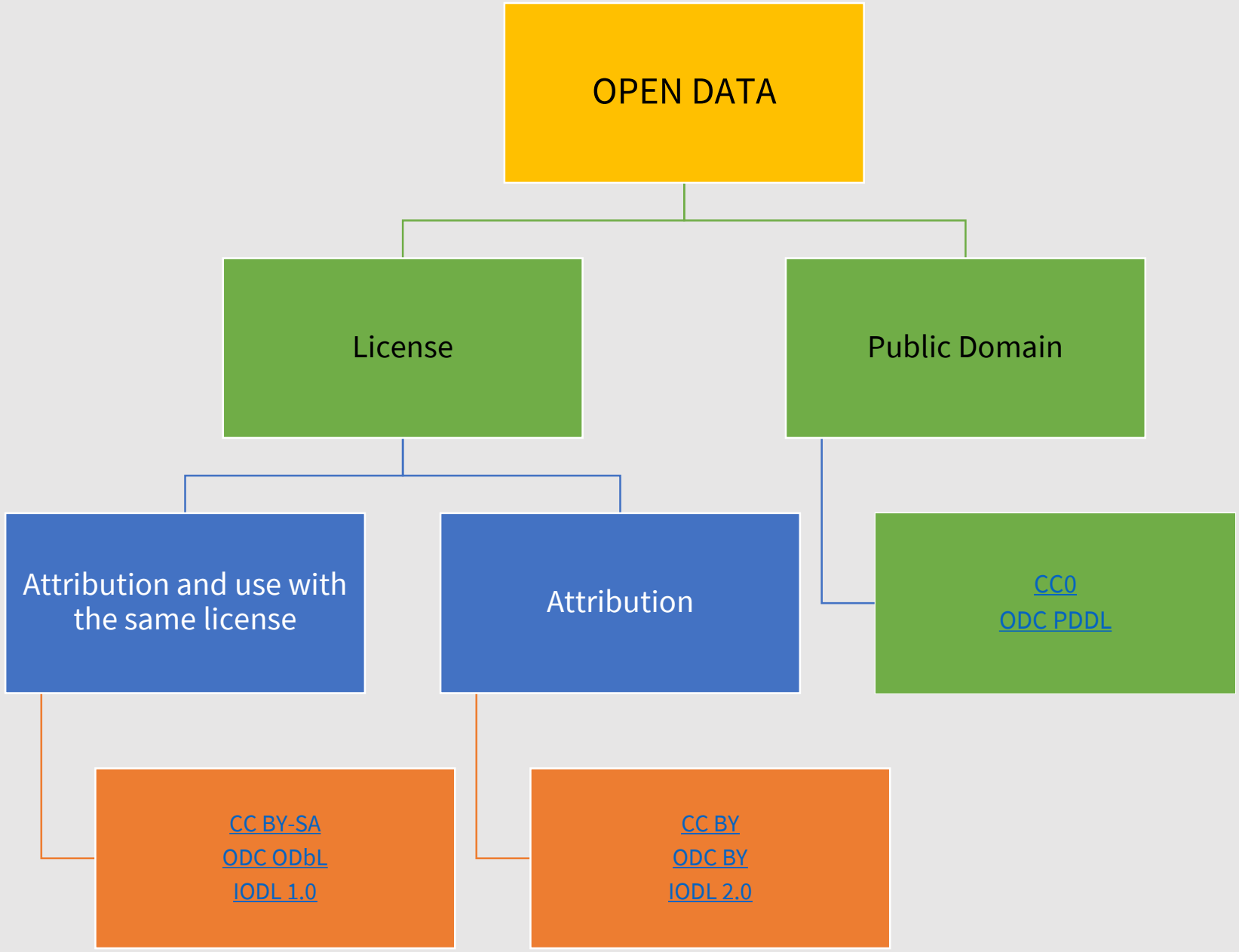
III. Minimise data size

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

Licenses for Open data



Here the main concerns about open data and the most effective answers

DMP and FAIRness

DMP = Data Management Plan

https://cerncourier.com/data-preservation-is-a-journey/

To be decided at the beginning of a project

Which data to preserve? In which formats?

Where preserving data?

Are there **costs** for preservation? (If yes, are they eligible inside research projects?)

Which data do **I want** to make accessible?

Which data do **I have** to make accessible?

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/

DMPTo
d your Data Management

Italian Open Science Support Group

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

OpenAIRE

- https://www.openaire.eu/what-isa-data-management-plan-and-how-do-i-create-one?highlight=WyJob3ciLCJ0byIsImNyZWF0ZSIsImRtcCIsImRtcCdzIiwiaG93IHRvIiwiaG93IHRvIGNyZWF0ZSIsInRvIGNyZWF0ZSJd

OpenAIRE

portage
VICES PARTAGÉS POUR LES DONNÉES DE RECHER
SHARED STEWARDSHIP OF RESEARCH DATA

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/ufficiserviziapplicazioni/course/view.php?id=112 (with SSO)

UNIVERSITÀ DEGLI STUDI DI PADOVA

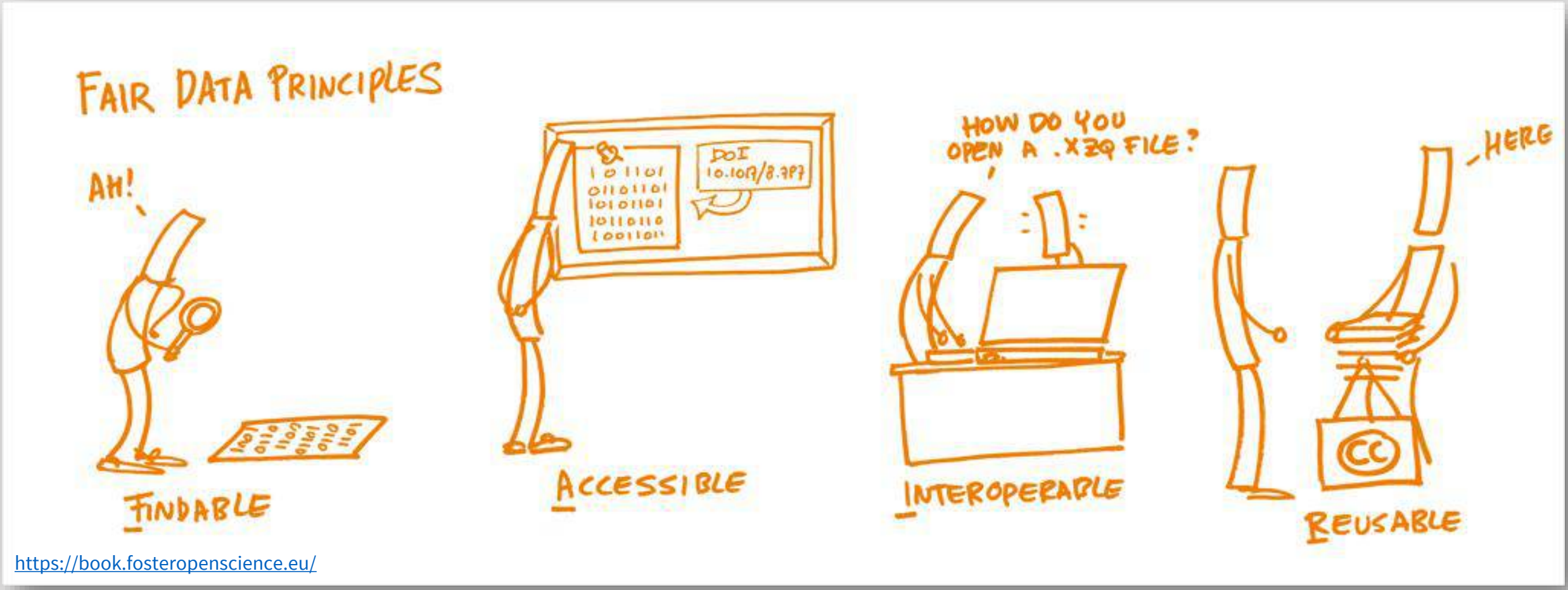
DMP: an example

CESSDA (Consortium of European Social Science Data Archives)

Link to pdf version here

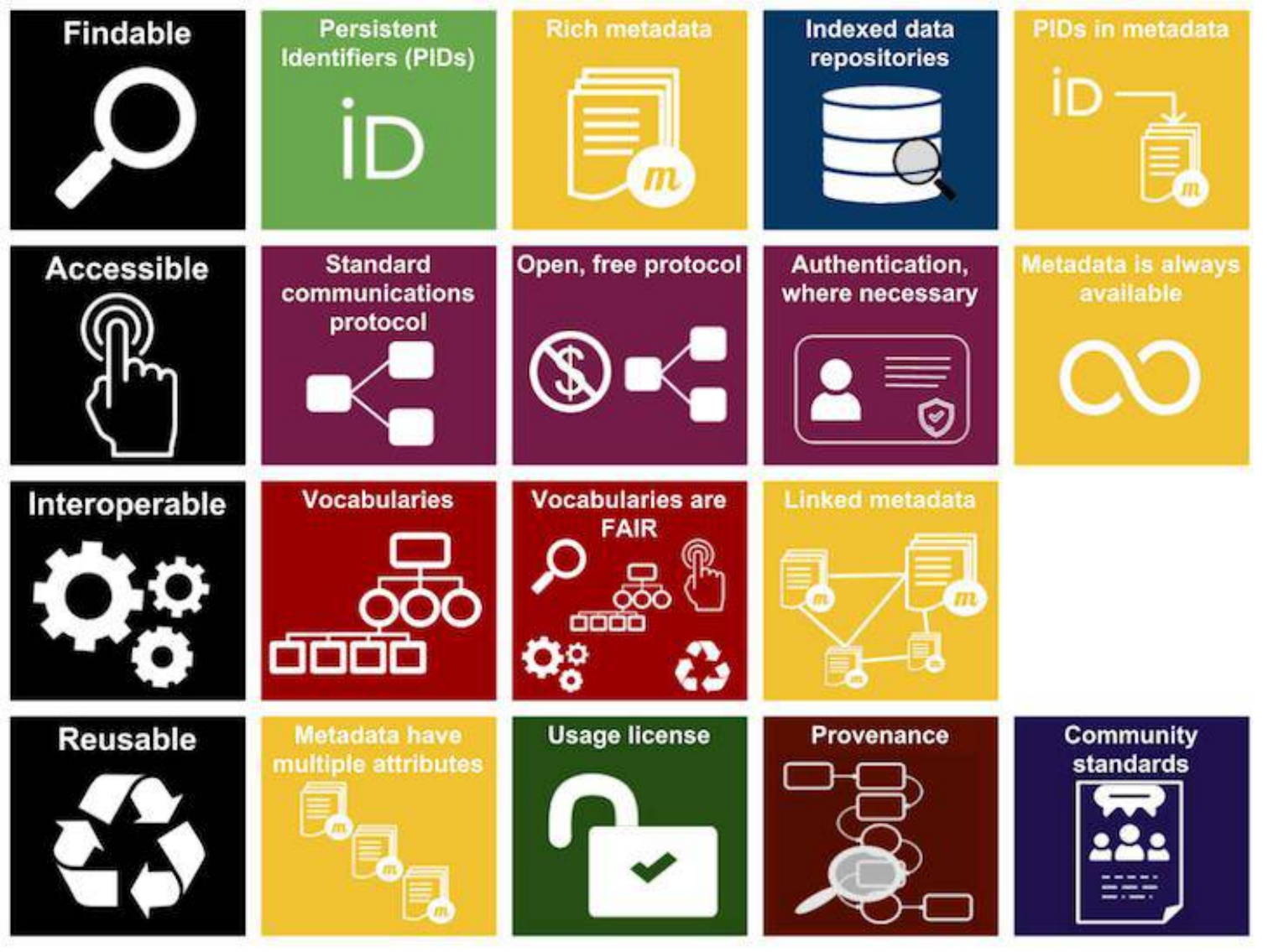
Link to editable version in this page

FAIR principles



https://book.fosteropenscience.eu/

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

Research Data Unipd

Open Data @ UniPD

The 1° December 2018 the Policy on the management of research data of the University of Padova entered into force.

WHO 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".

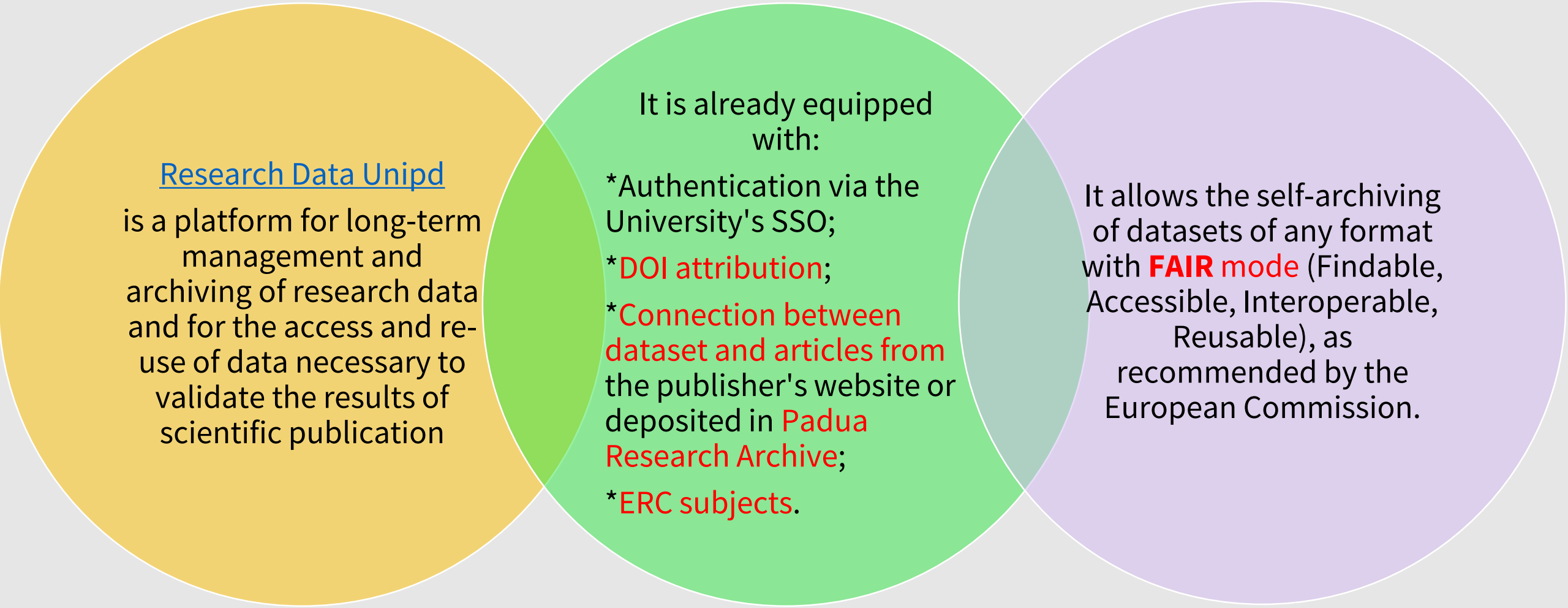
WHERE

“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".

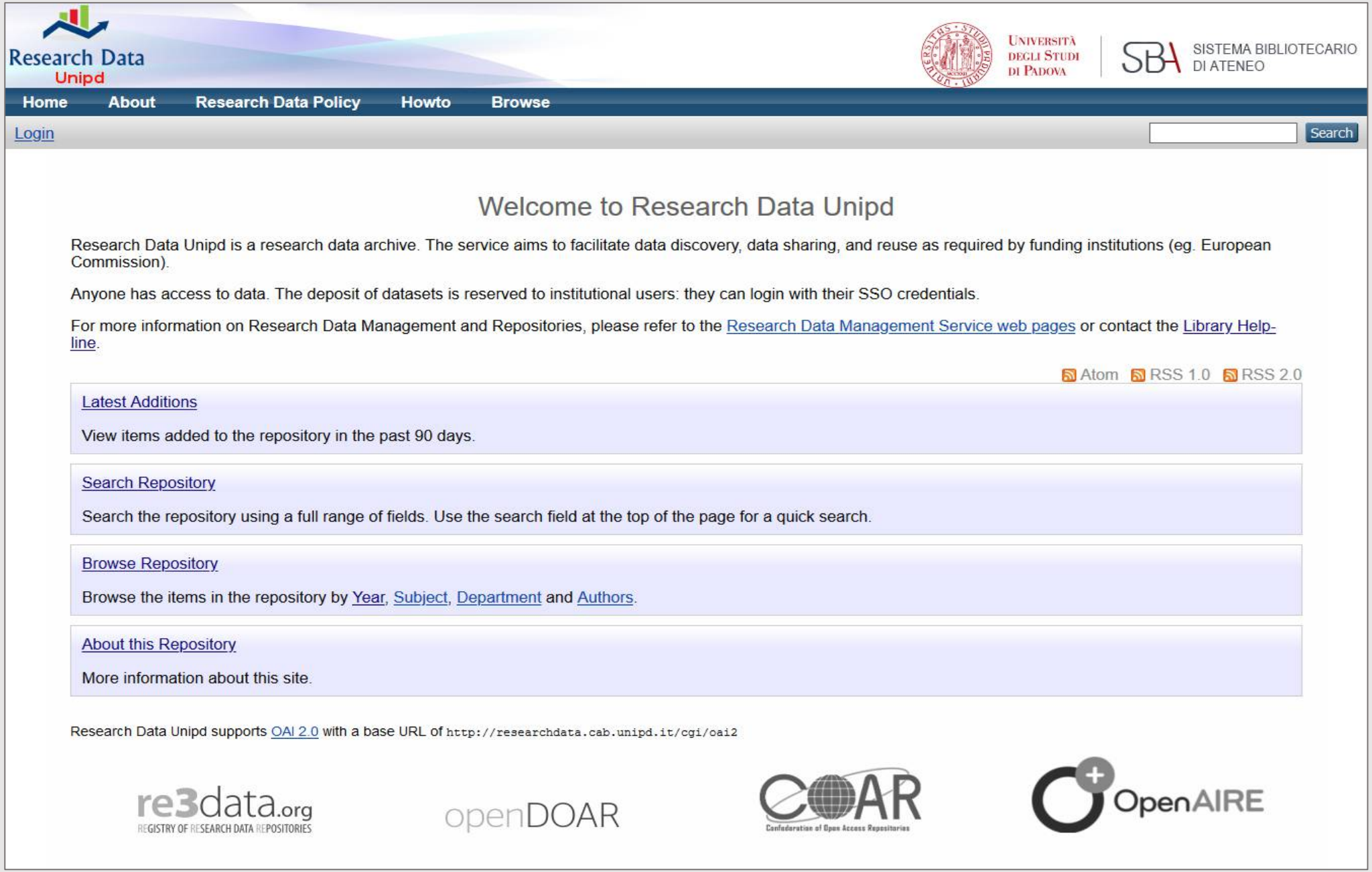
HOW

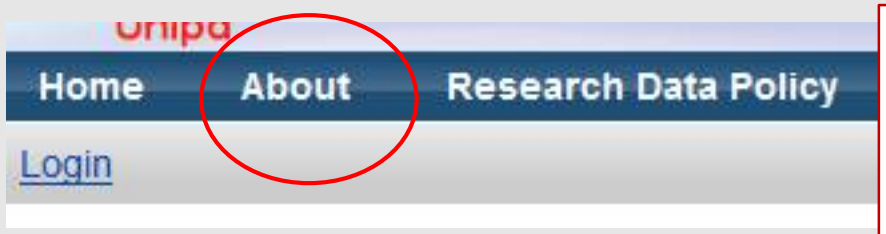
“Data must be stored correctly, completely, respecting their integrity. They must also be accessible, identifiable, traceable, interoperable and, where possible, available for subsequent use (FAIR principles)".

Research Data Unipd



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


Research Data
Unipd
Università degli Studi di Padova
SBA
Sistema Bibliotecario di Ateneo
Home
About
Research Data Policy
Howto
Browse
Login
Search
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
re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES
openDOAR
COAR
Confederation of Open Access Repositories
OpenAIRE



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

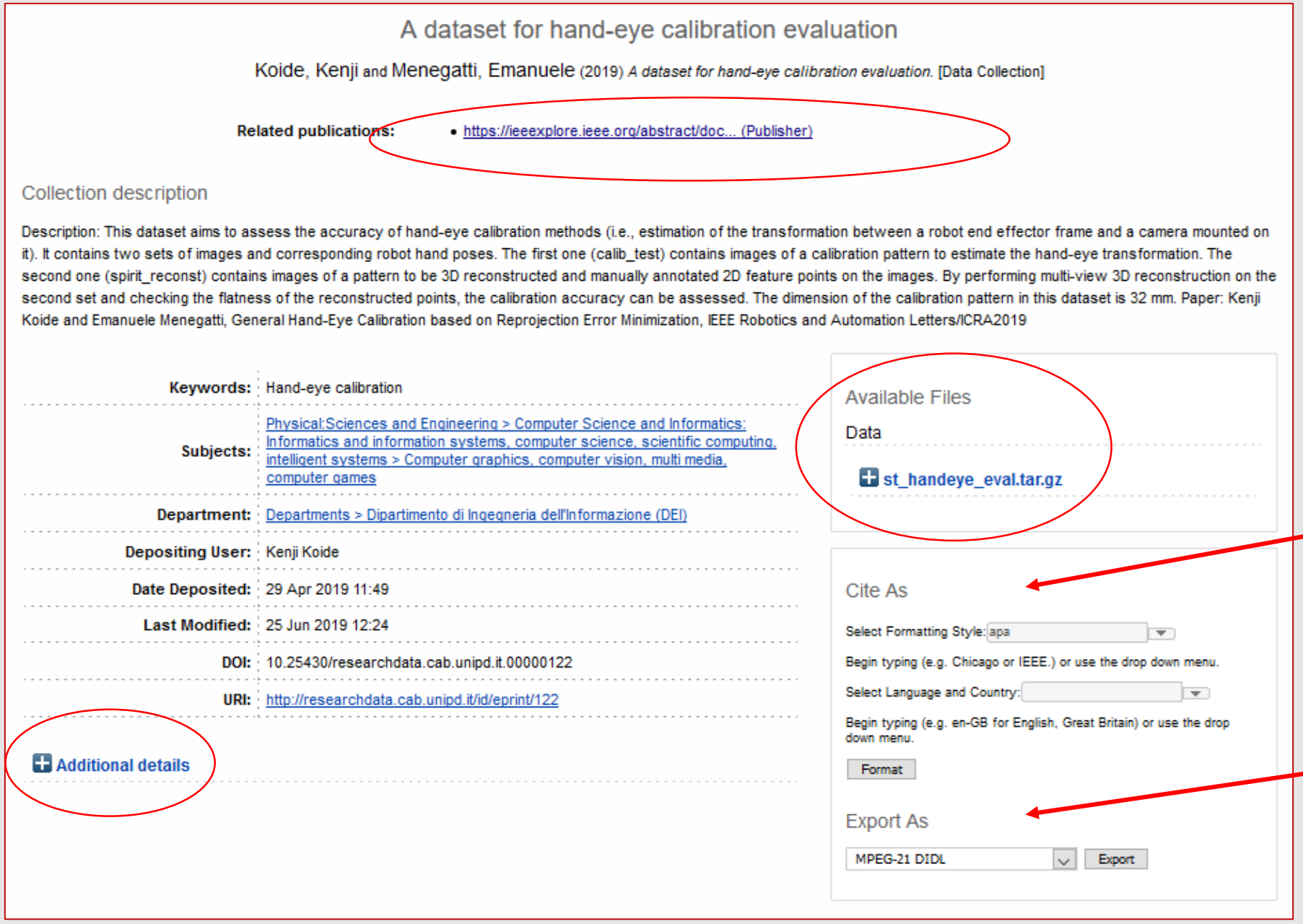
When you deposit data in the Research Data Unipd Archive, you will need to agree to the conditions below. This is done by clicking the "Deposit" button in the archive, before depositing the item.

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)

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: apa

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.

Format

Export As

MPEG-21 DIDL Export

Data Citation

Data citation refers to the practice of providing a reference to data in the same way as researchers routinely provide a bibliographic reference to outputs such as journal articles, reports and conference papers.

Main information required:

- Who produced the dataset (creator or author);
- The title of the dataset;
- The unique identifier of the dataset, preferably a Digital Object Identifier (DOI) or minimally a link to the dataset if it is online;
- The date the dataset was published and its version number, if it has one;
- The date and time the dataset was accessed;
- The distributor of the dataset.

Important elements in citing data, regardless of citation style, publisher or repository guidelines, can be found in this short overview by Purdue University.

https://www.ands.org.au/working-with-data/citation-and-identifiers/data-citation

Additional details and info on files

Additional details

Creators/Authors:	**Creators**	**Email**	**ORCID**
	Zane, Antonella	antonella.zane@unipd.it	orcid.org/0000-0001-7218-6068
Type of data:	Text		
Contributors:	**Contribution**	**Name**	**Email**
	Editor	Chavarria Arnau, Alexandra	UNSPECIFIED
	Editor	Brogiolo, Gianpietro	UNSPECIFIED
Collection period:	**From**	**To**	
	1999	2000	
Geographic coverage:	Italia - Veneto		
Data collection method:	Utilizzata microsonda eletronica (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 rivenute in Veneto.		

grafica, Alpi
ographic

Earth
geology,
eanography,
nvironmental
l resources
gneous

enze

000072

t/72

Available Files

Data

Monselice_ollar ... ci_Zane2017.PNG

Monselice_ollar ... io_Zane2017.PNG

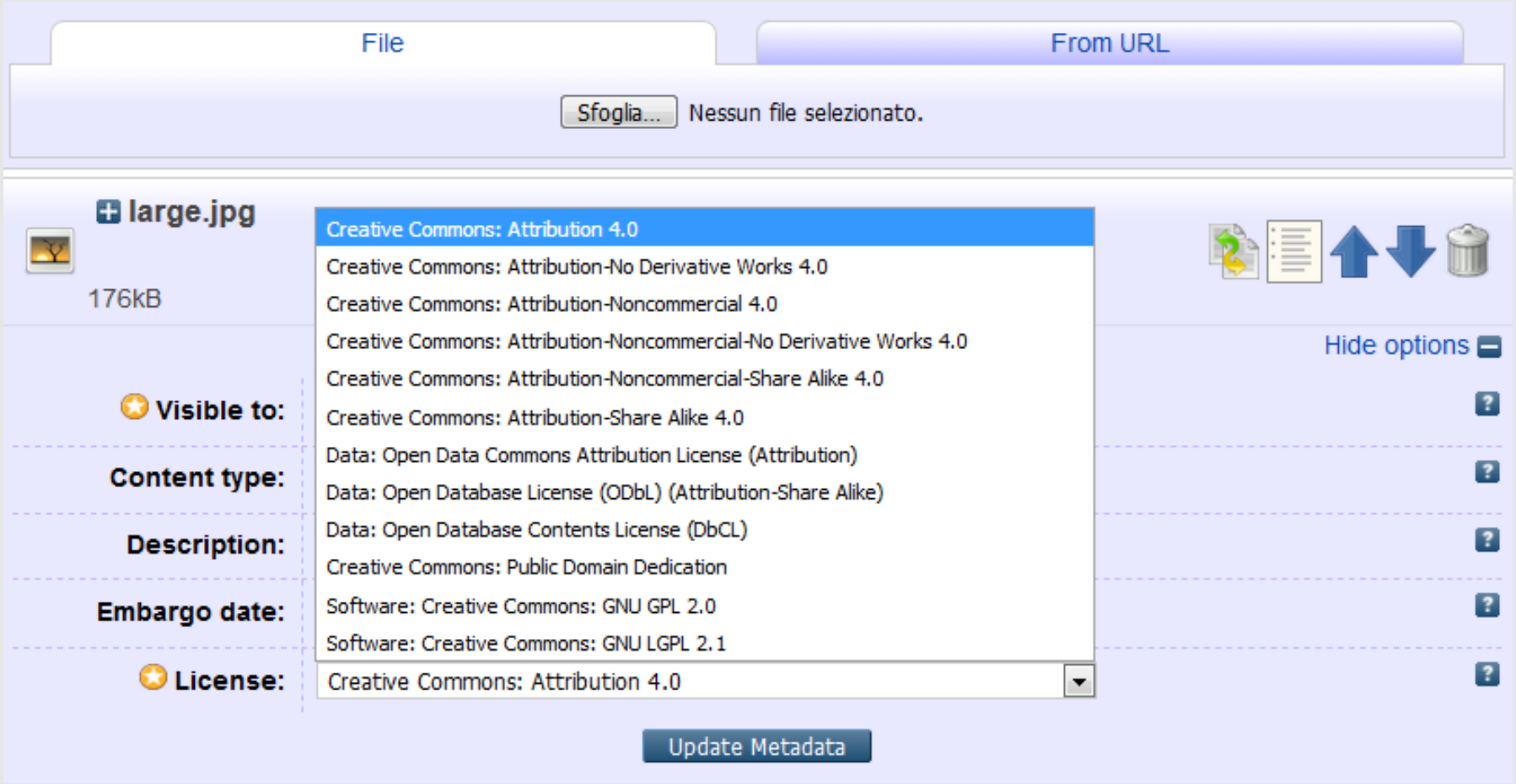
Visible to:	Anyone
Content type:	Data
Description:	microscopia
Metadata Revision:	3
Mime-Type:	image/png
License:	Creative Commons: Attribution 4.0
File size:	381kB

Read me

Monselice_readme_file.txt

Visible to:	Anyone
Content type:	["content_typename_readme" not defined]
Metadata Revision:	3
Mime-Type:	text/plain
License:	Creative Commons: Attribution 4.0
File size:	922B

Licenses to promote the reuse of data



Metadata (Details)

Creators/Authors

	Family Name	Given name / Initials	Email	ORCID
1.	Zane	Antonella	antonella.zane@unipd.	0000-0001-7218-6068
2.				
3.				
4.				

More input rows

Corporate creators

DOI

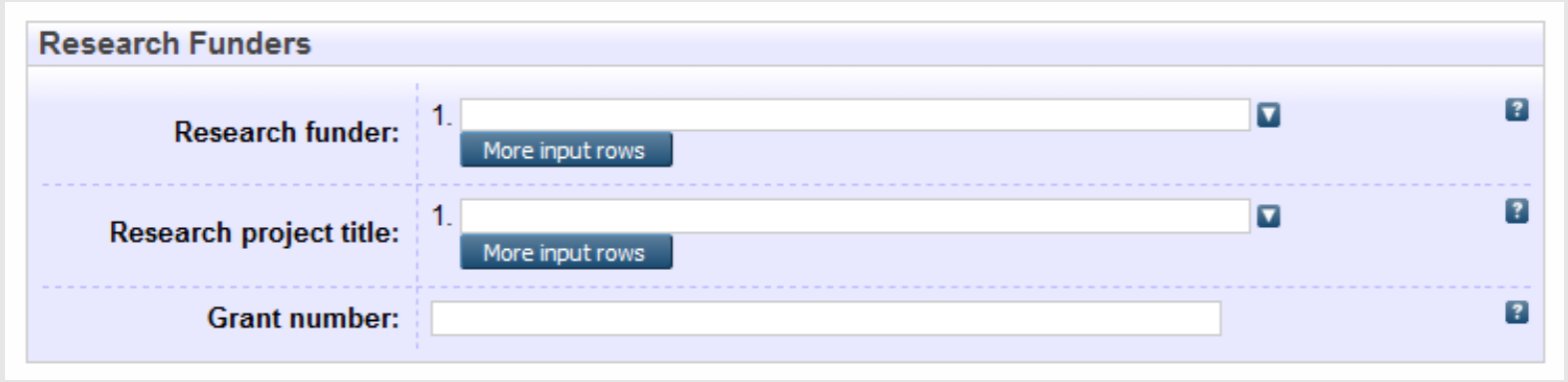
If this item has been given a Digital Object Identifier (DOI) when published elsewhere, please include it here.

10.25430/researchdata.cab.unipd.it.00000072

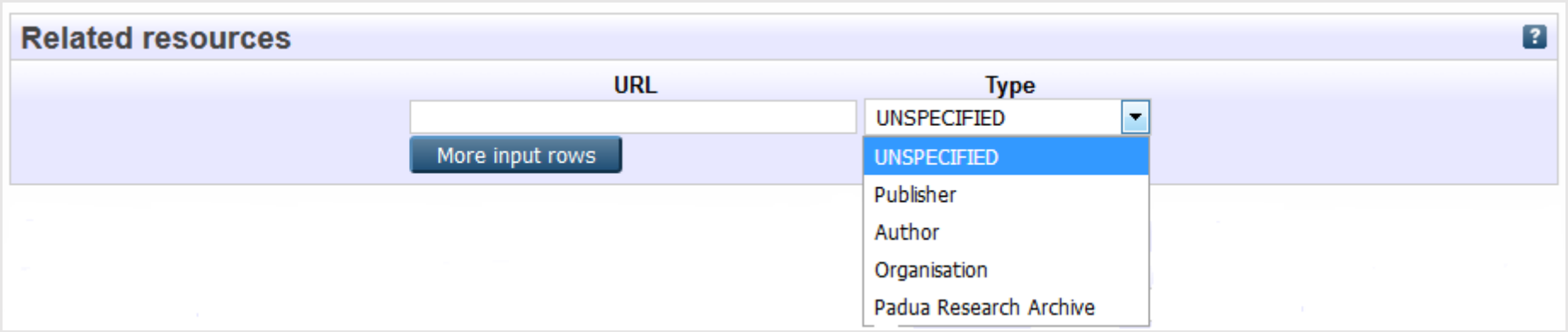
Type of data

- (●) **Text**
  If the dataset is mainly composed of text
- ( ) **Audio**
  If the dataset is mainly composed of audios
- ( ) **Video**
  If the dataset is mainly composed of videos
- ( ) **Image**
  If the dataset is mainly composed of images
- ( ) **Model**
  If the dataset is mainly composed of models
- ( ) **Software**
  If the dataset is mainly composed of software
- ( ) **Code**
  If the dataset is mainly composed of code
- ( ) **Machine/Instrument Log**
  If the dataset is mainly composed of Machine/Instrument log
- ( ) **Database**
  If the dataset is mainly composed of databases
- ( ) **Mixed**
  If the dataset is composed by mixed types
- ( ) **Other**
  If the dataset is mainly composed of other types not listed

Funders fields



Link to articles in publishers'websites or in Padua Research Archive (IRIS)



https://arxiv.org/abs/1907.02565

https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0230416

Subjects:	Digital Libraries (cs.DL)
DOI:	10.1371/journal.pone.0230416

The citation advantage of linking publications to research data

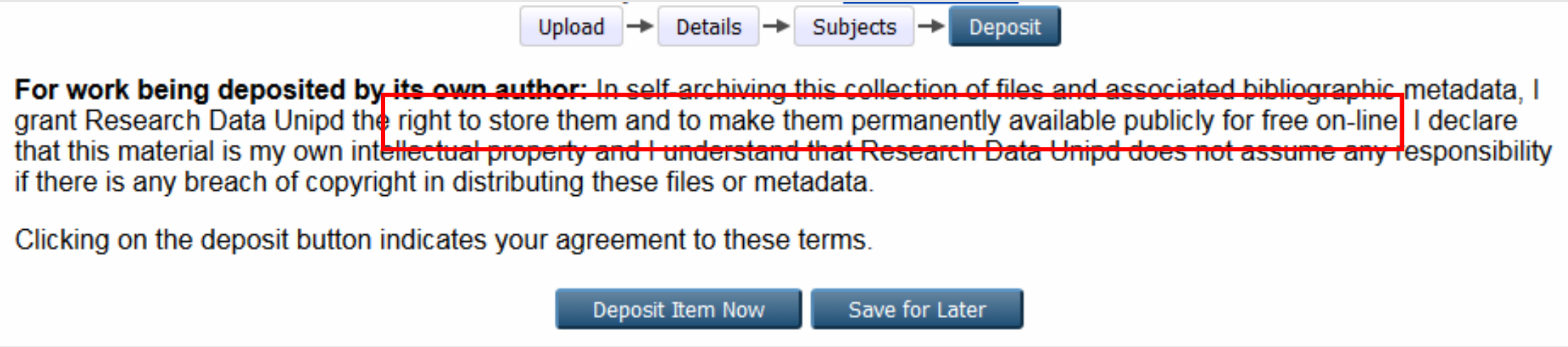
Giovanni Colavizza, Iain Hrynaszkiewicz, Isla Staden, Kirstie Whitaker, Barbara McGillivray

Efforts to make research results open and reproducible are increasingly reflected by journal policies encouraging or mandating authors to provide data availability statements. As a consequence of this, there has been a strong uptake of data availability statements in recent literature. Nevertheless, it is still unclear whaticles published by PLOS and BMC, develop an automatic system for labelling their data availability statements according to four categories based on their content and the type of data availability they display, and finally analyze the citation advantage of different statement categories via proportion of these statements actually contain well-formed links to data, for example via a URL or permanent identifier, and if there is an added value in providing such links. We consider 531,889 journal artregression. We find that, following mandated publisher policies, data availability statements become very common. In 2018 93.7% of 21,793 PLOS articles and 88.2% of 31,956 BMC articles had data availability statements. Data availability statements containing a link to data in a repository -- rather than being available on request or included as supporting information files -- are a fraction of the total. In 2017 and 2018, 20.8% of PLOS publications and 12.2% of BMC publications provided DAS containing a link to data in a repository. We also find an association between articles that include statements that link to data in a repository and up to 25.36% (

±

~1.07%) higher citation impact on average, using a citation prediction model. We discuss the potential implications of these results for authors (researchers) and journal publishers who make the effort of sharing their data in repositories. All our data and code are made available in order to reproduce and extend our results.

Licence to store and disseminate

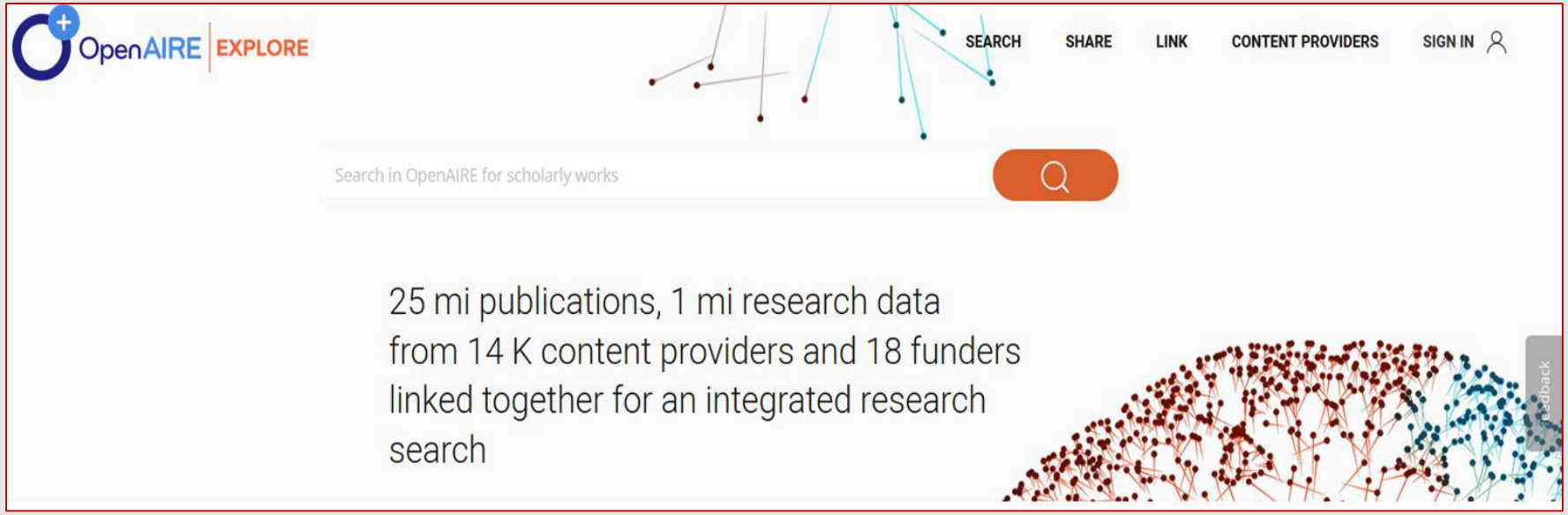


International registries of Open Access repositories

Research Data Unipd is indexed to re3data.org e OpenDOAR

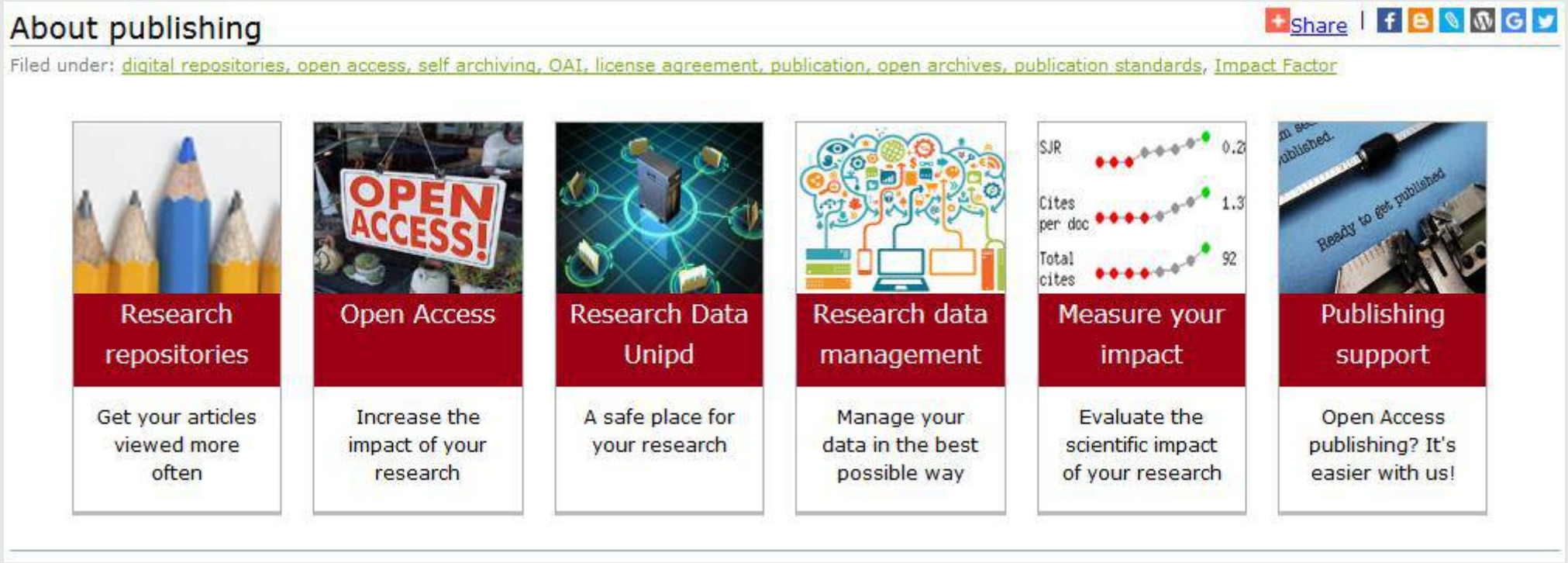
Make your content count with OpenAIRE

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.



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.



Library System support services

Authors can submit specific requests using the Library System Help Service, choosing between the following addresses:

- 09 Tesi di dottorato (Padua@research)
- 11 Supporto Open Access (Supporto Ricerca)

Kyle James ttps://www.flickr.com/photos/jameskm03/2711755476

Library System support services

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

Scholarly Communication
and principles of Open Science

a Training Course composed by 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).

Thank you!

Contact us:

pinali.moderna@unipd.it

biblioteca.digitale@unipd.it

Support:

https://bibliotecadigitale.cab.unipd.it/en/helpline