Research Data Management

Mario Locati

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 "Education and Research" - Component 2: "From research to business" - Investment 3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures"

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani

Index

1. Open Science
2. European Union plan
3. Italian plan
4. Research Data Management
5. FAIR data principles
6. Italian initiatives
7. Tools for DMPs

Open Science

Definition(s) of Open Science

Scientific knowledge of all kinds should be **openly shared as early as is practical in the discovery process**. This includes making research data and results freely available to the public and **encouraging collaboration** among scientists and with the public. Michael Nielsen, Reinventing Discovery, 2011

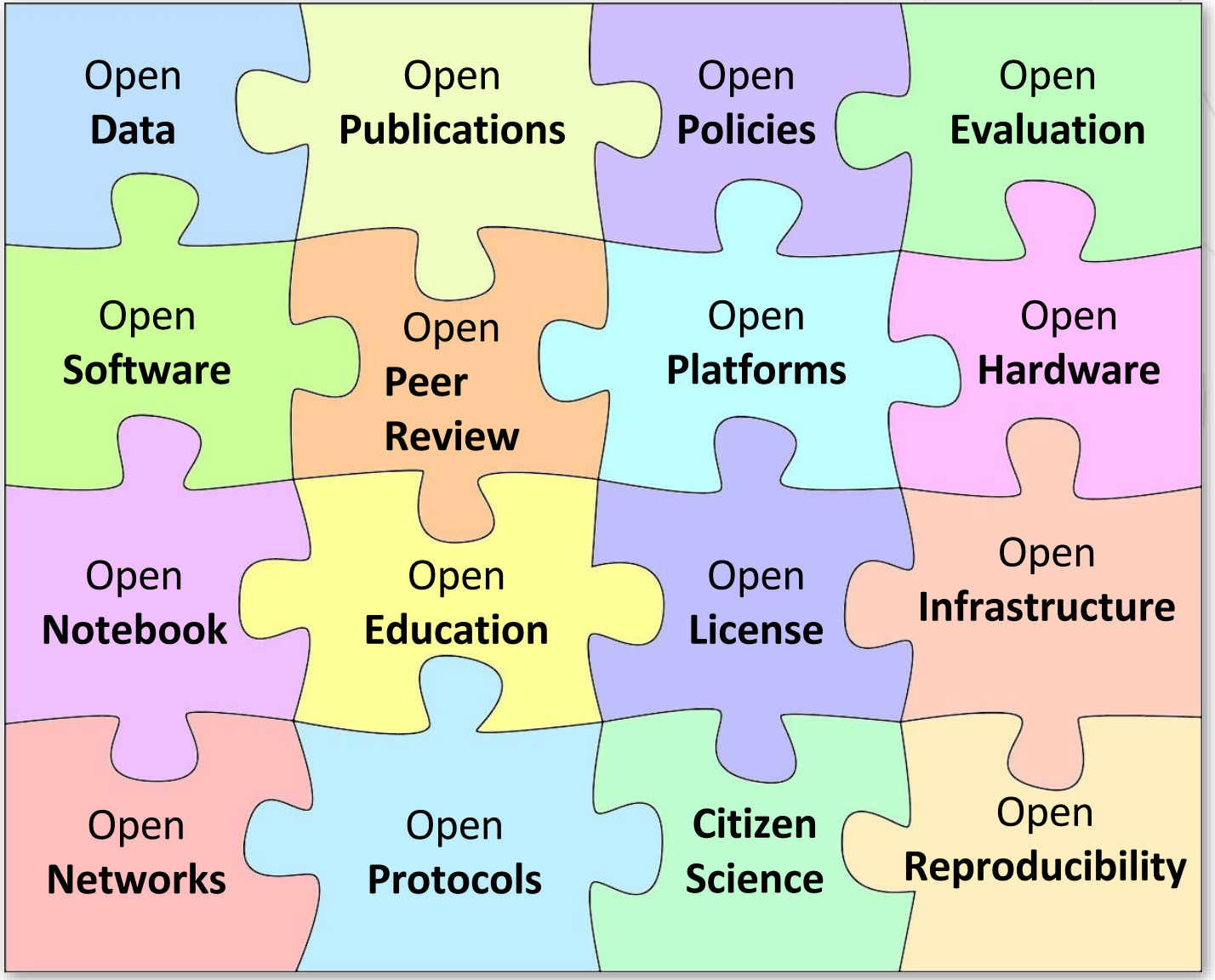
Practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, **under terms that enable reuse, redistribution and reproduction** of the research and its underlying **data and methods**. EC Project Foster, 2017-2019

Open science is a set of **principles and practices** that aim to make scientific research from all fields accessible to everyone for the **benefit of scientists and society** as a whole. Not only scientific knowledge should be accessible, but also the production of that **knowledge itself is inclusive, equitable and sustainable**. UNESCO, 2021

Components of Open Science

Cartoon by Nuthawut Somsuk at https://www.vecteezy.com/

Components of Open Science



CARE principles, society and value-driven data strategy

Collective benefit

Society should benefit from data and Research aligns with the **needs and goals of the community**.

Authority to control

Society must have the power to make decisions regarding **data governance** and how they are represented in the data. Data must be **accessible to society**.

Responsibility

Researchers are accountable to society and must be able to demonstrate how their use of data produced by society benefits the society and their self-determination. Evidence of benefits must be shared, and resources must be grounded supporting **multiple languages and world-views**.

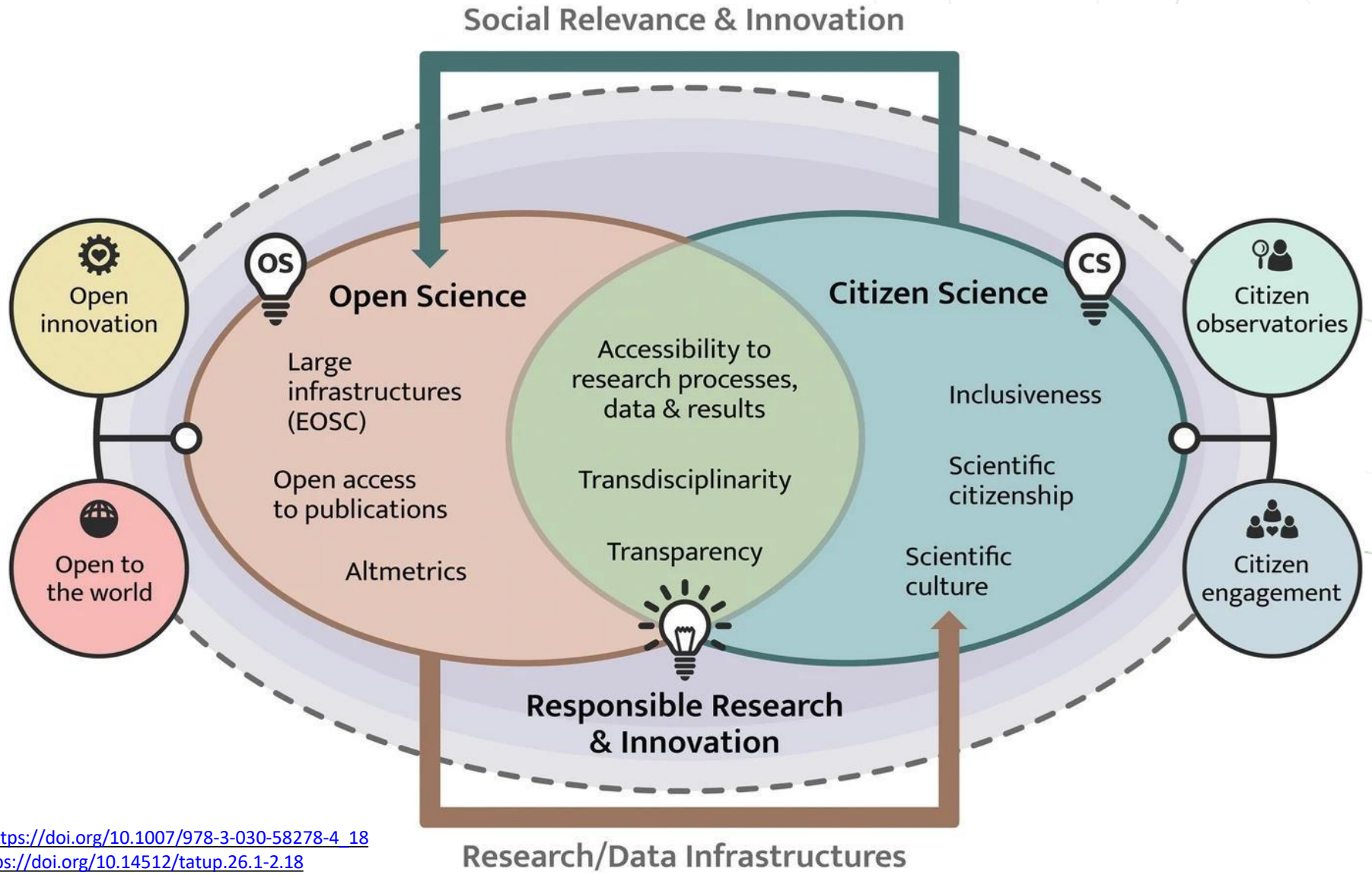
Ethics

Society **rights and well-being** are central during all stages of the data life-cycle.
Risk assessments must be considered from the society perspective.

Cartoon by Nuthawut Somsuk at https://www.vecteezy.com/

Source: Carroll et al. 2020, https://doi.org/10.5334/dsj-2020-043
Carroll et al. 2021, https://doi.org/10.1038/s41597-021-00892-0

Relationship between Open Science and Citizen Science



Source: Schade et al. 2021, https://doi.org/10.1007/978-3-030-58278-4_18
Vohland and Göbel 2017, https://doi.org/10.14512/tatup.26.1-2.18

European Union plan

European Union - Decisions

European Council
Represents the interests of National Governments
General political direction and priorities of the EU

Council
Ministries of the EU Countries
Negotiate legislation in conjunction with the EU Parliament

Parliament
Represents the interests of the Citizens
Legislative, Supervisory, Budgetary roles

Commission
Represents the interests of the Union
Legislative, Supervisory, Budgetary roles

DECISIONS

European Research Area ERA
Directorate-General for
Research and Innovation (RTD)

Single Market

Digital Single Market

European Union - Data Strategy

European Research Area ERA
Directorate-General for
Research and Innovation (RTD)

Single Market
Digital Single Market

Digital Programme Europe

Common European Data Spaces

Directorate-General for Communications Networks, Content and Technology (CNECT)

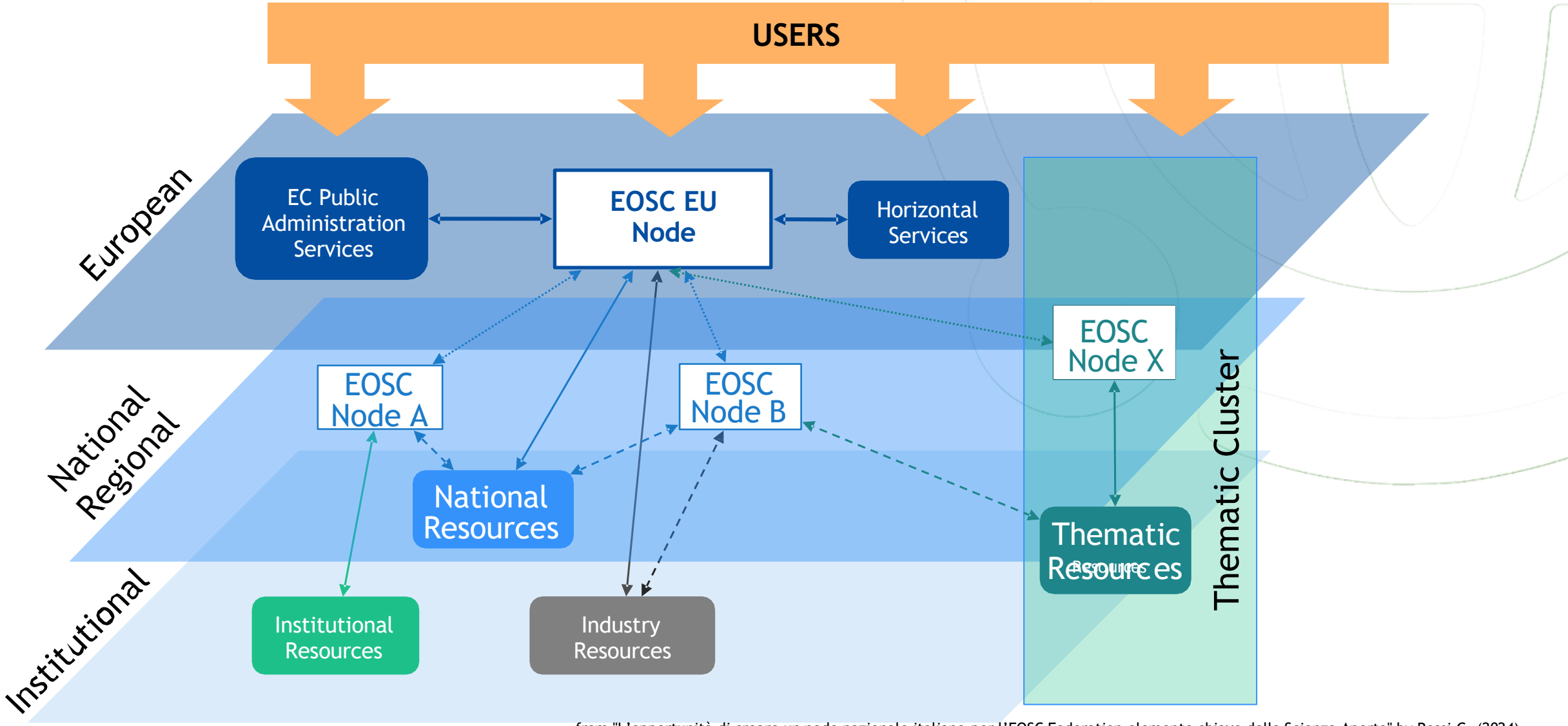
DATA SPACES SUPPORT CENTRE

Health | Industrial & Manufacturing | Agriculture | Culture | Mobility | Green Deal | Security | Public Administration | Finance | Skills | Energy | Media

European Open Science Cloud (EOSC)
Directorate-General for Research and Innovation (RTD)

EOSC EU Node
Directorate-General for Communications Networks, Content and Technology (CNECT)

EOSC Federated “System of Systems”



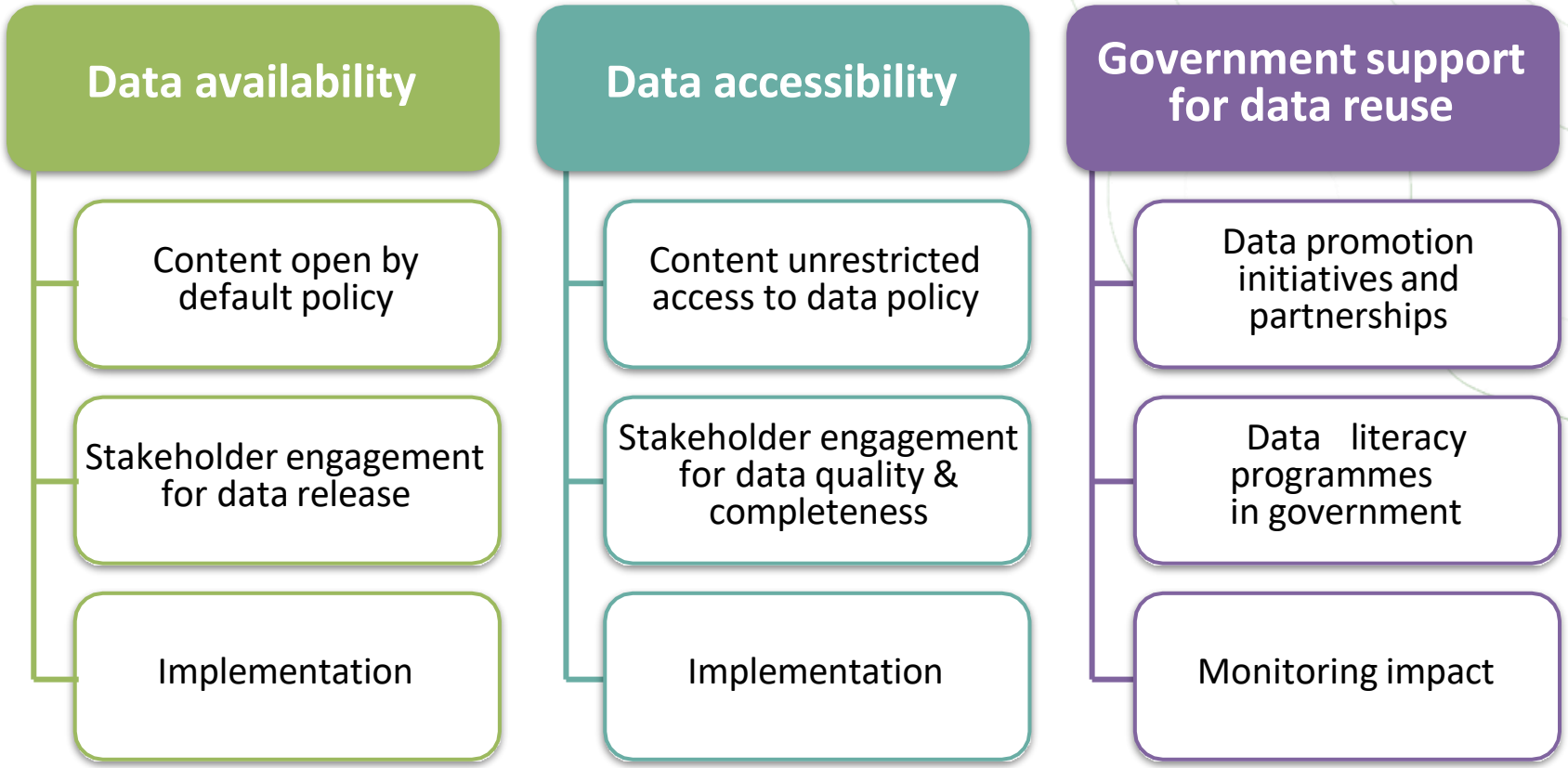
from "L’opportunità di creare un nodo nazionale italiano per l’EOSC Federation elemento chiave della Scienza Aperta" by Rossi G. (2024).
https://agenda.infn.it/event/42629/contributions/239395/attachments/123838/181791/Presentazione%20NODO%20EOSC%20CoPER-CRUI.pdf

EU legislation of interest

1. Commission Recommendation (EU) 2018/790 of 25/04/2018 on access to and preservation of scientific information
2. Regulation (EU) 2018/1807 of the EU Parliament and of the Council of 14/11/2018 on a framework for the free flow of non-personal data in the EU
3. Directive (EU) 2019/790 of the EU Parliament and of the Council of 17/04/2019 on copyright and related rights in the Digital Single Market
4. Directive (EU) 2019/1024 of the EU Parliament and of the Council of 20/06/2019 on open data and the re-use of public sector information. **Public Sector Information, PSI**
5. European Commission, DG for Research and Innovation (2022). Tackling R&I foreign interference: staff working document
6. Regulation (EU) 2022/868 of the EU Parliament and of the Council of 30/05/2022 on EU data governance and amending Regulation (EU) 2018/1724. **Data Governance Act**
7. Regulation (EU) 2022/1925 of the EU Parliament and of the Council of 14/09/2022 on contestable and fair markets in the digital sector. **Digital Markets Act**
8. Regulation (EU) 2022/2065 of the EU Parliament and of the Council of 19/10/2022 on a Single Market For Digital Services. **Digital Services Act**
9. Council Recommendation (EU) 2022/2415 of 2/12/2022 on the guiding principles for knowledge valorisation
10. Commission Implementing Regulation (EU) 2023/138 of 21/12/2022 laying down a list of specific **high-value datasets** and the arrangements for their publication and re-use
11. Regulation (EU) 2023/2854 of the EU Parliament and of the Council of 13/12/2023 on harmonised rules on fair access to and use of data. **Data Act**
12. Commission Recommendation (EU) 2023/499 of 1/03/2023 on the management of intellectual assets for knowledge valorisation in the European Research Area
13. Regulation (EU) 2024/903 of the EU Parliament and of the Council of 13/03/2024 for a high level of public sector interoperability across the Union. **Interoperable Europe Act**
14. Regulation (EU) 2024/1689 of the EU Parliament and of the Council of 13/06/2024 laying down harmonised rules on artificial intelligence. **Artificial Intelligence Act**
15. Regulation (EU) 2024/2847 of the EU Parliament and of the Council of 23/10/2024 on horizontal cybersecurity requirements for products. **Cyber Resilience Act**

Open by design and by default

Directive (EU) 2019/1024 on open data and the re-use of public sector information, Article 5, Paragraph 2: *Member States shall encourage public sector bodies and public undertakings to produce and make available documents in accordance with the principle of "open by design and by default". Link here*



Lafortune G., Ubaldi B. (2017). OECD 2017 OURdata Index: Methodology and Results. https://doi.org/10.1787/2807d3c8-en

Definitions from the EU legislation

Research Data
Documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process or are commonly accepted in the research community as necessary to validate research findings and results.

Dynamic data
Documents in a digital form, subject to frequent or real-time updates, in particular because of their volatility or rapid obsolescence; data generated by sensors are typically considered to be dynamic data

Standard licence
A set of predefined re-use conditions in a digital format, preferably compatible with standardised public licences available online.

Definitions from the EU legislation

Research data can be defined as the evidence that underpins the answer to research questions and can be used to validate findings regardless of their form. These might be quantitative information or qualitative statements collected by researchers in the course of their work by esperimentation, observations modelling, interview, or other methods, or information derived from existing evidence.
UK Reasearch and Innovation, 2016

Definitions from the EU legislation

access
access rights
anonymisation
application programming interface (API)
authorised representative
bulk download
cloud computing service
common specifications
competent authority
critical infrastructure
cross-border interoperability
cross-border processing
data
data altruism
data holder
data intermediation service
data processing service
data sharing
data user
dissemination to the public
distributor
dynamic data
end user
exploitation
exportable data
formal open standard
granularity
harmonised standard
high-value datasets
identification service
infrastructure for spatial information
input data
intended purpose
intermediary service
interoperability
legal entity
legal representative
machine-readable format
metadata
non-personal data
content-sharing service provider
online platform
online search engine
open access
open format
open interoperability specification
open science
open-source licence
personal data breach
processing
product data
professional user
profiling
provider
public authority
readily available data
real-world testing plan
related service data
research data
research infrastructures
research organisation
research output
re-use
risk
search results
secure processing environment
service provider
spatial data
spatial data services
spatial data set
standard licence
terms and conditions
text and data mining
training data
validation data
validation data set

Locati M. (2024). Definitions of terms extracted from data-related European Union laws. https://doi.org/10.5281/zenodo.12791247

Definitions from the EU legislation

Metadata
Data about data. Information about data such as its title and description, method of collection, author or publisher, area and time-period covered, licence, date and frequency of release, funding projects, closely related data and publications, etc.

Dataset
Any organised collection of data. ‘Dataset’ is a flexible term and may refer to an entire database, a spreadsheet or other types of data file, or a related collection of data resources.

Database
1) An organised collection of data may be considered a database. In this sense, is synonymous with dataset.
2) A software for processing and managing data, including features to extend or update, transform and query the data (e.g. the open-source MySQL, PostgreSQL, or the proprietary Microsoft Access).

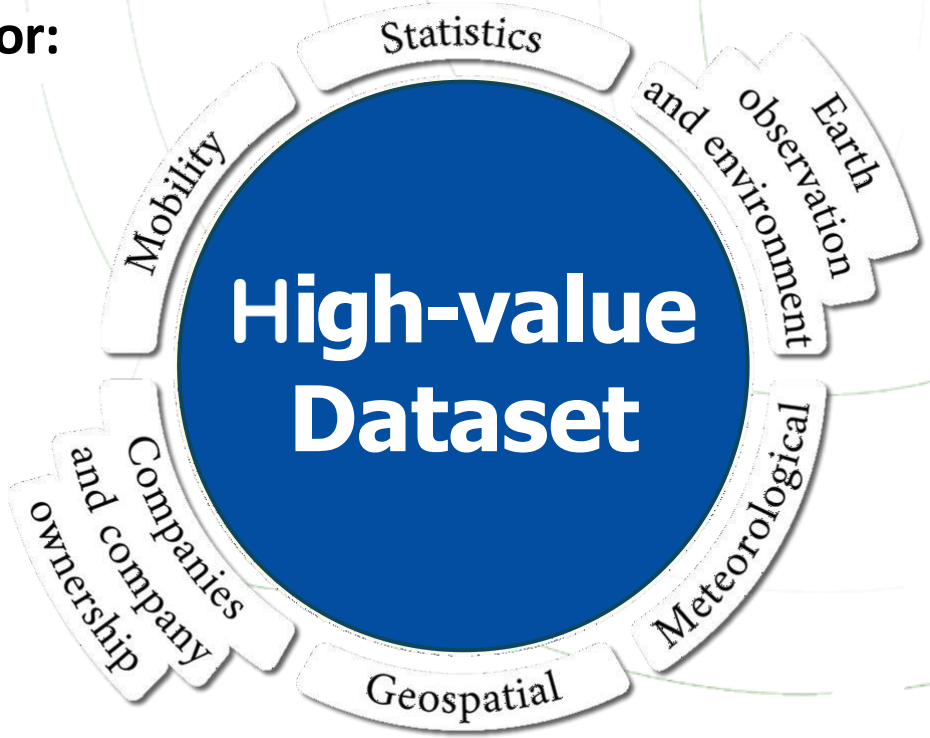
Open format
A file format that is platform-independent (not tied to a specific software), well documented, and made available to the public without any restriction that impedes the full re-use of documents.

Data that must not, should be, and must be shared

As open as possible,

as closed as necessary

Sharing compulsory for:



Re-usable dataset that may generate important benefits for society, the environment and the economy, in particular because of their suitability for the creation of value-added services, applications and new, high-quality and decent jobs for many potential beneficiaries.

Sharing limitations apply to:

- Personal data
- Data related to public security, defence and national security

Sharing recommendation for:

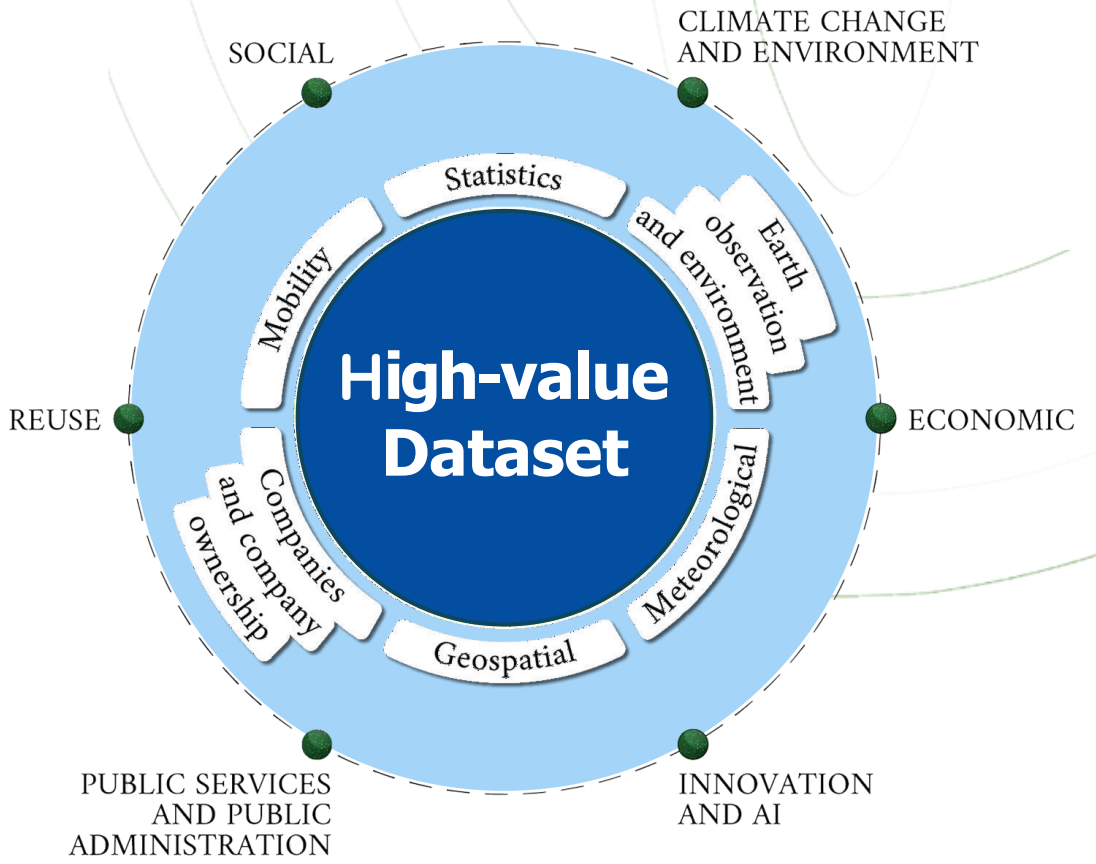
- Data generated using public fundings
- No re-use limitations, even for commercial purposes

Data that must not, should be, and must be shared

As open as possible,

as closed as necessary

Sharing compulsory for:



Sharing limitations apply to:

- Personal data
- Data related to public security, defence and national security

Sharing recommendation for:

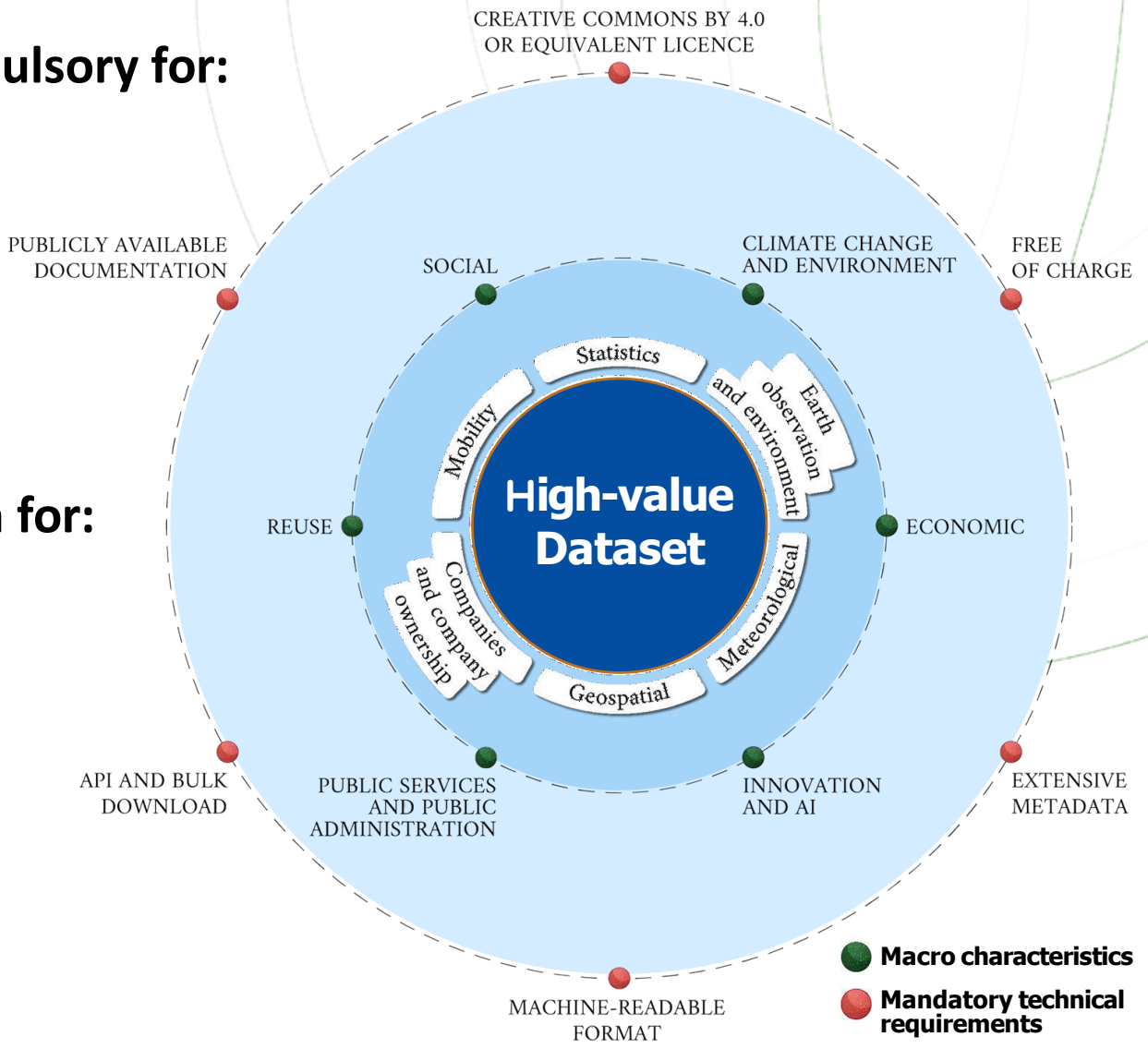
- Data generated using public fundings
- No re-use limitations, even for commercial purposes

Data that must not, should be, and must be shared

As open as possible,

as closed as necessary

Sharing compulsory for:



Sharing limitations apply to:

- Personal data
- Data related to public security, defence and national security

Sharing recommendation for:

- Data generated using public fundings
- No re-use limitations, even for commercial purposes

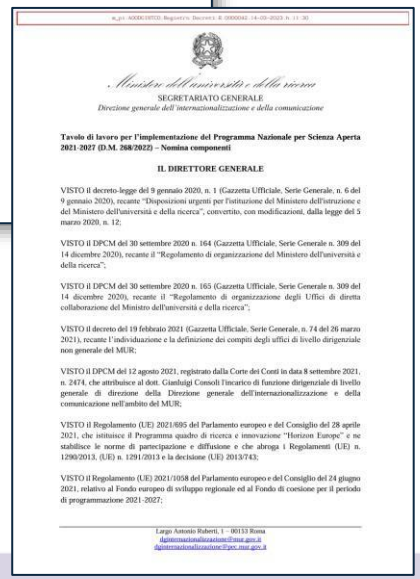
Italian plan

Italian plan for Open Science 2021-2027

Link to PDF

The Ministry of University and Research (MUR) selected 5 areas of intervention for the period 2021-2027:

1. Scientific publications
2. Scientific research data
3. Research evaluation
4. Open science, the community and European participation
5. Open data from research on SARS-CoV-2 and COVID-19



Working group for the implemer
National Programme for Open S

- Donatella Castelli
- Emma Lazzeri
- Giorgio Rossi
- Francesca Di Donato
- Roberto Cimino

Link to PDF

Italiadomani
PIANO NAZIONALE
DI RIPRESA E RESILIENZA

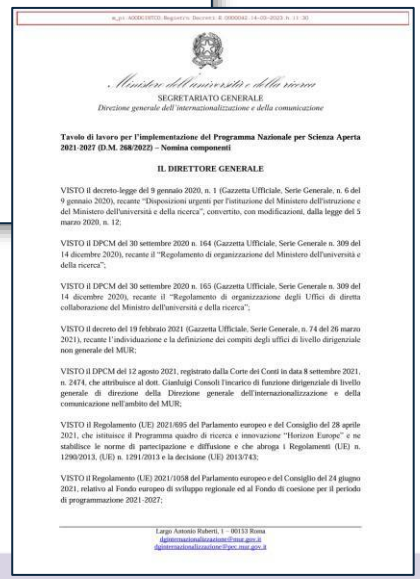
[...] strengthening rapid exchange of data between researchers, including access to results and evidence at no charge to the user and Open Science [...]

Italian plan for Open Science 2021-2027

Link to PDF

The Ministry of University and Research (MUR) selected 5 areas of intervention for the period 2021-2027:

1. Scientific publications
2. Scientific research data
3. Research evaluation
4. Open science, the co
5. Open data from rese


Ministero dell'università e della ricerca
SEGRETARIATO GENERALE
Direzione generale dell'internazionalizzazione e della comunicazione

Tavolo di lavoro per l'implementazione del Programma Nazionale per Scienza Aperta 2021-2027 (D.M. 268/2022) – Nomina componenti

IL DIRETTORE GENERALE

VISTO il decreto-legge del 9 gennaio 2020, n. 1 (Gazzetta Ufficiale, Serie Generale, n. 6 del 9 gennaio 2020), recante "Disposizioni urgenti per l'istituzione del Ministero dell'istruzione e del Ministero dell'università e della ricerca", convertito, con modificazioni, dalla legge del 5 marzo 2020, n. 12;

VISTO il DPCM del 30 settembre 2020 n. 164 (Gazzetta Ufficiale, Serie Generale n. 309 del 14 dicembre 2020), recante il "Regolamento di organizzazione del Ministero dell'università e della ricerca";

VISTO il DPCM del 30 settembre 2020 n. 165 (Gazzetta Ufficiale, Serie Generale n. 309 del 14 dicembre 2020), recante il "Regolamento di organizzazione degli Uffici di diretta collaborazione del Ministro dell'università e della ricerca";

VISTO il decreto del 19 febbraio 2021 (Gazzetta Ufficiale, Serie Generale, n. 74 del 26 marzo 2021), recante l'individuazione e la definizione dei compiti degli uffici di livello dirigenziale non generale del MUR;

VISTO il DPCM del 12 agosto 2021, registrato dalla Corte dei Conti in data 8 settembre 2021, n. 2474, che attribuisce al dott. Gianluigi Consoli l'incarico di funzione dirigenziale di livello generale di direzione della Direzione generale dell'internazionalizzazione e della comunicazione nell'ambito del MUR;

VISTO il Regolamento (UE) 2021/695 del Parlamento europeo e del Consiglio del 28 aprile 2021, che istituisce il Programma quadro di ricerca e innovazione "Horizon Europe" e ne stabilisce le norme di partecipazione e diffusione e che abroga i Regolamenti (UE) n. 1290/2013, (UE) n. 1291/2013 e la decisione (UE) 2013/743;

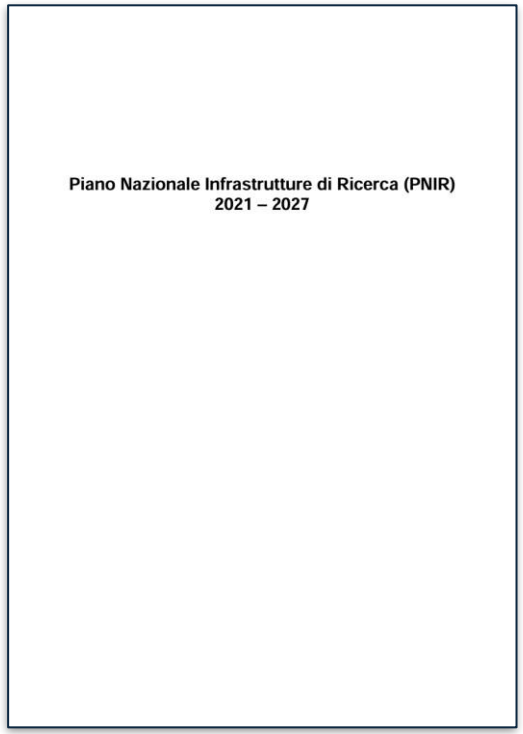
VISTO il Regolamento (UE) 2021/1058 del Parlamento europeo e del Consiglio del 24 giugno 2021, relativo al Fondo europeo di sviluppo regionale ed al Fondo di coesione per il periodo di programmazione 2021-2027;

Largo Antonio Ruberti, 1 – 00153 Roma

Working group for
National Programr

- Donatella Caste
- Emma Lazzeri
- Giorgio Rossi
- Francesca Di Dc
- Roberto Cimino

Link to PDF


Piano Nazionale Infrastrutture di Ricerca (PNIR)
2021 – 2027

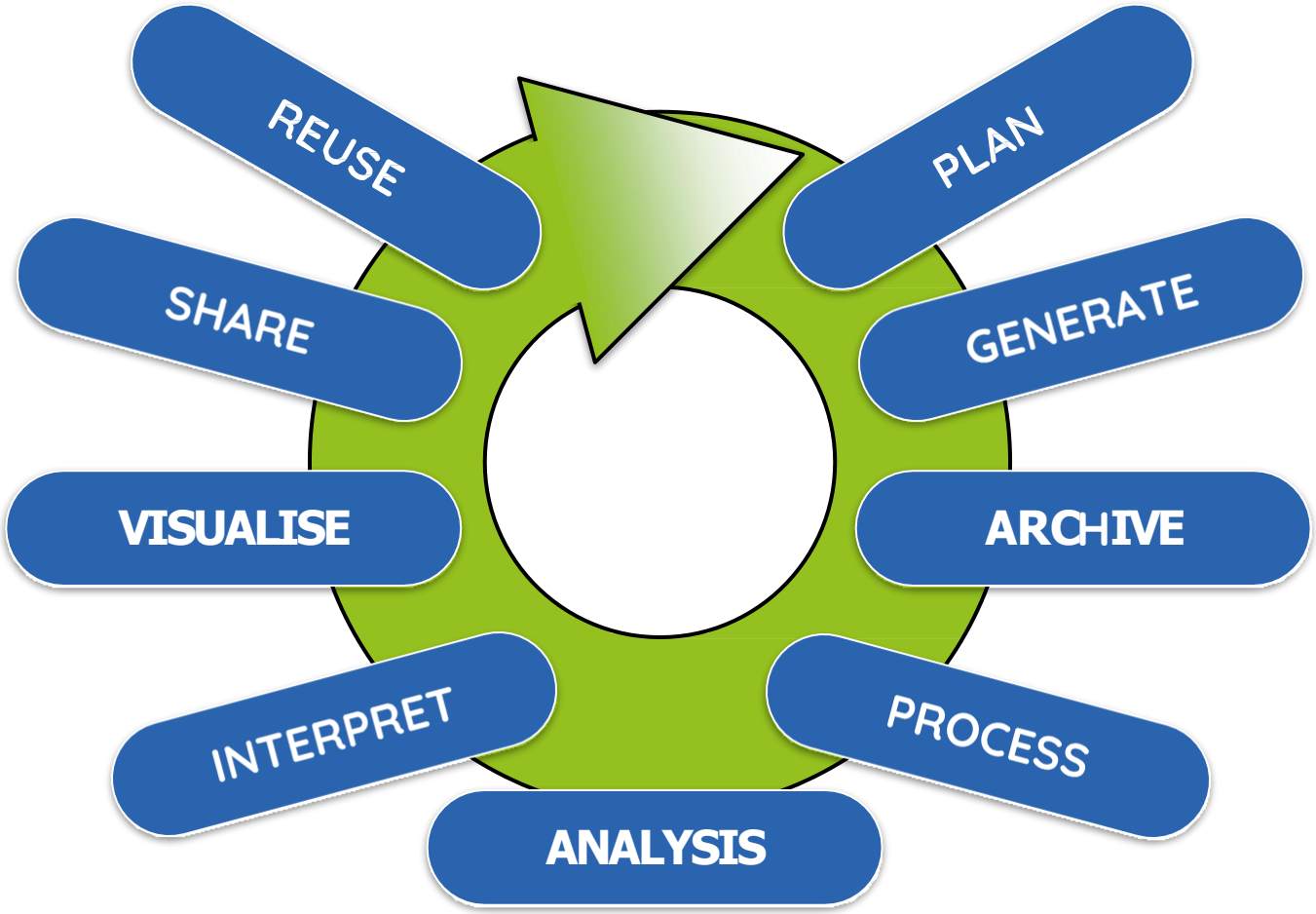
National Plan for Research Infrastructure

- National and European context
- Strategic guidelines
- National priorities

Link to PDF

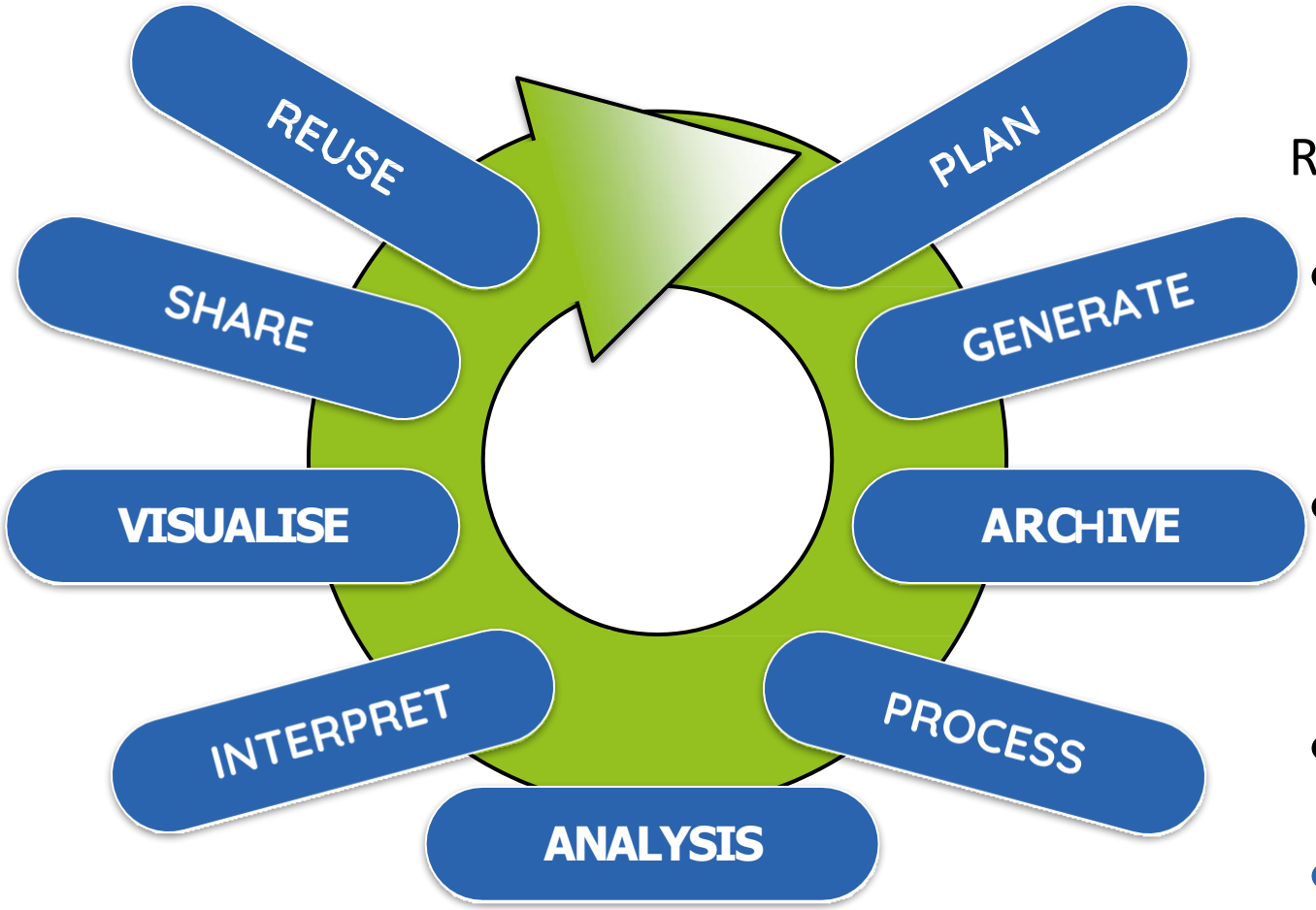
Research Data Management

Research Data Life Cycle



Cartoon by Nuthawut Somsuk at https://www.vecteezy.com/

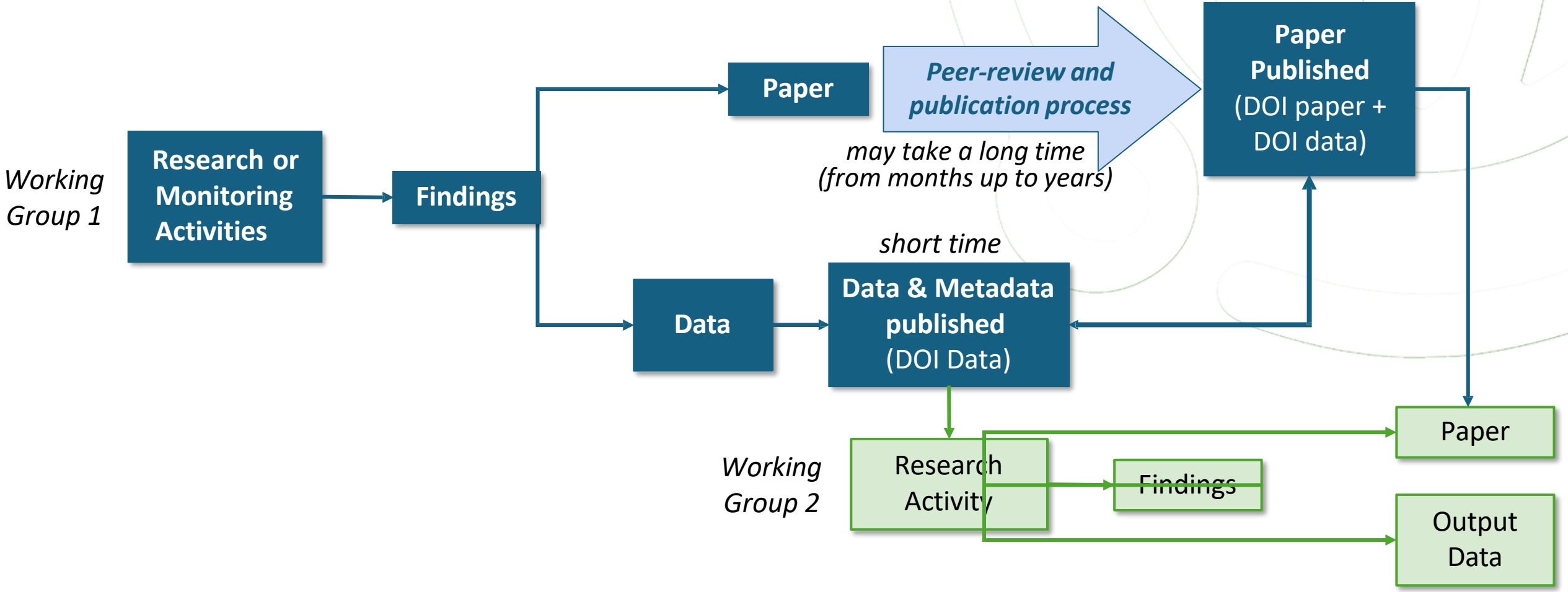
Research Data Life Cycle



Research Data Management involves:

- **Scientific Community** for data generation and curation (**Data Providers** and **Data Curators**)
- **Technological Community** for making data Interoperable (**IT** and **Data Stewards**)
- Governance and DMP (**Data Managers**)
- **Financial Managers**

A differentiated path depending on the type of output



Data classification, it depends on the point-of-view

- Disciplinary specific **vs** Multi-disciplinary data
- Primary (raw) data **vs** Secondary (refined) data
- Data from natural phenomena **vs** Data from logical models
- Reproducible **vs** Not reproducible
- Static **vs** Dynamic
- Big data **vs** Small data
- Personal **vs** Anonymous
- Qualitative **vs** Quantitative
- Structured **vs** Unstructured
- Fully owned **vs** Owned by others
- Openly available **vs** Restricted
- Public domain **vs** Copyrighted
- Open formats **vs** Closed formats
- Relational **vs** Hierarchical

Data classification, it depends on the point-of-view

- Disciplinary specific **vs** Multi-disciplinary data
- **Primary (raw) data vs Secondary (refined) data**
- Data from natural phenomena **vs** Data from logical models
- Reproducible **vs** Not reproducible
- Static **vs** Dynamic
- Big data **vs** Small data
- Personal **vs** Anonymous
- Qualitative **vs** Quantitative
- Structured **vs** Unstructured
- Fully owned **vs** Owned by others
- Openly available **vs** Restricted
- Public domain **vs** Copyrighted
- Open formats **vs** Closed formats
- Relational **vs** Hierarchical

Data Processing Levels (DPL)

0: raw data, or basic data

Represent the output of the sensors or actuators used to measure the physical properties or phenomena.

1: data from nearly automated procedures

applied to raw data at level 0.

2: data from scientists' investigations,

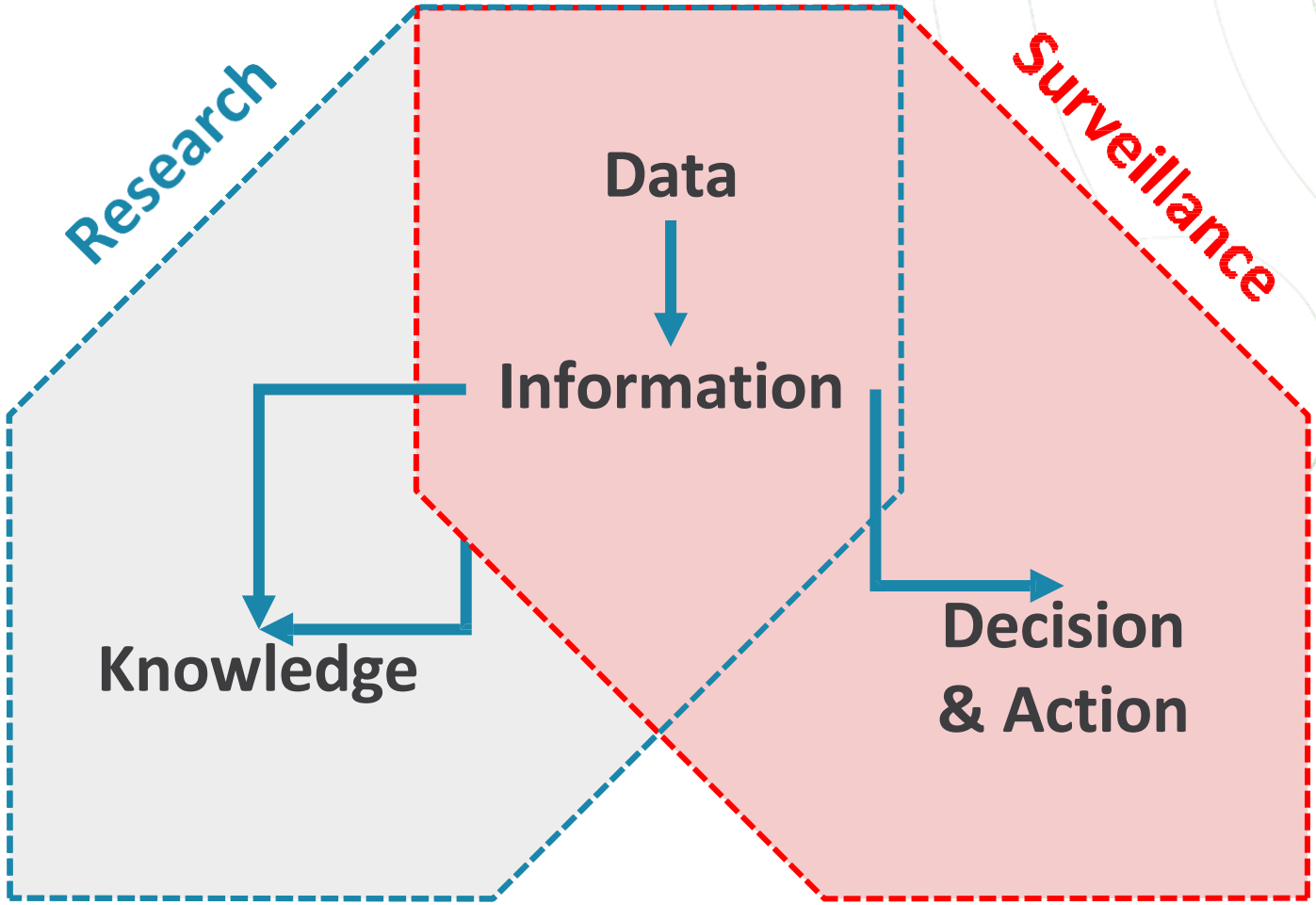
it requires scientists that interpret and add values to Level 0 or Level 1 data.

3: integrated data

complex analyses or community shared products, and collaborative processes in a considerable time span.

Bailo et al. (2023) https://doi.org/10.1038/s41597-023-02697-9

Data for Research and for Surveillance



Adapted from Chiolero & Buckeridge 2020
https://doi.org/10.1136/jech-2018-211654

Data management

Research Data Management (RDM)

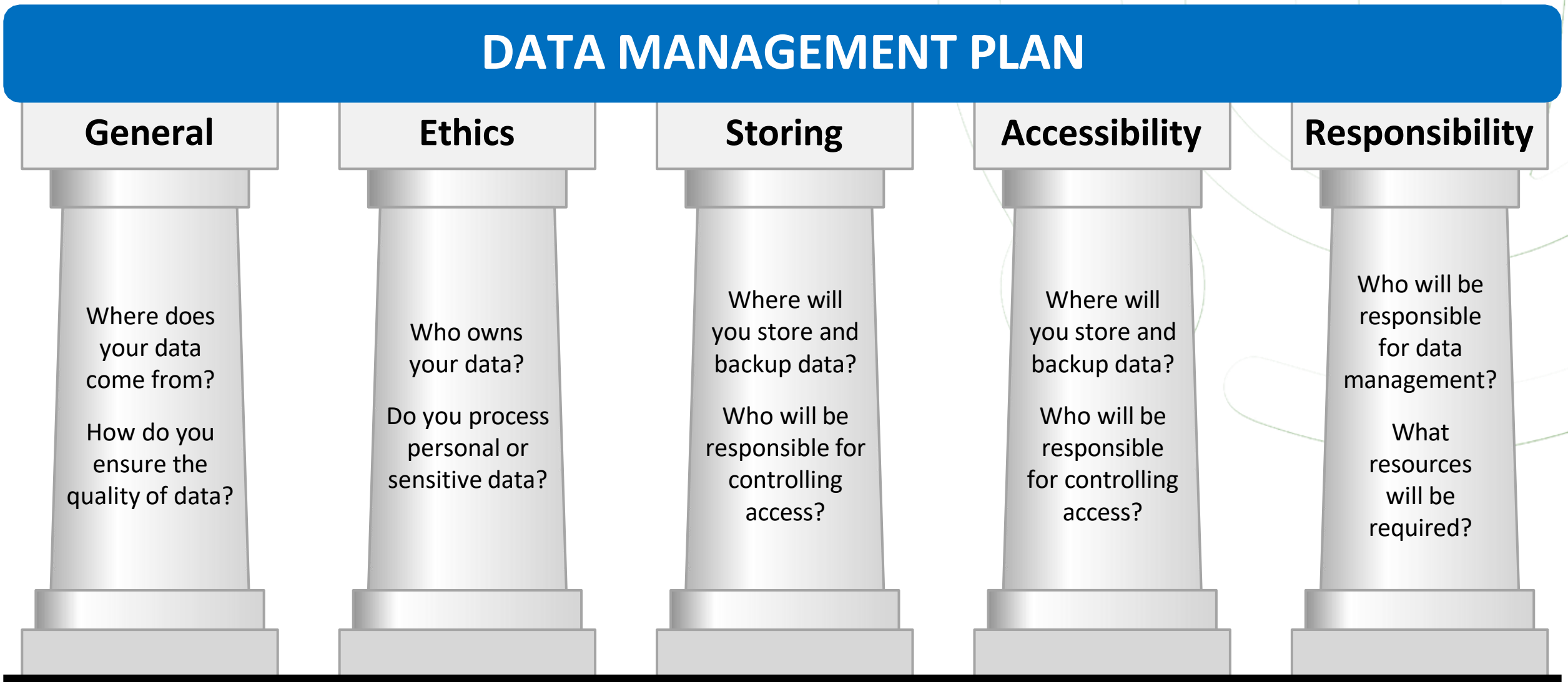
Organization, storage, preservation, and sharing of data in a research environment throughout the **entire data life cycle and beyond**.
RDM ensures the efficiency, transparency, and reproducibility of research and include **practices and policies that aim to improve the quality and value of research data**.

Data Management Plan (DMP)

A formal document that outlines **how data is handled during and after a research project**.
Details how data is collected, analysed, preserved, and shared, including details about data formats, metadata standards, data storage, and data sharing policies.
It is **often a required component of grant proposals**, emphasizing the importance of responsible data management for funders and institutions.

Cartoon by Nuthawut Somsuk at https://www.vecteezy.com/

Basic components of a DMP



Scheme modified from "Basics of Research Data Management", University of Eastern Finland, https://sites.uef.fi/rdm/data-management-plan-dmp/

Map your stakeholders and their needs

Scientific Community

Public Bodies

Professionals

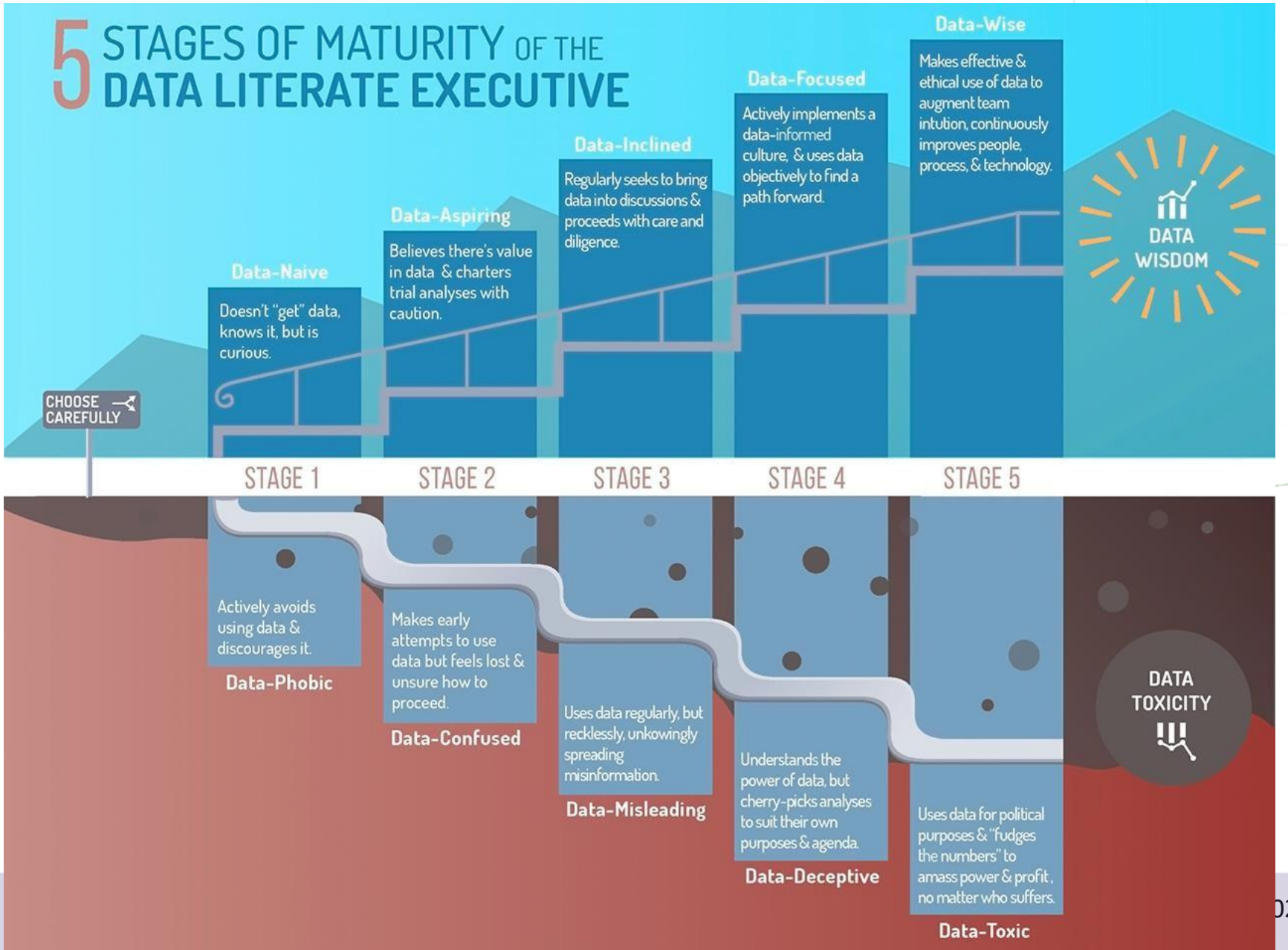
Companies

Media

General Public

Cartoon by Nuthawut Somsuk at https://www.vecteezy.com/

Know your Data Literacy Maturity Level



Data Literacy, 2023
https://dataliteracy.com/the-data-leadership-maturity-model/

Institutional Data Policy

Any organisation should set up a policy adapting Data Regulations, the Open Science paradigm and FAIR data principles to its own environment and practices, providing its employees with:

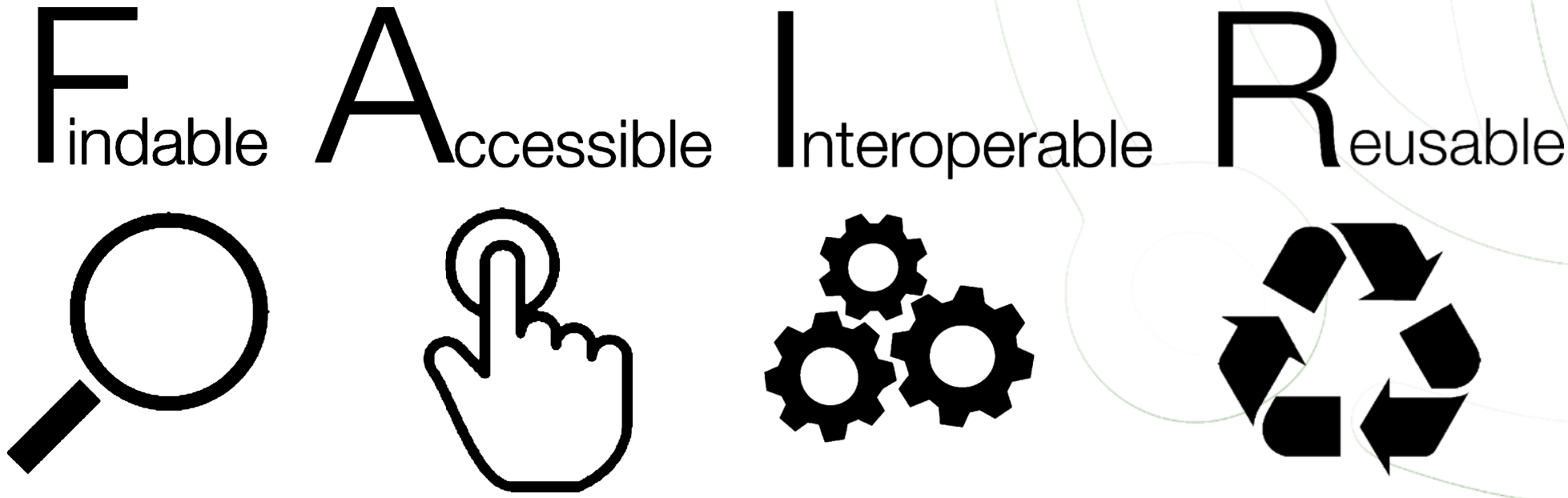
- A **translation layer** to the complex external legal framework related to data
- A clear framework facilitating the adoption of **Research Data Management**
- **Guidelines for peculiarities** not covered by any external framework
- Identify **roles, and responsibilities**, providing internal and external contact points
- May establish a **Data Registry**, also called “metadata catalogue”, a list describing all data shared
- **Practical rules** on how to share data internally and externally
- Predefined and **ready-to-use solutions** covering
  - Accountability
  - Provenance
  - Licensing
  - Preservation

Cartoon by Nuthawut Somsuk at https://www.vecteezy.com/

FAIR data principles

FAIR data principles

Wilkinson et al., 2016



Findable: easy to find by humans and computers thanks to metadata and unique persistent identifiers

Accessible: stored for easy access and downloading

Interoperable: ready to be combined with other datasets by humans and computer systems

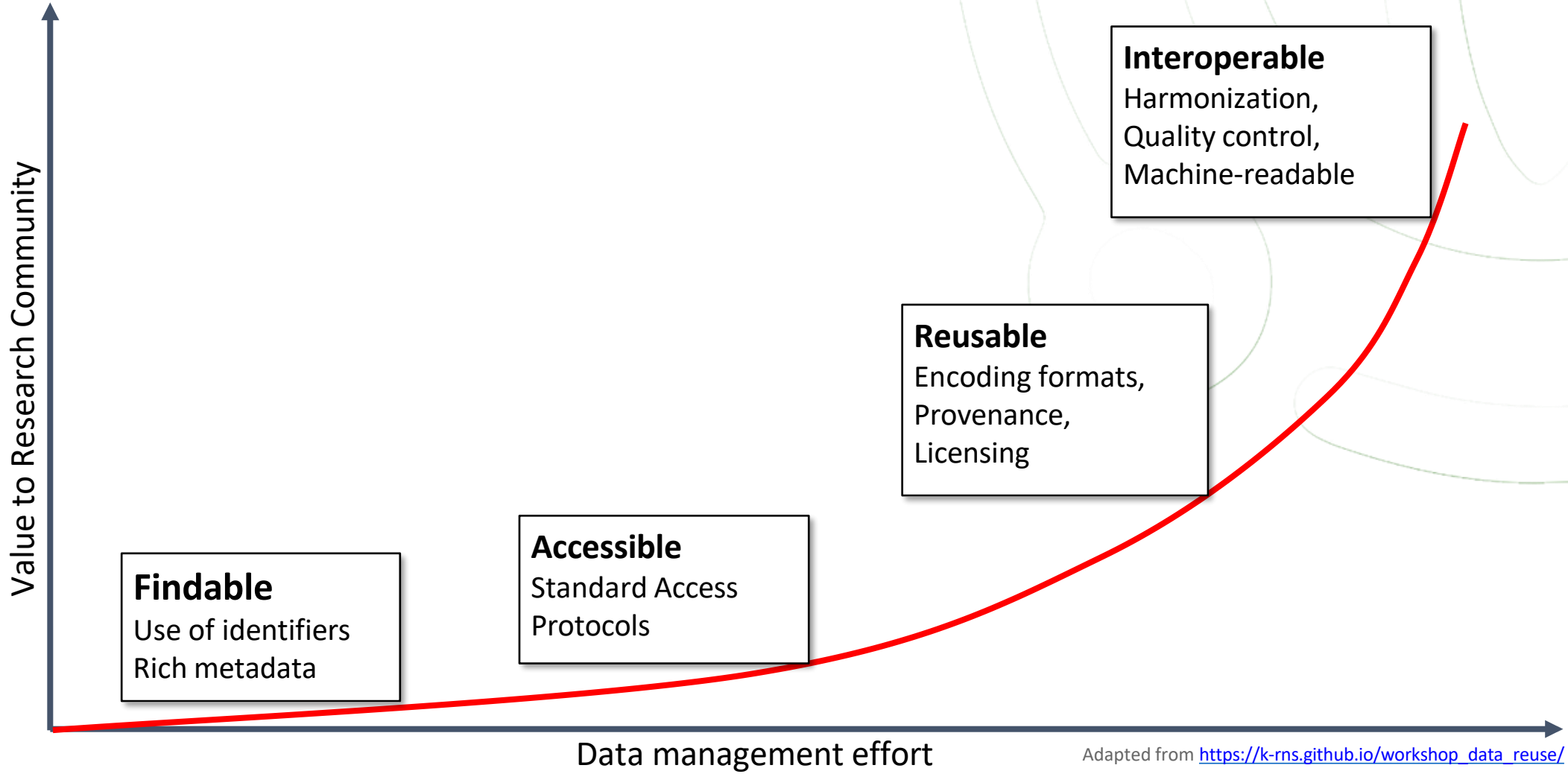
Re-usable: ready for reuse thanks to detailed, accurate documentation and clear usage license

Graphical representation by Door SangyaPundir - Eigen werk, https://commons.wikimedia.org/w/index.php?curid=53414062

Advantages of FAIR data

- **Future-proof**, because data is well documented and is provided using open data formats
- Improve the **transparency**, **reliability** and **reproducibility** of research
- Increase the **visibility**, more people are aware that your data exists
- Facilitate data sharing and **collaborations**
- Prevent **data loss**
- Maximise **potential** from data assets, increase re-use of data
- Maximise research **impact**, return-on-investment for public funding
- Adhere to **Open Science** paradigm, more efficiency in accessing data

Effort to implement FAIR data



Adapted from https://k-rns.github.io/workshop_data_reuse/

Factors weakening the FAIRness of data

Findable. Not easy to find **documentation**, no permission to have **access**, no understandable **licenses**, no appropriate **metadata**.

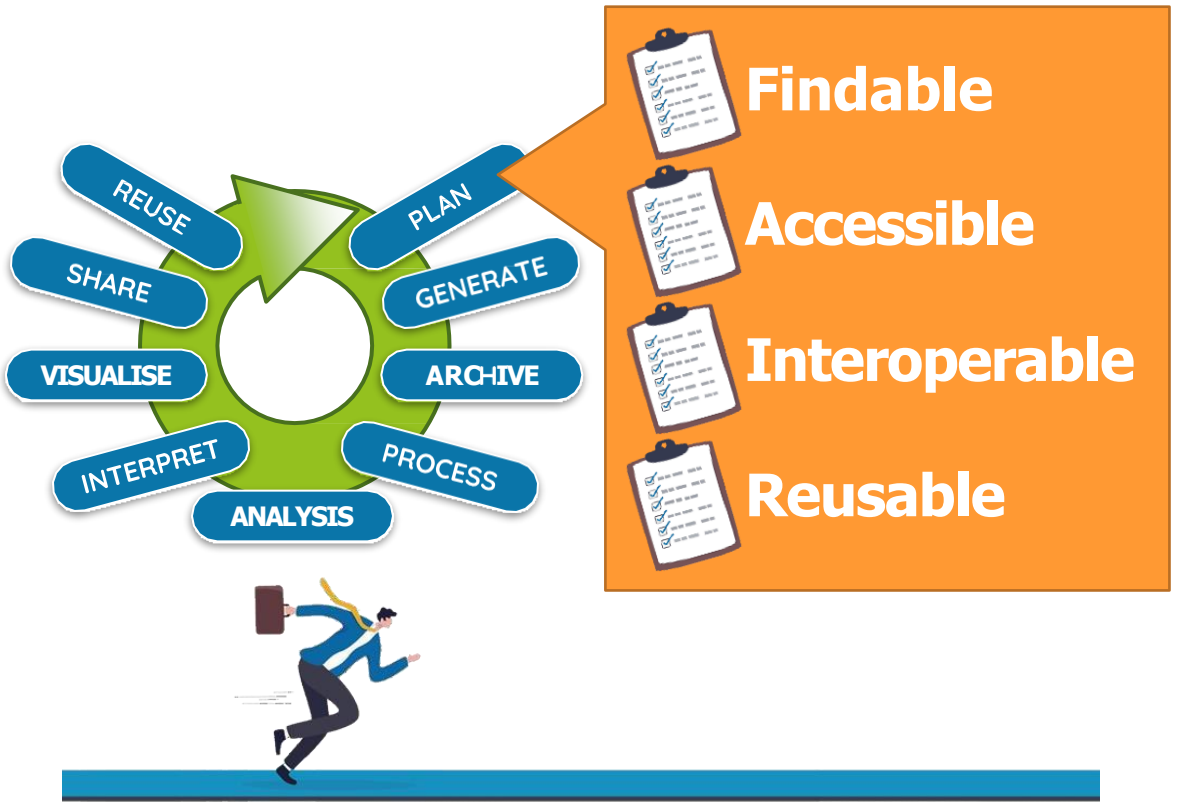
Accessible. Lack of clear **rules** to make datasets/data available, no **identification** of providers/users for access, no open **licenses**, not easy to find **vocabularies** or **ontologies**.

Interoperable. Lack of **documentation**, abuse of **IPRs** or **copyright restrictions**, factors preventing **machine actionable** (access), **readable** (encoding) and **interpretable** data (use).

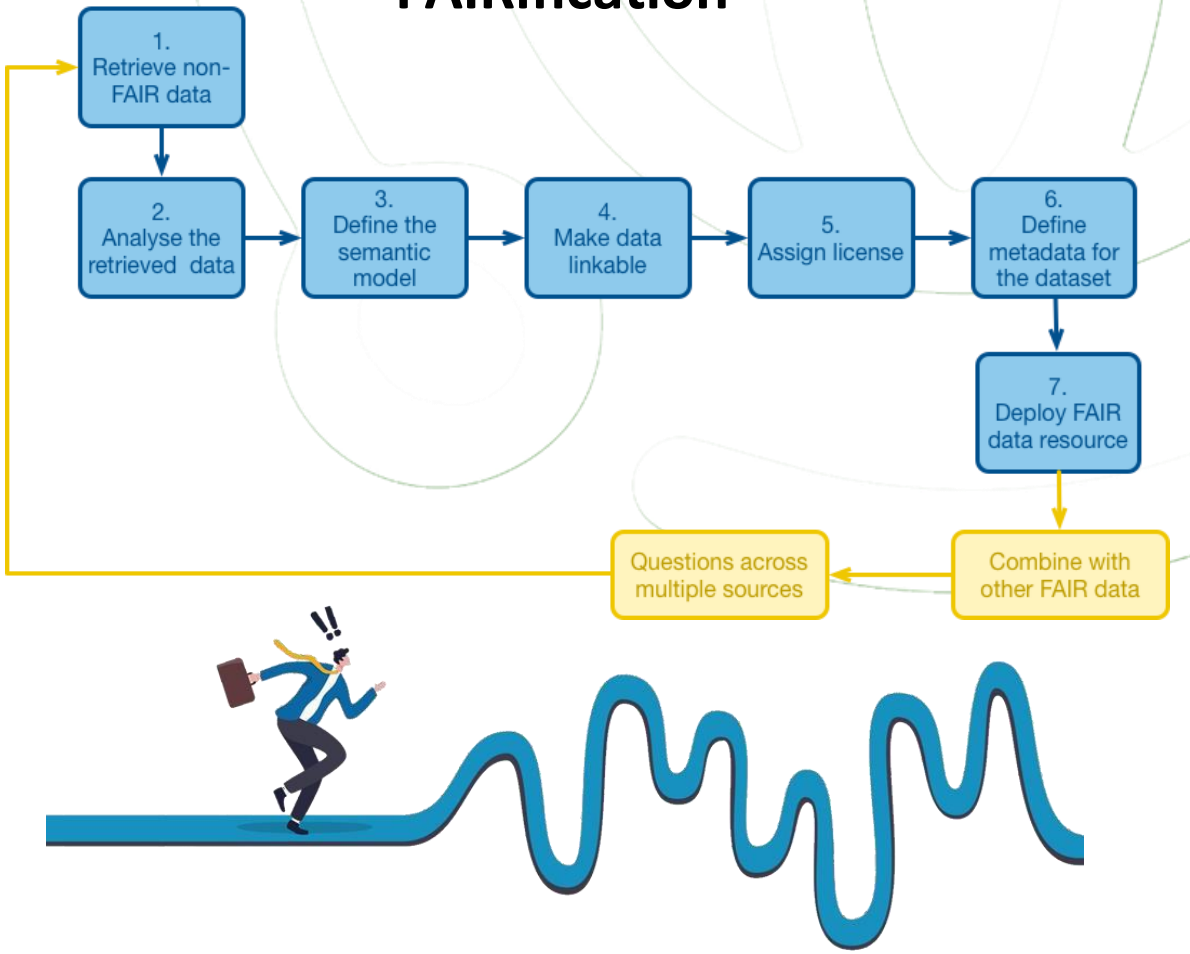
Re-usable. Exclusive arrangements for documentation, no secure environment for access, no appropriate **licenses** for re-use, limited and vague semantic.

FAIR by design vs FAIRification

FAIR-by-design



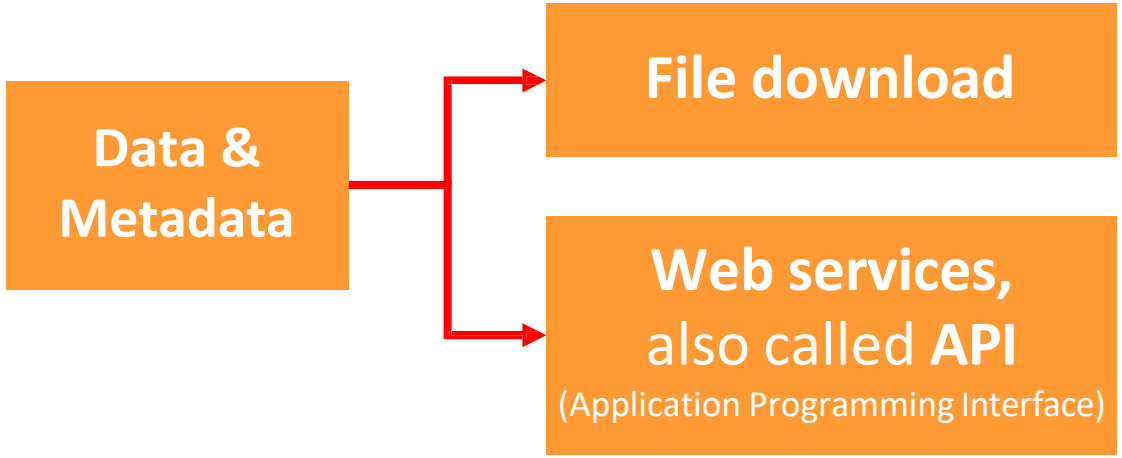
FAIRification



Cartoon by Nuthawut Somsuk at https://www.vecteezy.com/

Data access and data encoding

Two ways to access Data & Metadata:



Web services accept parameters for filtering at the source the downloaded data, providing users only with the data they need.

Spatial data can be provided via web services compliant to the standard **OGC (Open Geospatial Consortium)**

Data can be encoded using multiple formats

A **table** with a dataset can be encoded using:
CSV, XLSX, XML, JSON, ODS, RDF, HDF5

An **image** can be encoded using:
PNG, TIFF, TGA, JPG, GIF, PDF

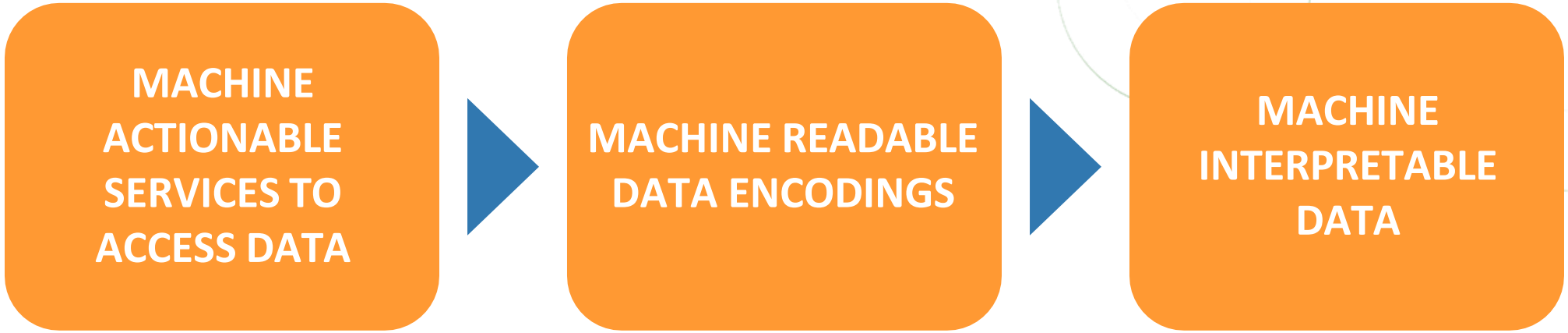
The encoding format must be selected carefully, depending on a variety of factors (i.e. potential use, storing capabilities), some remarks:

- some format adopt a data compression that may alter the original data (e.g. images compressed using JPG)
- some format may store metadata together with data (e.g. RDF, XML, JSON, HDF5)

Open Science: efficient access and use of data

Adopting the **Open Science** paradigm should increase the **efficiency** in accessing and using Research outputs, taking advantage of modern **communication and analysis technologies**.

The **less manual actions** are required, the more efficient the overall system is.



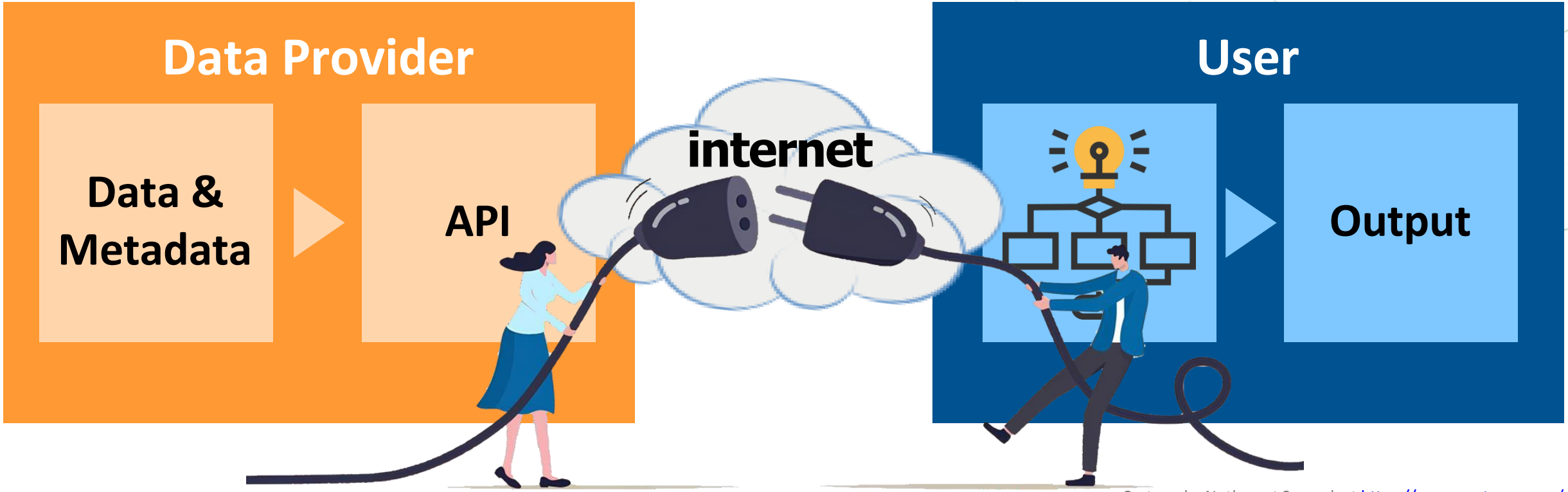
The goal can be achieved adopting the **FAIR data principles**.

Adapted from https://k-rns.github.io/workshop_data_reuse/

Web services simplify the end user's workflow

Reliable machine actionable web services allows users to:

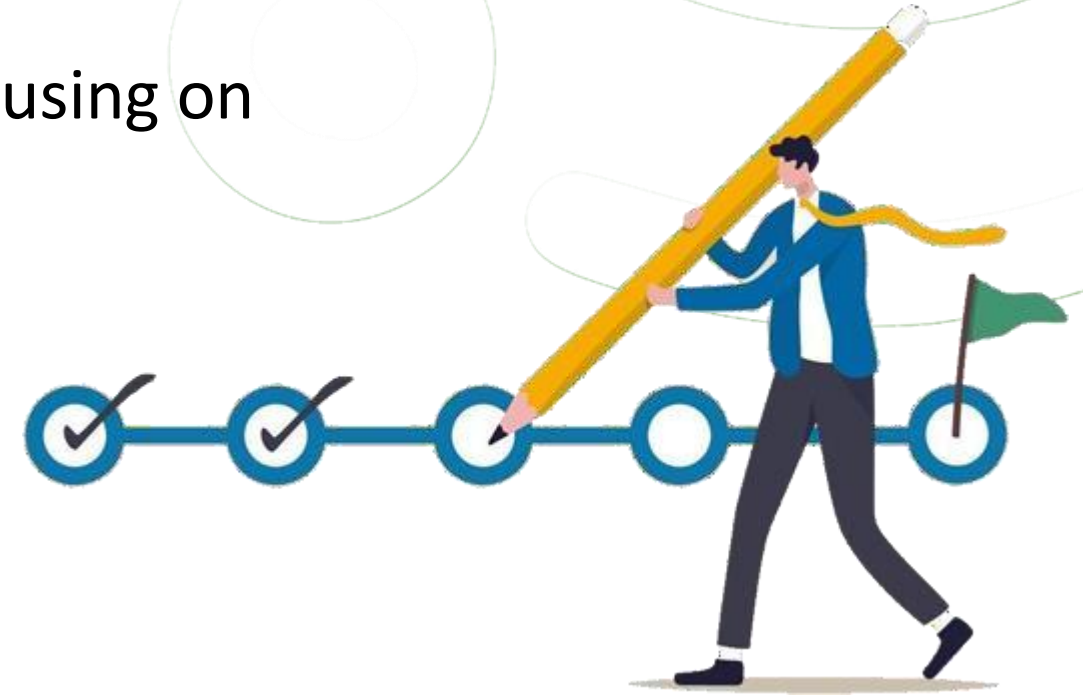
- no need to build another data infrastructure
- ensure output based on the latest data



Cartoon by Nuthawut Somsuk at https://www.vecteezy.com/

Traceability of provenance

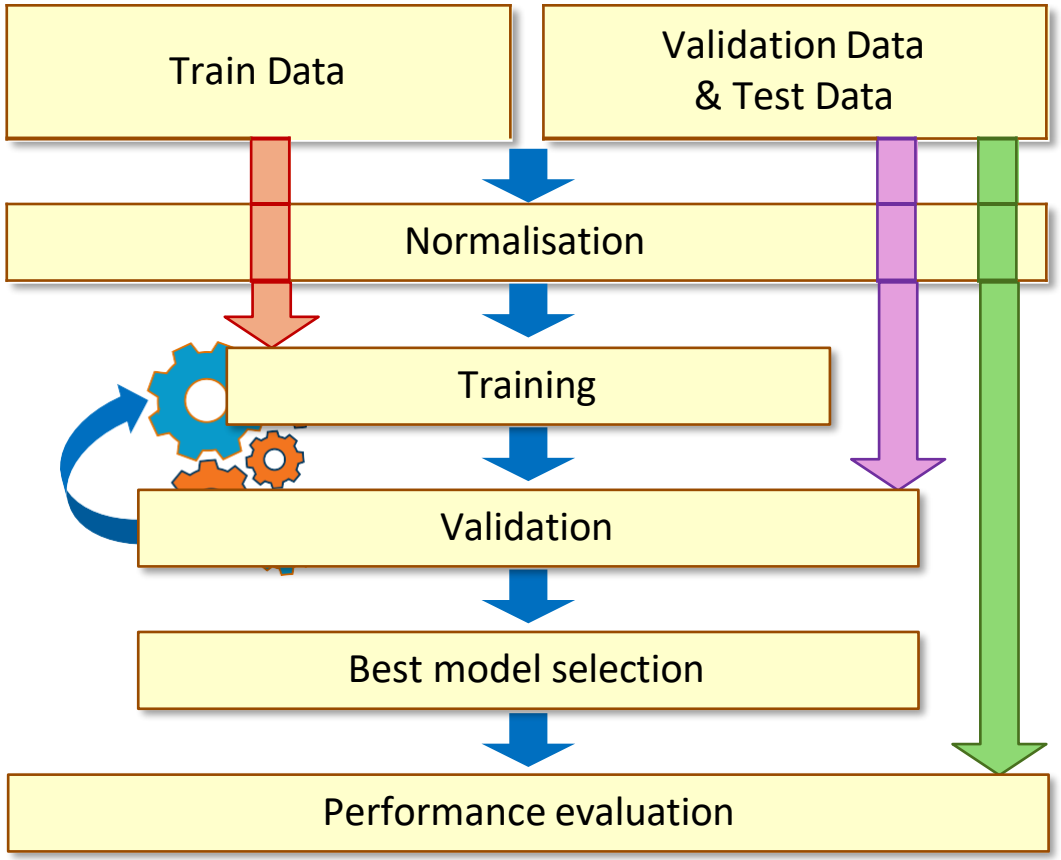
- Track every step of the data life cycle
- Heavily relies on metadata
- Easy to implement while generating data
- Hard once data is published
- Increase the perceived trust and quality focusing on
  - Reliability
  - Transparency of procedures
- Facilitate
  - Replicability
  - Identification of errors



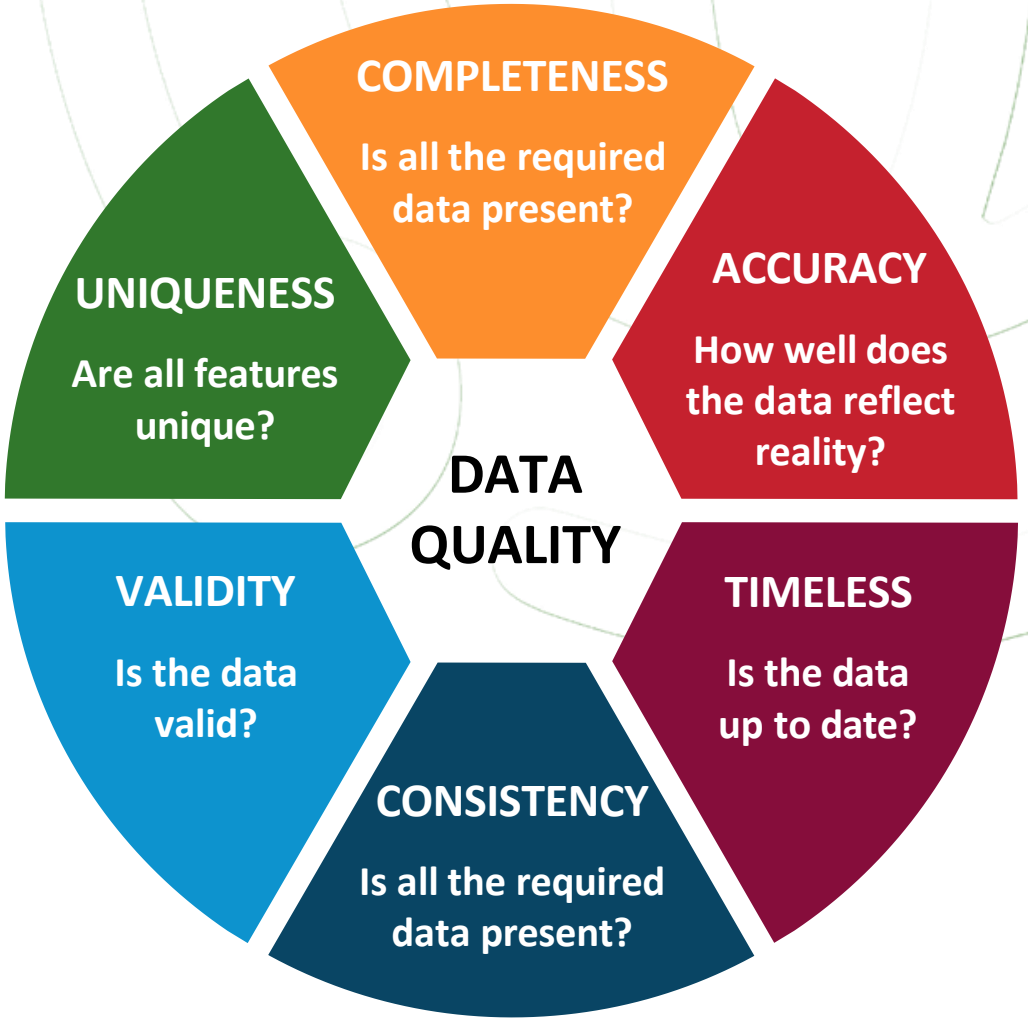
Cartoon by Nuthawut Somsuk at https://www.vecteezy.com/

Data Quality, a key to improve Machine Learning

Reduce the garbage in, garbage out
for a better training of Machine Learning algorithms



Andy McDonald, Data Quality Considerations for Machine Learning Models, https://towardsdatascience.com/data-quality-considerations-for-machine-learning-models-dcbe9cab34cb



Source: adapted from Karapanagiotis & Krebber 2023, https://doi.org/10.3390/s23136187

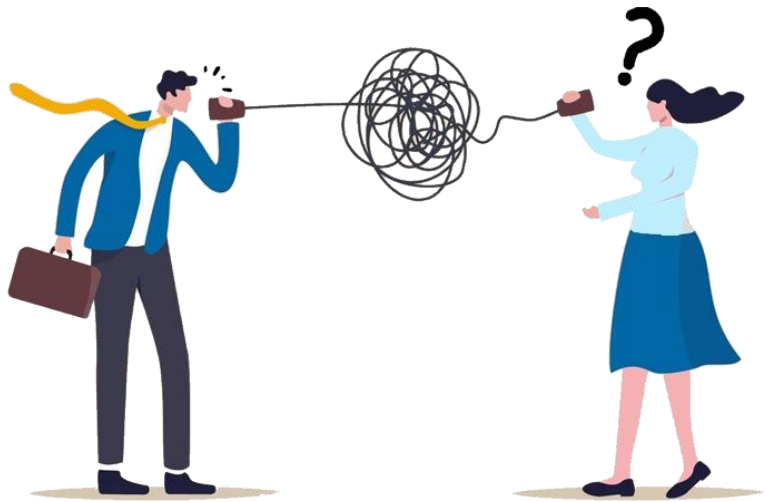
Controlled vocabularies and ontologies

Controlled Vocabularies

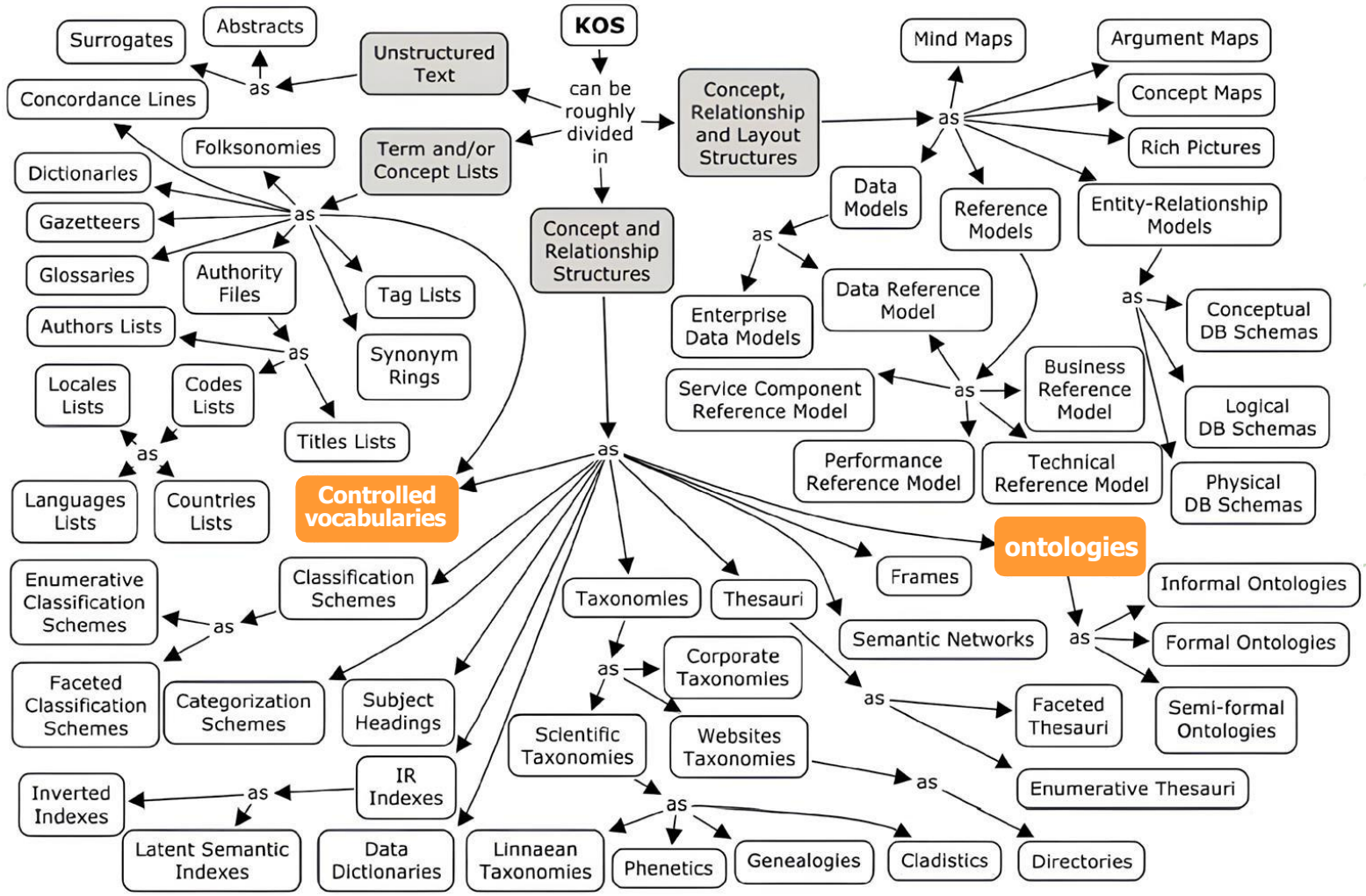
are used to ensure that data is **described consistently**, terms are defined without a context.

Ontologies

are used to create a **shared understanding** of the data, often in a specific domain of interest, concepts are defined and interlinked.



Cartoon by Nuthawut Somsuk at https://www.vecteezy.com/



Scheme describing the building blocks of a Knowledge Organization Systems (KOS) adapted from Souza, Tudhope & Almeida, 2010

5-star deployment scheme for Open Data

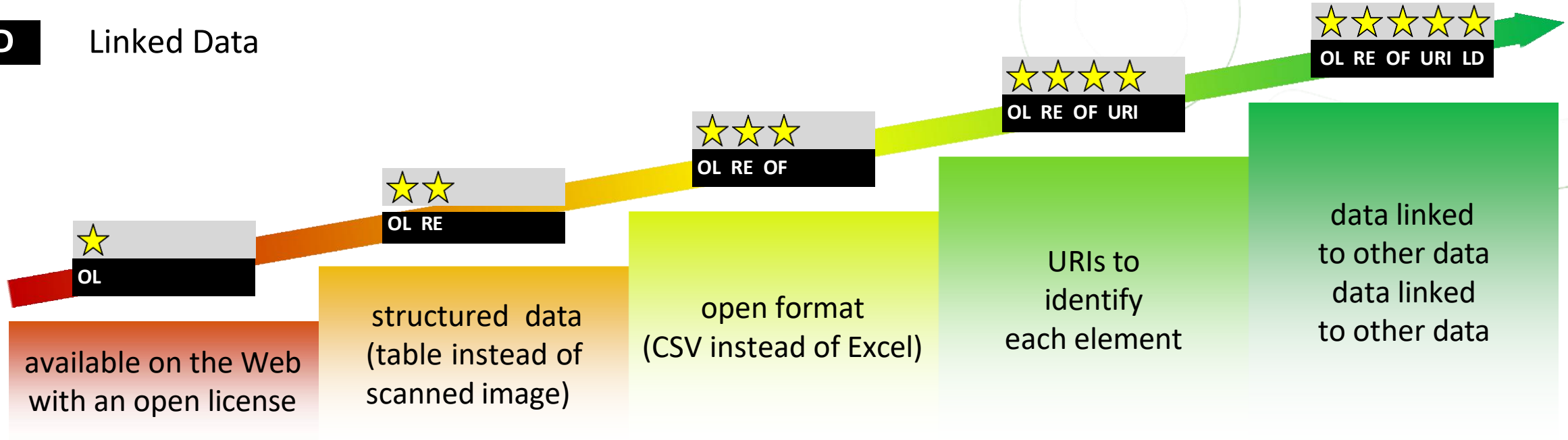
OL Open License

RE Readable

OF Open Format

URI Uniform Resource Identifier

LD Linked Data



Tim Berners-Lee 2010, https://5stardata.info/

Cartoon by Nuthawut Somsuk at https://www.vecteezy.com/

Data preservation, integrity and security

Policy defining where and how data is stored and by whom is managed

Roles and responsibilities for the security of data storage infrastructure

Measure in place for preventing unauthorized access

Frequency of data accessibility checks

Adoption of future-proof data encodings

Frequency of backups

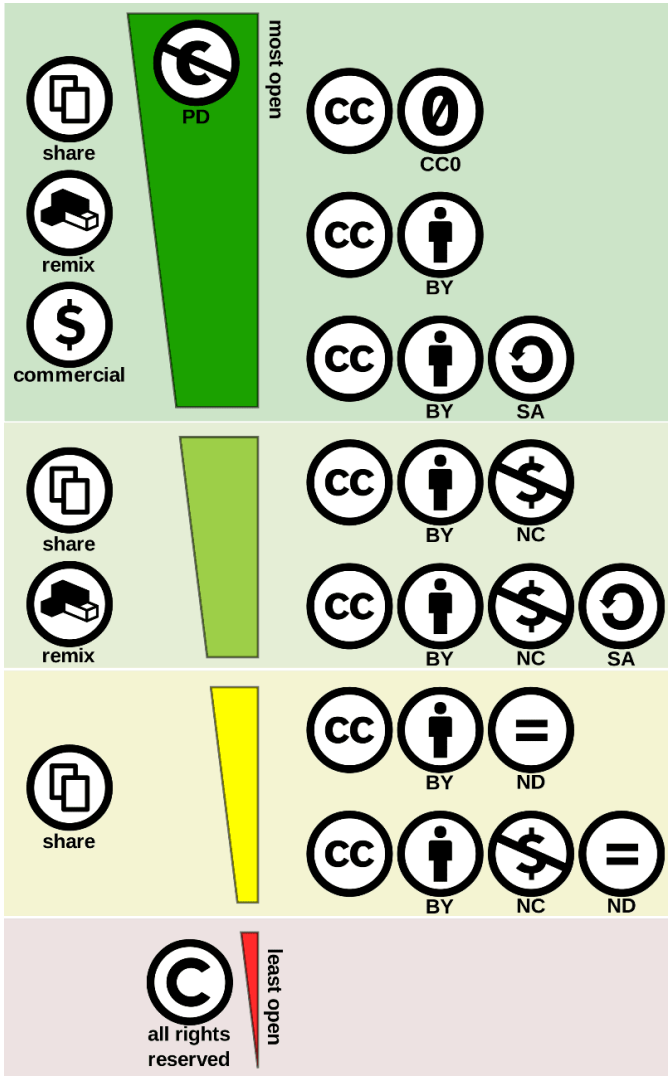
Geographical redundancy

Incident reporting and reaction

Disaster data recovery plan

Cartoon by Nuthawut Somsuk at https://www.vecteezy.com/

Licensing data - The Creative Commons licenses



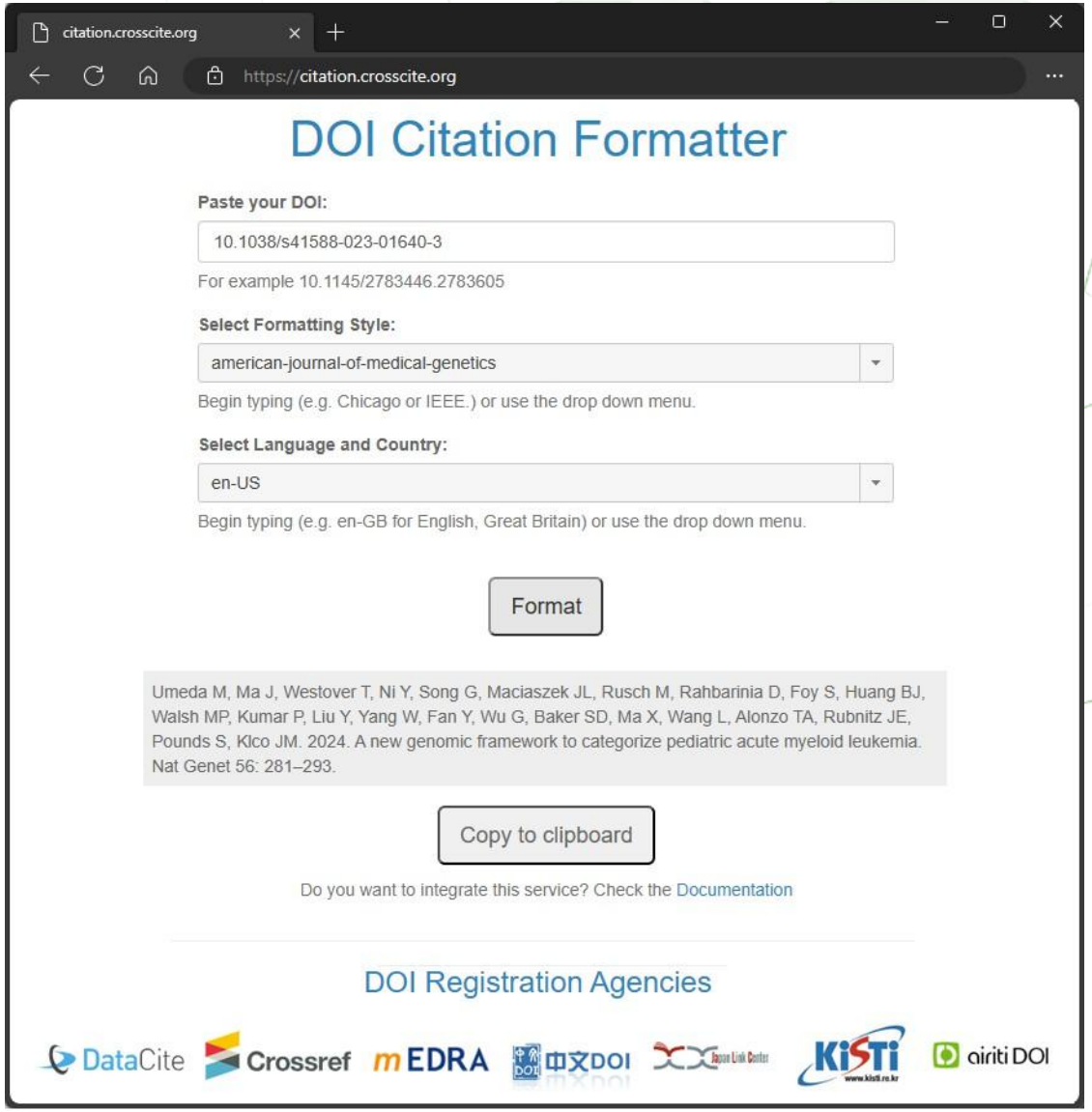
	PUBLIC DOMAIN	PUBLIC DOMAIN	BY	BY SA	BY NC	BY ND	BY NC SA	BY NC ND
PUBLIC DOMAIN	✓	✓	✓	✓	✓	✗	✓	✗
PUBLIC DOMAIN	✓	✓	✓	✓	✓	✗	✓	✗
BY	✓	✓	✓	✓	✓	✗	✓	✗
BY SA	✓	✓	✓	✓	✗	✗	✗	✗
BY NC	✓	✓	✓	✗	✓	✗	✓	✗
BY ND	✗	✗	✗	✗	✗	✗	✗	✗
BY NC SA	✓	✓	✓	✗	✓	✗	✓	✗
BY NC ND	✗	✗	✗	✗	✗	✗	✗	✗

Data citation

Citing source data is a compulsory requirement by most data licenses
(the “BY” component in a Creative Commons license)

A bibliographic citation of any digital object associated to a DOI can be obtained entering the code at

https://citation.crosscite.org/



FAIR data vs Open Data

Does following the FAIR principles mean that
data has to be shared openly with everyone? **NO!**

Data can be FAIR but not Open

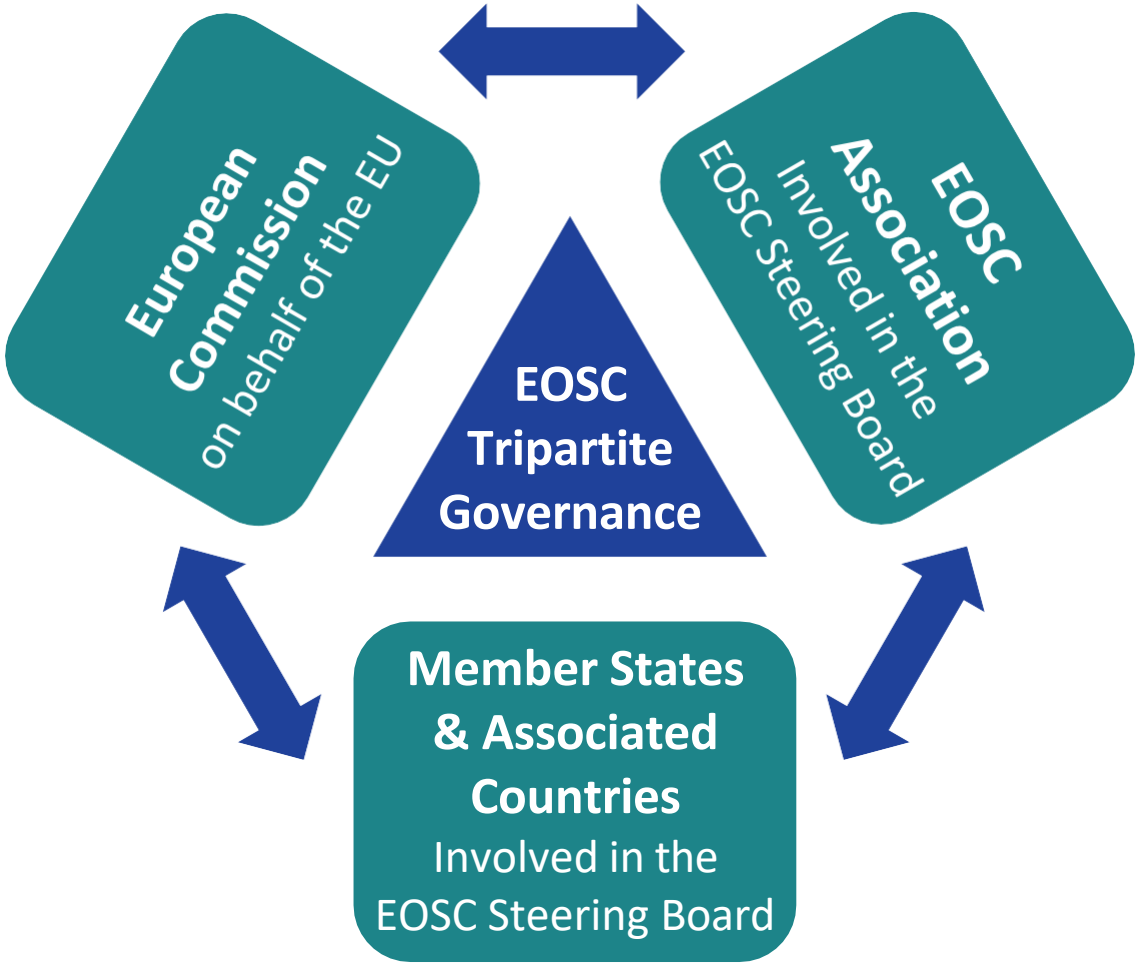
Data could meet the FAIR principles,
but be private or only shared under certain restrictions.

Open Data may not be FAIR

Publicly available data may lack sufficient documentation to
meet the FAIR principles, such as licensing for clear reuse.

Italian Initiatives

Italian participation in EOSC



ITA eosc

Italian Tripartite Assembly on the European Open Science Cloud

Roma, 5 giugno 2023, https://open-science.it/de/itaeosc2023

Mandated Organisation

Consortium GARR - ICDI, Italian Computing and Data Infrastructure

Members

Università del Piem. Orientale
Università della Tuscia
Università dell'Aquila
Università di Bologna
Università di Catania
Università di Ferrara
Università di Milano-Bicocca
Università di Padova
Università di Pisa
Università di Torino
Università S.O. Ben. di Napoli
Politecnico di Milano
Politecnico di Torino

Area Science Park
Consiglio Nazionale delle Ricerche (CNR)
Istituto Nazionale di Fisica Nucleare (INFN)
Istituto Nazionale di Astrofisica (INAF)
Istituto Nazionale di Geof. e Vulc. (INGV)
Istituto Nazionale di Oce. e di Geo. Sp. (OGS)
Gran Sasso Science Institute (GSSI)
Elettra-Sincrotrone Trieste S.C.p.A.
Scuola Intern. Sup. di Studi Avanzati (SISSA)
Scuola Normale Superiore di Pisa
Consorzio per Va. Biologiche e Farm. (CVBF)
Trust IT Service provider for research

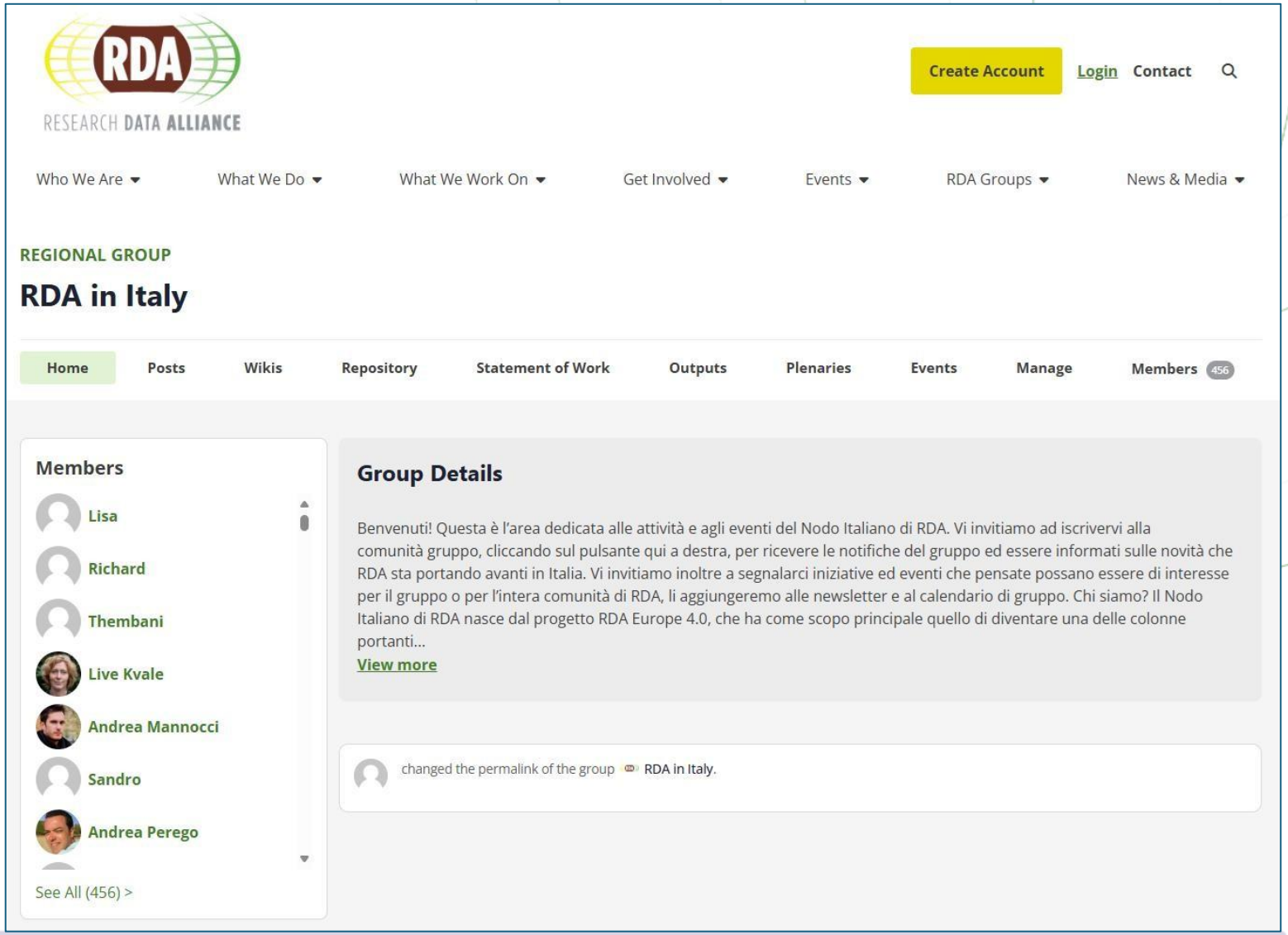
RDA - Research Data Alliance

International organisation founded in 2013 to promote open data sharing across technologies, disciplines, and countries.

Its mission is to build the social and technical bridges necessary for global data sharing and collaboration.

- Working Groups (62)
- Interest Groups (68)
- Communities of Practice (2)

Italian Community
https://www.rd-alliance.org/groups/rda-italy/activity/49462/



ICDI, the Italian Computing and Data Infrastructure

Forum created by representatives of major Italian Research Infrastructures and e-Infrastructures, with the aim of **promoting synergies** at national level, and **optimising** the Italian participation to European and global challenges in this field, including the European Open Science Cloud (EOSC), the European Data Infrastructure (EDI) and HPC.
https://www.icdi.it

ICDI Competence Centre

https://www.icdi.it/en/activities/competence-centre/open-science-cafe

Skills 4 eosc
https://www.skills4eosc.eu/

Italian Community of Data Stewards (CIDS)
https://open-science.it/

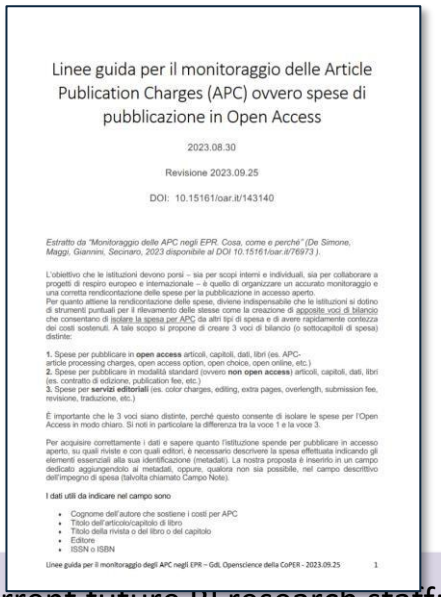
CoPER, Open Science Working Group

Facilitate the cooperation among Public Research Institutions in the production of documents and joint actions for the promotion and support of Open Science policies in Italy.

https://home.infn.it/coper/openscience.html

coper.openscience@lists.infn.it

Guidelines on the monitoring of Article Publication Charges (APC)



Linee guida per il monitoraggio delle Article Publication Charges (APC) ovvero spese di pubblicazione in Open Access

2023.08.30

Revisione 2023.09.25

DOI: 10.15161/oar.it/143140

Estratto da "Monitoraggio delle APC negli EPR. Cosa, come e perché" (De Simone, Maggi, Giannini, Secinaro, 2023 disponibile al DOI 10.15161/oar.it/76973).

L'obiettivo che le istituzioni devono porsi – sia per scopi interni e individuali, sia per collaborare a progetti di respiro europeo e internazionale – è quello di organizzare un accurato monitoraggio e una corretta rendicontazione delle spese per la pubblicazione in accesso aperto.
Per quanto attiene la rendicontazione delle spese, diviene indispensabile che le istituzioni si dotino di strumenti puntuali per il rilevamento delle stesse come la creazione di apposite voci di bilancio che consentano di isolare la spesa per APC da altri tipi di spesa e di avere rapidamente contezza dei costi sostenuti. A tale scopo si propone di creare 3 voci di bilancio (o sottocapitoli di spesa) distinte:

1. Spese per pubblicare in **open access** articoli, capitoli, dati, libri (es. APC- article processing charges, open access option, open choice, open online, etc.)
2. Spese per pubblicare in modalità standard (ovvero **non open access**) articoli, capitoli, dati, libri (es. contratto di edizione, publication fee, etc.)
3. Spese per **servizi editoriali** (es. color charges, editing, extra pages, overlength, submission fee, revisione, traduzione, etc.)

È importante che le 3 voci siano distinte, perché questo consente di isolare le spese per l'Open Access in modo chiaro. Si noti in particolare la differenza tra la voce 1 e la voce 3.

Per acquisire correttamente i dati e sapere quanto l'istituzione spende per pubblicare in accesso aperto, su quali riviste e con quali editori, è necessario descrivere la spesa effettuata indicando gli elementi essenziali alla sua identificazione (metadati). La nostra proposta è inserirlo in un campo dedicato aggiungendolo ai metadati, oppure, qualora non sia possibile, nel campo descrittivo dell'impegno di spesa (talvolta chiamato Campo Note).

I dati utili da indicare nel campo sono

- Cognome dell'autore che sostiene i costi per APC
- Titolo dell'articolo/capitolo di libro
- Titolo della rivista o del libro o del capitolo
- Editore
- ISSN o ISBN

Linee guida per il monitoraggio degli APC negli EPR – GdL Openscience della CoPER - 2023.09.25 1

Permanent monitoring of institutional Open Science policies (e.g. open access to publications, open data, platforms, rewarding)

Monitoraggio sulle politiche di gestione istituzionale dei dati scientifici

Pagina a cura del Gruppo di Lavoro Open Science della CoPER.

Il rapporto che descrive il sondaggio originale condotto tra marzo e giugno 2023 è citabile come
Locati M., Chiodetti A.G., Saraò A., Sala M., Palma D., Scano R., Cipolloni C., Menasce D., Bianco S. (2023). Risultati del sondaggio sulle politiche di gestione istituzionale dei dati scientifici. https://doi.org/10.15161/oar.it/77195

Abilita l'ente per cui desideri mostrare le informazioni disponibili

☐ AreaSciencePark ☐ ASI ☑ CNR ☐ CREA ☐ CREF ☐ ENEA ☐ IISG ☐ INAF ☐ INAPP ☐ INDAM ☐ INDIRE ☐ INFN ☑ INGV
☐ INRIM ☐ INVALSI ☑ ISPRA ☐ ISS ☐ ISTAT ☑ OGS ☐ SZN

Download

Nome dell'ente	Consiglio Nazionale delle Ricerche (CNR)	Istituto Nazionale di Geofisica e Vulcanologia (INGV)	Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA)	Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS)
Homepage	https://www.cnr.it/	https://www.ingv.it/	https://www.isprambiente.gov.it/	https://www.ogs.it/
Codice ROR	04zaypm56	00qps9a02	022zv0672	04y4t7k95
Ente vigilante	Ministero dell'Università e della Ricerca (MUR)	Ministero dell'Università e della Ricerca (MUR)	Ministero dell'Ambiente e della Sicurezza Energetica (MASE)	Ministero dell'Università e della Ricerca (MUR)
Data di ultimo aggiornamento	28/04/2023	19/04/2023	21/04/2023	19/04/2023
email del compilatore	openaccess@cnr.it	mario.locati@ingv.it	carlo.cipolloni@isprambiente.it	asarao@ogs.it
1. Politica sulla Gestione dei Dati	No	Si	Si	In preparazione
1.1 Link a uno o più documenti o descrizione della Politica sulla Gestione dei Dati	--	La Politica dei Dati dell'INGV / The INGV Data Policy, Link	Link	--
2. Politica su dati FAIR	No	Si	Si	In preparazione
2.1 Link a uno o più documenti o descrizione della Politica su dati FAIR	--	La Politica dei Dati dell'INGV / The INGV Data Policy, Link	Data Policy approvata da CdA e pubblicata Link ; Link	--
3. Politica su Open Data	No	Si	Si	In preparazione
3.1 Link a uno o più documenti o descrizione della Politica su Open Data	--	La Politica dei Dati dell'INGV / The INGV Data Policy, Link	Data Policy approvata da CdA e pubblicata Link	--
4. Catalogo di metadati	No	Si	Si	No
4.1 Link al Catalogo di Metadati	--	INGV Open Data Portal, Link . Il catalogo è chiamato "Registro Dati" ed è citabile come: INGV Data Management Office (2020). INGV Open Data Registry, the metadata catalogue of the Istituto Nazionale di Geofisica e Vulcanologia. Istituto Nazionale di Geofisica e Vulcanologia (INGV). Link	Link	--
5. Archivio dati istituzionale	No	In preparazione	In preparazione	No

AISA
Italian association for the promotion of Open Science

Encourage the values of open access to knowledge by:

- disseminate a culture of Open Science;
- publish studies on the implementation of its principles;
- provide staff training programs;
- organise conferences and seminars;
- engage international cooperation;
- promote participation in international projects;
- raise awareness among decision makers.

https://aisa.sp.unipi.it/about-aisa/

COARA, Coalition for Advancing Research Assessment

70 Italian organisations signed the Agreement on Reforming Research Assessment (ARRA).

50 Italian Research are member and participate to the COARA National Chapter activities.

11 Italian organisations adopted and shared an action plan.

https://www.coara-italia.it/

SIGNATORY

optional membership

MEMBER

ENDORSE PUBLICLY
Publicly signal the need and your willingness to reform the current ways of research assessment.

ACHIEVE CHANGE
Implement the 10 CoARA commitments in your organisation by implementing an action plan.

JOIN THE COMMUNITIES
Benefit from communities of practice; join Working Groups and National Chapters

HAVE A VOICE
Shape CoARA's strategy by participating in governance and decision-making

NETWORK AND LEARN
Interact and share best practices with the CoARA network of members

AgID, the Agency for Digital Italy

Technical agency of the Presidency of the Council of Ministers that guarantees the achievement of the objectives of the **Italian Digital Agenda**, coordinating **Italian Public Administrations**.
https://www.agid.gov.it/

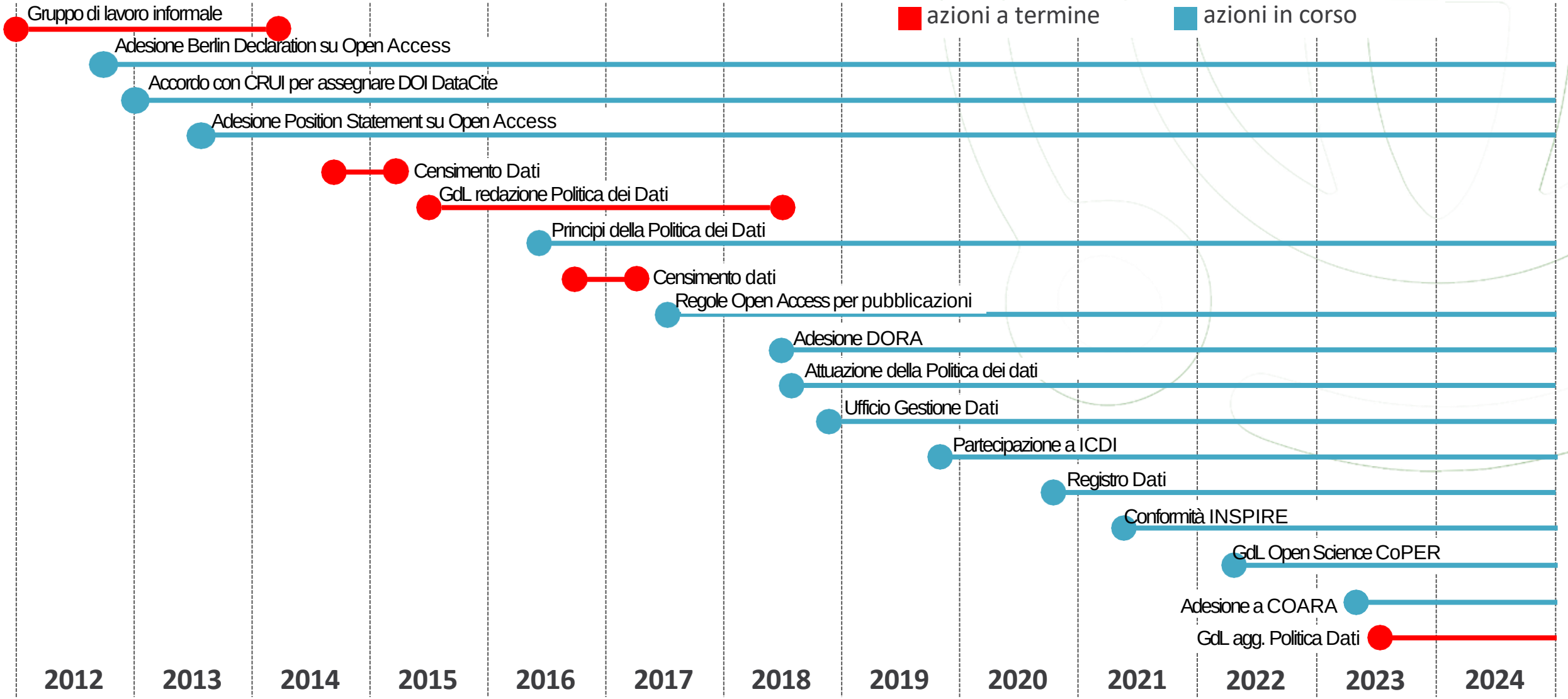
Areas of intervention:

- Digital transition
- Digital skills and education
- Monitoring and Surveillance

Guidelines:

- Accessibility of IT tools
- Open data and re-use of Public Sector Information
- Design websites and digital services
- Management electronic documents
- Management of the National Index of Digital Domiciles
- e-procurement
- Interoperability standards and technologies
- Acquisition and reuse of software
- Implementation of the eIDAS regulation
- Cybersecurity
- SPID, public digital identity system

Temporal evolution of the INGV data policy



https://istituto.ingv.it/ufficio-gestione-dati

Tools

Online tools for managing DMPs

Tool	Description
DMPTool	A free, open-source tool that provides detailed guidance and resources for creating, reviewing, and sharing data management plans that meet institutional and funder requirements.
DMPonline	Similar to DMPTool, DMPonline helps researchers create, review, and share data management plans. It is customized to meet the requirements of various funders and institutions, primarily used in the UK and Europe.
Argos	An open-source tool developed by OpenAIRE, aimed at simplifying the process of creating, sharing, and exporting data management plans. It focuses on the European research area but is open to global users.
RDMO	A tool that enables researchers and institutions to plan, implement, and manage the life cycle of research data. RDMO offers a structured approach to creating and maintaining DMPs.
easyDMP	Provided by Sigma2, it is an online tool for creating data management plans, offering templates and guidance to ensure compliance with various requirements.
Data Stewardship Wizard (DSW)	A tool that provides a dynamic and user-friendly interface for creating data management plans, integrating knowledge models to tailor DMPs according to specific project needs and funder requirements.

Credits

Many slides in this presentation were prepared by me in the framework of these two projects

EU Horizon Europe project

https://www.skills4eosc.eu/
https://cordis.europa.eu/project/id/101058527

PNRR Research Infrastructure Project “MEET” - Monitoring Earth’s Evolution and Tectonics
Mission 4 “Education and Research”

THANKS!

IR0000032 – ITINERIS, Italian Integrated Environmental Research Infrastructures System
(D.D. n. 130/2022 - CUP B53C22002150006) Funded by EU - Next Generation EU PNRR-Mission 4 "Education and Research" - Component 2: "From research to business" - Investment 3.1: "Fund for the realisation of an integrated system of research and innovation infrastructures"

Finanziato dall'Unione europea NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani