IBEC

Institute for Bioengineering of Catalonia

Open Science Workshops

OSW2 · Research Data Management

May 2025

Fidel Bellmunt
Knowledge Manager
fbellmunt@ibecbarcelona.eu

SHAPING THE PRESENT AND FUTURE OF MEDICINE WITH
BIOENGINEERING

IBEC
Institute for Bioengineering of Catalonia

Data management workshop

Program

Institute for Bioengineering of Catalonia

9:30 Start

1. Introduction, the Open Science framework
2. Definitions and the Research Data Lifecycle
3. Research Data Management policies, funders current requirements
4. The FAIR principles

11:00 Coffee-break

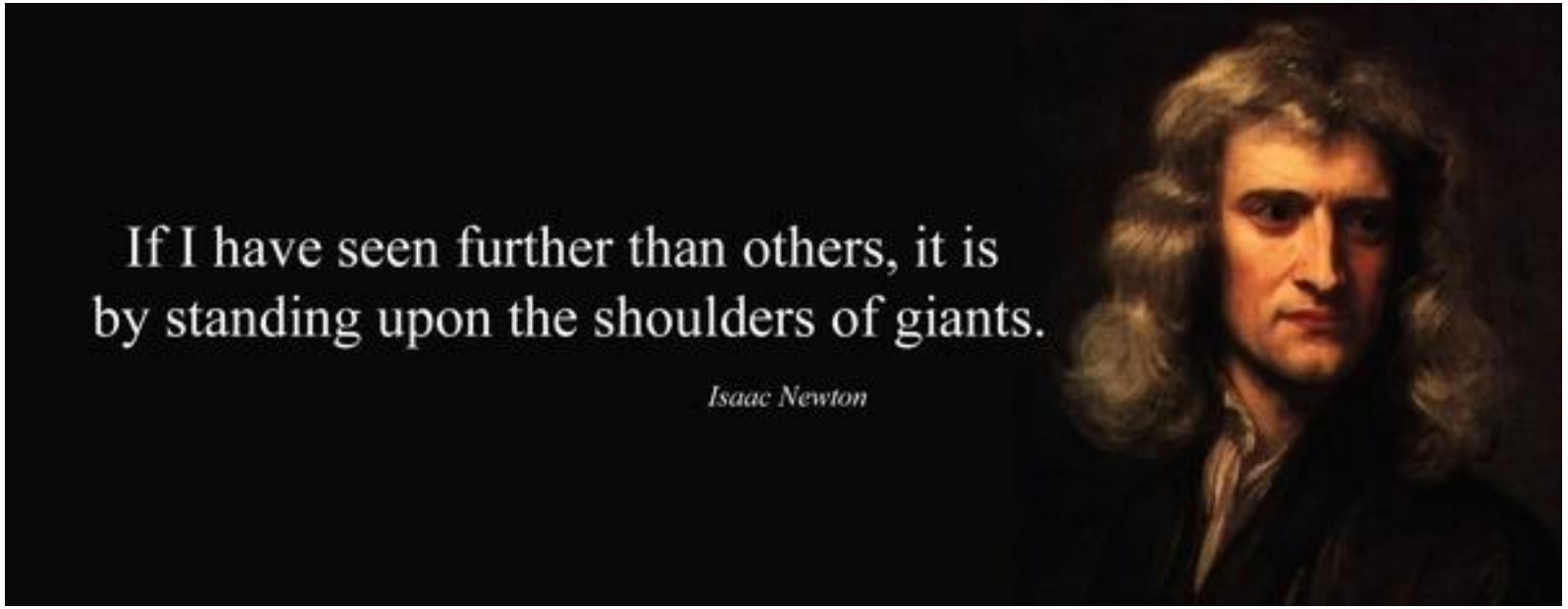
5. IBEC's Research Data Management Policy, procedures and tools
6. The Data Management Plan
7. Choosing a data repository
8. Licenses and copyright

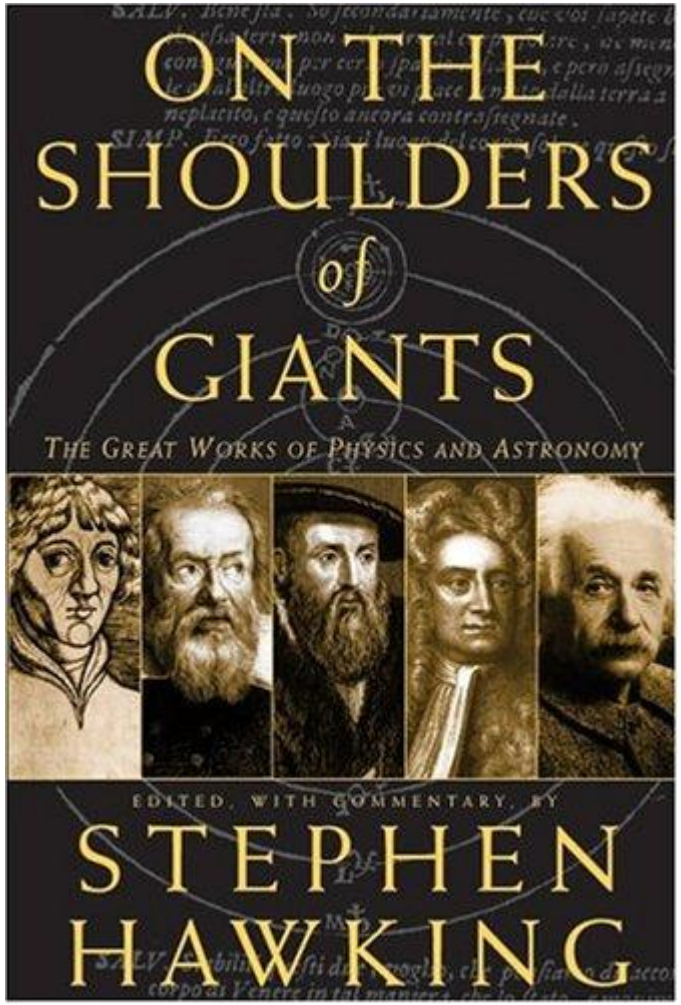
12:00 Q&A and closing remarks

Data management workshop

1. Introduction, the Open Science framework

IBEC
Institute for Bioengineering of Catalonia





1. Introduction, the Open Science framework

Institute for Bioengineering of Catalonia

Definitions:

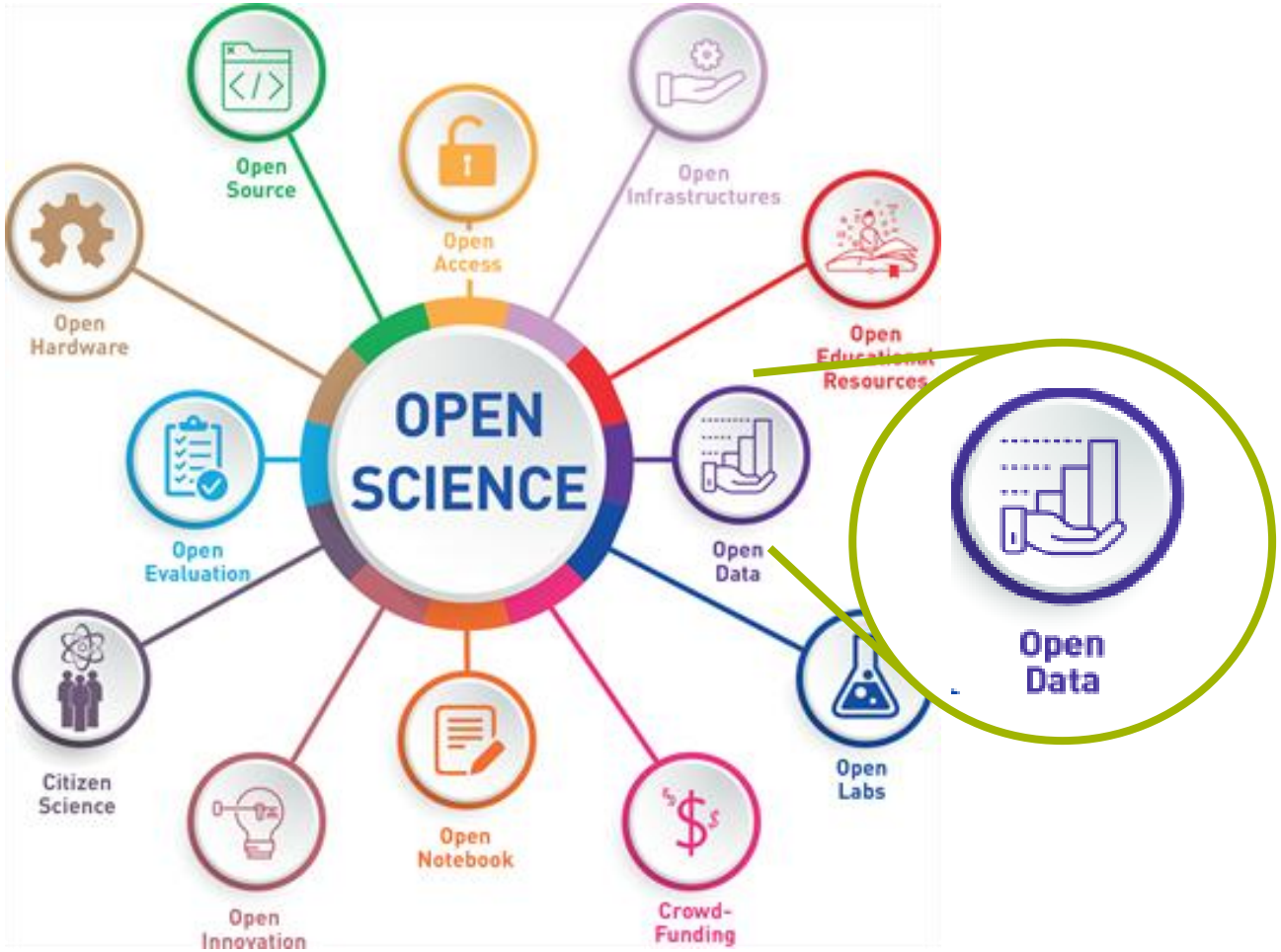
Open Science is science done right. (popular)

Open science refers to a new approach to the scientific process based on cooperative work and new ways of disseminating knowledge, improving accessibility to and re-usability of research outputs by using digital technologies and new collaborative tools. (EC, 2018)

Open Science is transparent and accessible knowledge that is shared and developed through collaborative networks. (Vicente-Saez y Martínez, 2018)

Interesting resources about OS:

- Podcast **Open Science Stories** by Heidi Seibold: https://anchor.fm/opensciencestories
- **Passport for Open Science** – A Practical Guide for PhD Students: https://www.ouvrirlascience.fr/passport-for-open-science-a-practical-guide-for-phd-students/

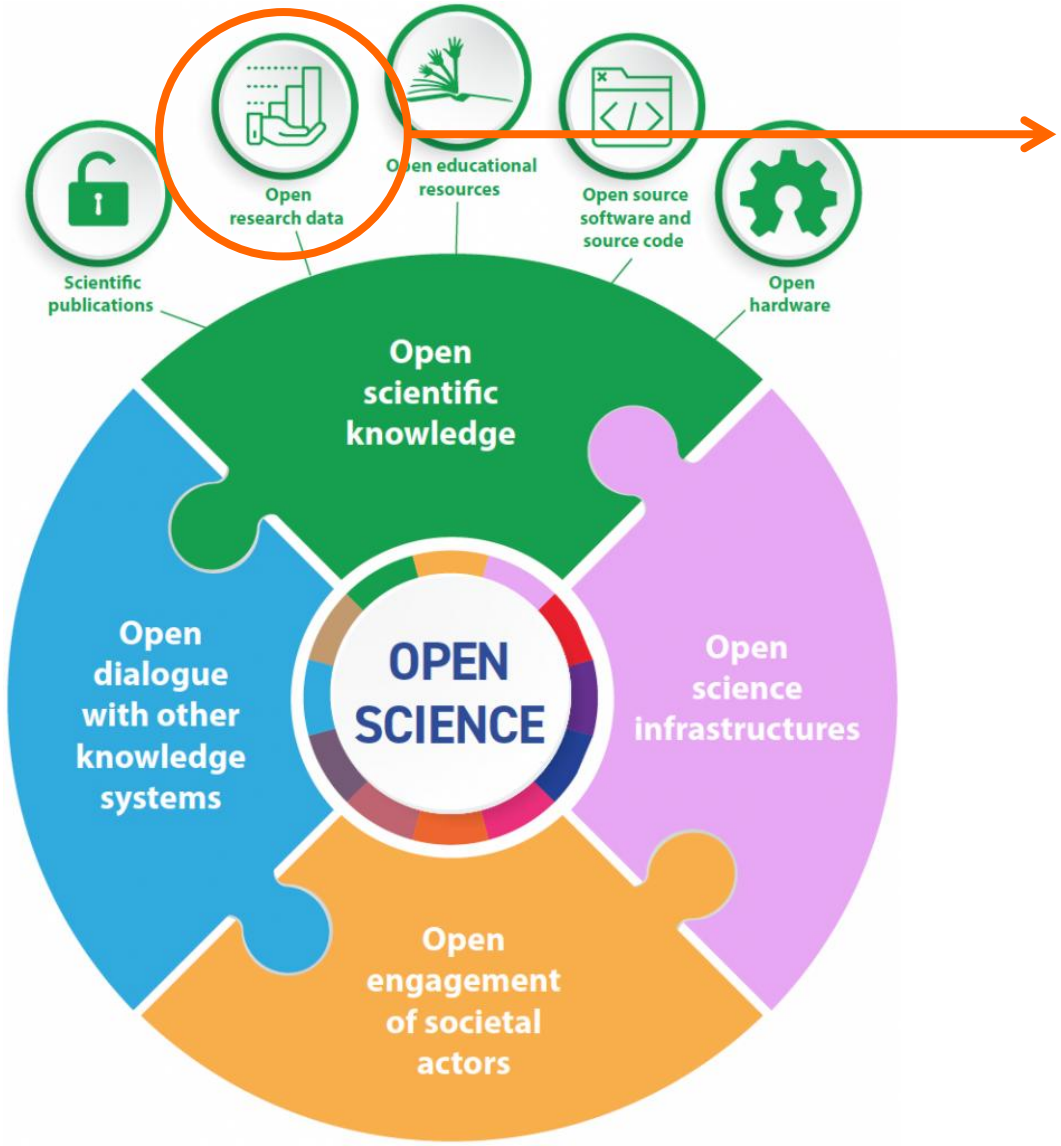


Open science is defined as an inclusive construct that combines various movements and practices aiming to make multilingual scientific knowledge openly available, accessible and reusable for everyone, to increase scientific collaborations and sharing of information for the benefits of science and society, and to open the processes of scientific knowledge creation, evaluation and communication to societal actors beyond the traditional scientific community. It comprises all scientific disciplines and aspects of scholarly practices, including basic and applied sciences, natural and social sciences and the humanities, and it builds on the following key pillars: open scientific knowledge, open science infrastructures, science communication, open engagement of societal actors and open dialogue with other knowledge systems.

(UNESCO Recommendation on Open Science, 2021): https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en

2. Definitions and the Research Data Life Cycle

IBEC
Institute for Bioengineering of Catalonia



b. Open research data that include, among others, digital and analogue data, both raw and processed, and the accompanying metadata, as well as numerical scores, textual records, images and sounds, protocols, analysis code and workflows that can be openly used, reused, retained and redistributed by anyone, subject to acknowledgement. Open research data are available in a timely and user-friendly, human- and machine-readable and actionable format, in accordance with principles of good data governance and stewardship, notably the FAIR (Findable, Accessible, Interoperable, and Reusable) principles, supported by regular curation and maintenance.

(UNESCO Recommendation, 2021, p. 9)

IBEC
Institute for Bioengineering of Catalonia

Research data is any information that has been collected, observed, generated or created to validate original research findings. Research data may be arranged or formatted in a such a way as to make it suitable for communication, interpretation and processing. Data comes in many formats, both digital and physical.

The research data lifecycle describes the different stages research data go through before, during, and after a research project. Various data management activities take place within each stage of the data lifecycle, and the choices made in one stage influence the next one.

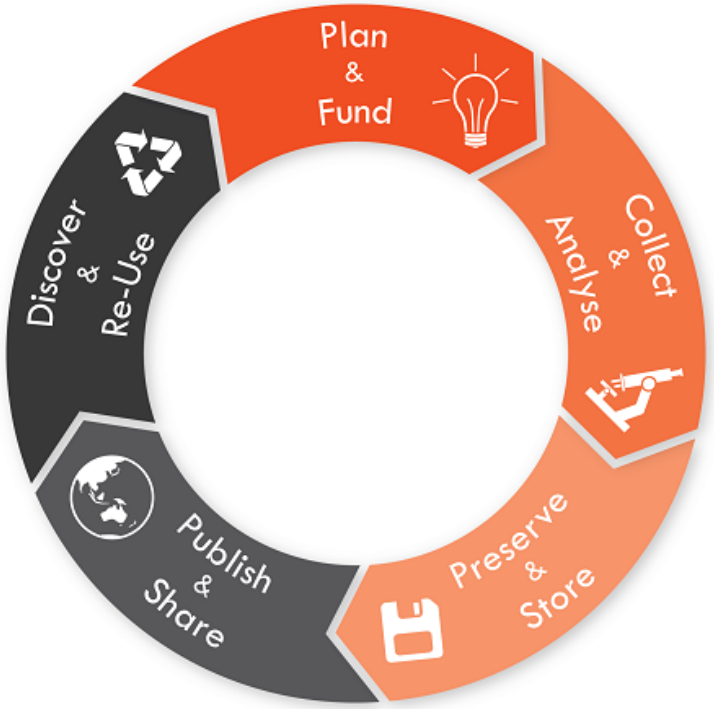
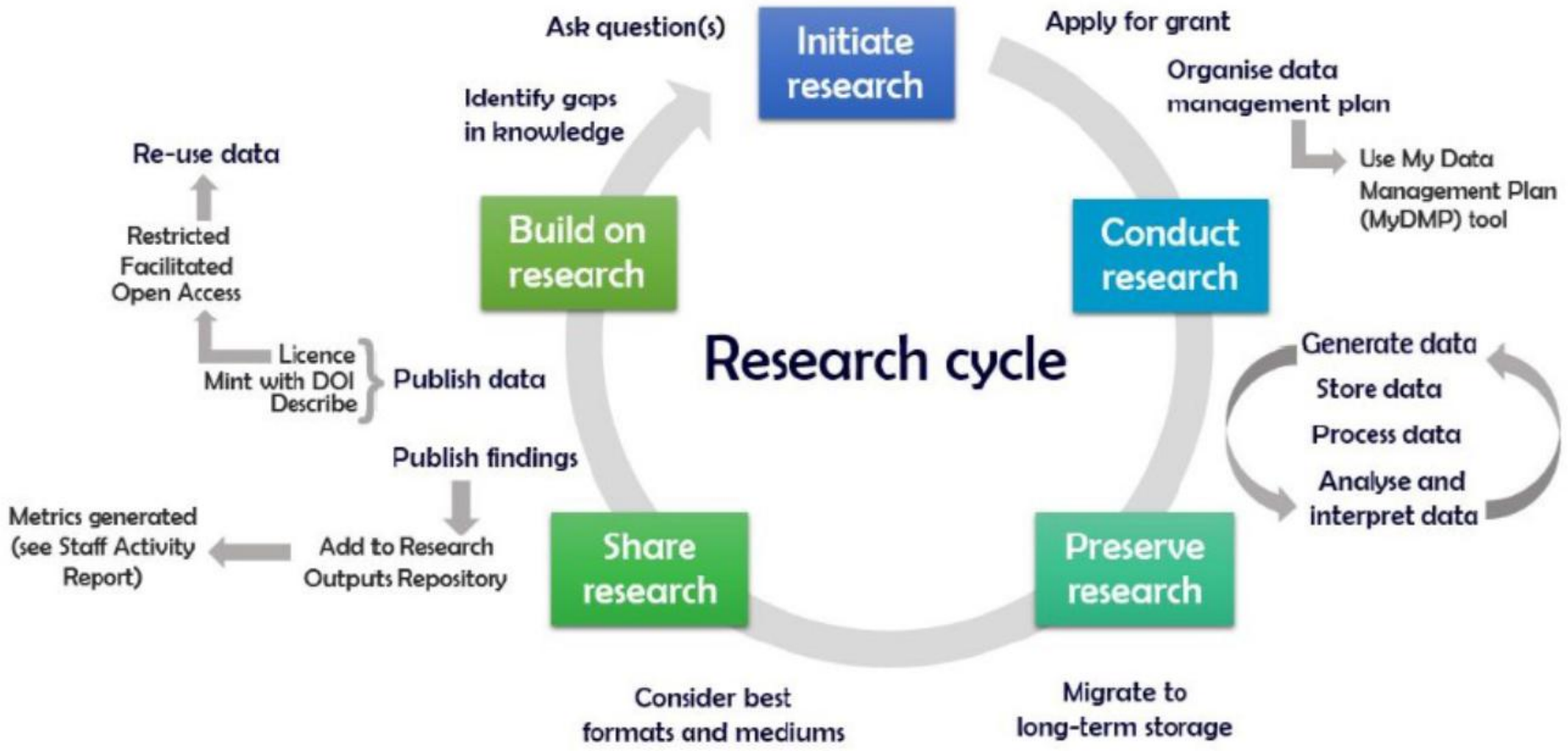


Image from OSF: https://osf.io/fr8pd/wiki/home/

Actions to carry on during the research data lifecycle:



Infographic by: Ignasi Labastida, Universitat de Barcelona

Institute for Bioengineering of Catalonia

Research data management (RDM) refers to the organization, storage and preservation of data created during a research project. Through its lifecycle, it covers initial planning, day-to-day processes and long-term archiving and sharing. It's required to *make excellent science* and to comply with *funders requirements*:

Ensure research integrity and replication
Ensure research data and records are accurate, complete, authentic and reliable
Increase your research efficiency
Save time and resources in the long run
Enhance data security and minimize the risk of data loss
Prevent duplication by enabling others to use your data
Comply with practices conducted in industry and commerce
Protect your institution from reputational, financial & legal risk
Fulfill publisher requirements
Fulfill funding body grant requirements

IBEC
Institute for Bioengineering of Catalonia

Research data types

Infographic by Science Nature:
https://researchdata.springernature.com/documents/web_a92645_what-are-research-data-revision

- documents (text, Word), spreadsheets;
- questionnaires, transcripts, codebooks;
- audiotapes, videotapes;
- photographs, films;
- test responses;
- slides, artefacts, specimens, samples;
- digital objects acquired and generated during research;
- data files;
- database contents;
- models, algorithms, scripts;
- contents of an application (input, output, logfiles for analysis software, simulation software, schemas);
- methodologies and workflows.

Institute for Bioengineering of Catalonia

What should be done to manage research data correctly?

Planning and describing data-related work *before* it takes place;
Documenting your data (and processing/workflows) so that others can find and understand it;
Choosing open (or at least standardised) file formats where possible;
Storing data safely during a project;
Depositing data in a trusted archive at the end of the research;
Creating metadata records for datasets and licensing them appropriately;
Linking publications to the datasets (and increasingly the code and protocols).

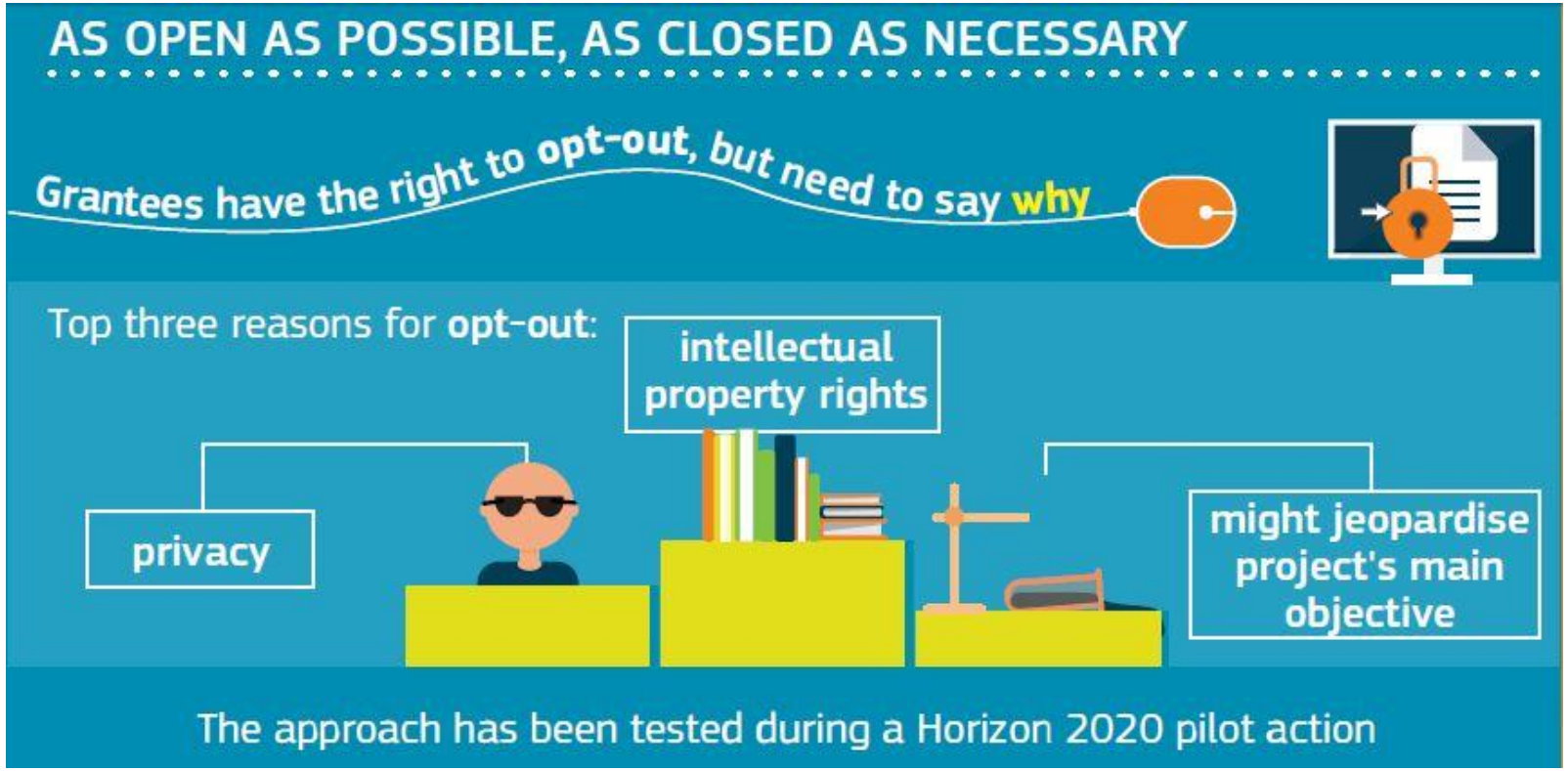
Need of:

- IT infrastructures,
- Methodologies,
- Tools,
- Standards,
- Protocols.

3. Research Data Management policies, funders current requirements

Institute for Bioengineering of Catalonia

Many funders, especially the EC, ask for open access to research data under the principle *as open as possible, as closed as necessary*, and in either case, a Data Management Plan.



Data management benefits

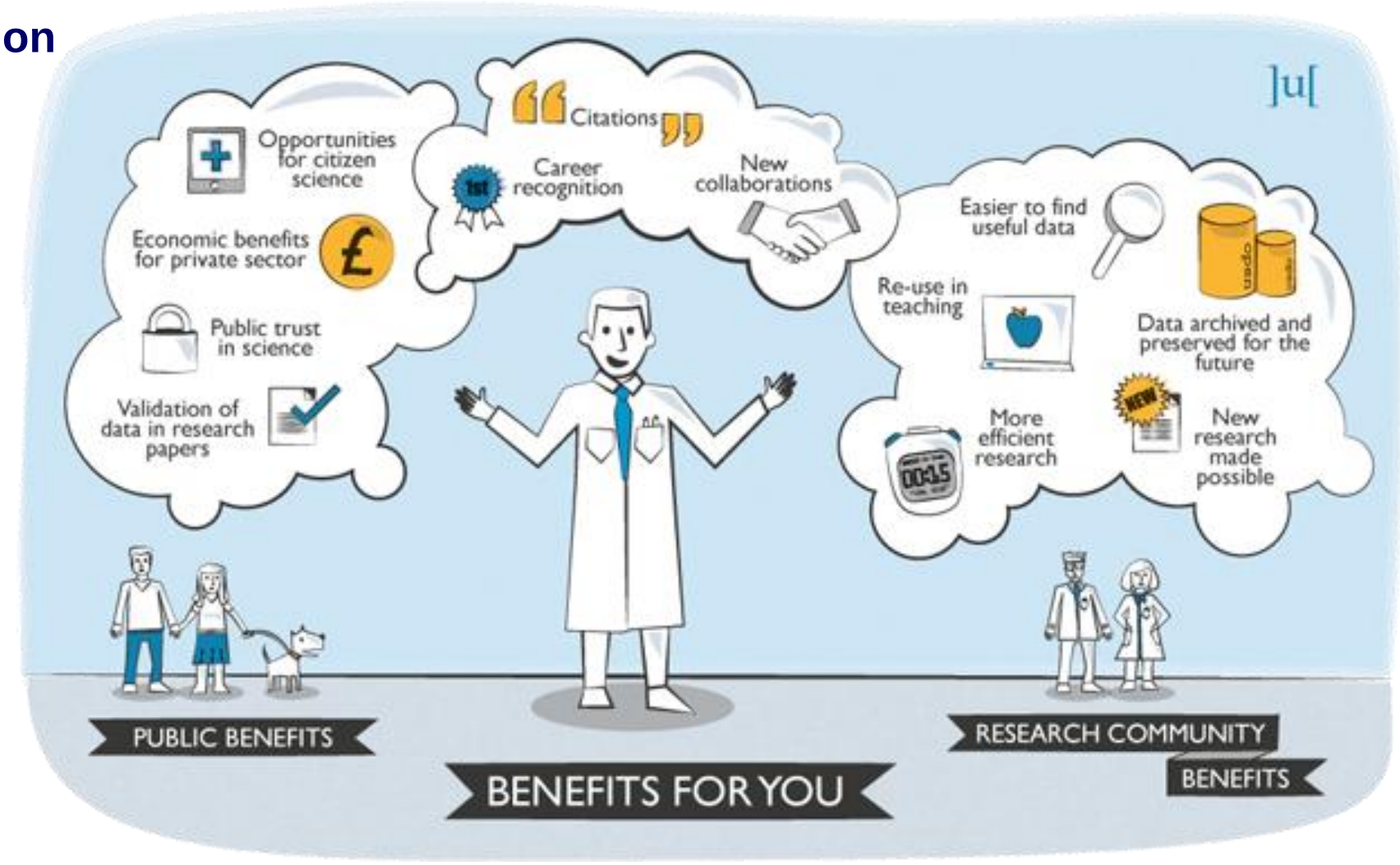
There are numerous reasons why research data management has become a policy requirement of major funders and research institutions: fundamentally and most importantly it improves academic quality, transparency and the robustness of the scholarly record. But also:

- DURABILITY: Simply put, fewer important datasets will be lost
- SPEED: The research process becomes faster
- IMPACT and LONGEVITY: Data linked to publications receive more citations, over longer periods
- EFFICIENCY: Data collection can be funded once, and used many times for a variety of purposes
- ACCESSIBILITY: Interested third parties can (where appropriate) access and build upon publicly-funded research outputs with minimal barriers to access

Data management workshop - 3. Research Data Management policies, funders current requirements

IBEC
Institute for Bioengineering of Catalonia

Benefits on opening data



(c) poster: Hole, Brian (2012) Poster: The Journal of Open Archaeology Data. Figshare.
https://doi.org/10.6084/m9.figshare.96890.v1

Funders - European Comission

Mandatory open science practices

[…]

responsible management of research data in line with the FAIR principles of 'Findability', 'Accessibility', 'Interoperability' and 'Reusability', notably through the generalised use of data management plans, and open access to research data under the principle 'as open as possible, as closed as necessary', under the conditions required by the grant agreement;

information about the research outputs/tools/instruments needed to validate the conclusions of scientific publications or to validate/re-use research data;

digital or physical access to the results needed to validate the conclusions of scientific publications, unless exceptions apply; o in cases of public emergency, if requested by the granting authority, immediate open access to all research outputs under open licenses or, if exceptions apply, access under fair and reasonable conditions to legal entities that need the research outputs to address the public emergency.

Source:
Horizon Europe Guide https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/programme-guide_horizon_en.pdf / Open Science section, on page 41

Funders - Spanish context

Ley 17/2022, de 5 de septiembre, por la que se modifica la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación.

https://www.boe.es/eli/es/l/2022/09/05/17/con

Treinta y nueve. Se modifica el artículo 37, que queda redactado en los siguientes términos: Artículo 37. Ciencia abierta.

➔ 2. El personal de investigación del sector público o cuya actividad investigadora esté financiada mayoritariamente con fondos públicos y que opte por diseminar sus resultados de investigación en publicaciones científicas, deberá depositar una copia de la versión final aceptada para publicación y los **datos asociados a las mismas** en repositorios institucionales o temáticos de acceso abierto, de forma simultánea a la fecha de publicación.

LLEI 9/2022, del 21 de desembre, de la ciència.

https://portaljuridic.gencat.cat/eli/es-ct/l/2022/12/21/9

Article 79. Informació i transmissió del coneixement

1. [...] L'estratègia de ciència oberta de Catalunya ha de girar al voltant dels sis eixos següents:

b) La gestió de les dades de recerca, amb l'objectiu que es publiquin de manera tan oberta com sigui possible i seguint els principis FAIR, per tal de garantir-ne la preservació i reutilització i de desenvolupar bones pràctiques en llur gestió.

[…]

Data management workshop

4. The FAIR principles

Institute for Bioengineering of Catalonia

FAIR Data Principles (Findable, Accessible, Interoperable, Re-usable) support knowledge discovery and innovation as well as data and knowledge integration and promote sharing and reuse of data. The principles help data and metadata to be 'machine readable', supporting new discoveries through the harvest and analysis of multiple datasets.

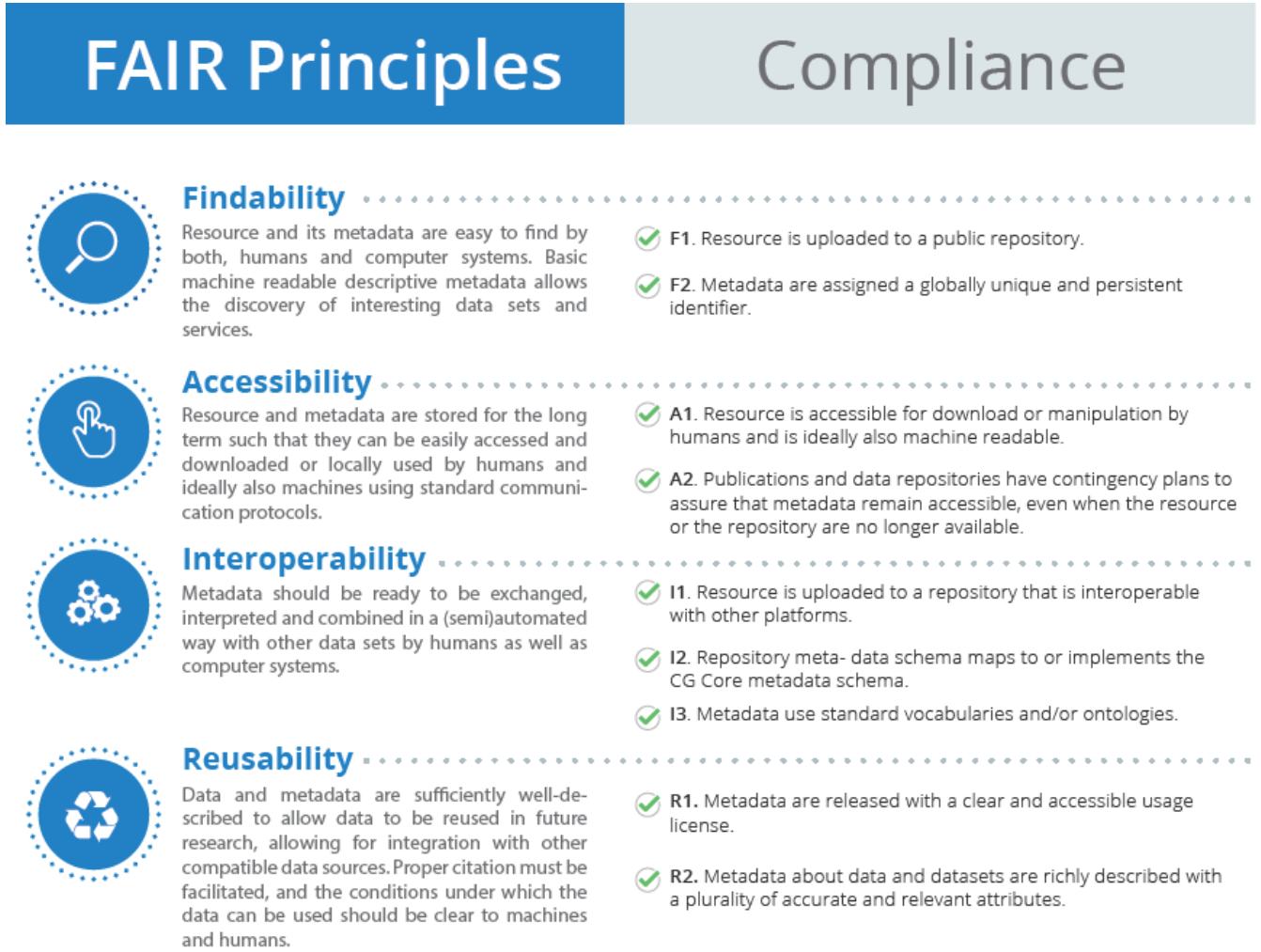
Origin:
Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

If you cannot make your data open, at least make it FAIR!

Credit: vector graphics in the image are licensed for free use by rawpixel.com / Freepik

Data management workshop - 4. The FAIR principles

IBEC
Institute for Bioengineering of Catalonia



https://ccafs.cgiar.org/open-access-and-fair-principles

IBEC
Institute for Bioengineering of Catalonia

Data creation and organization best practices

(folder structure, file formats, file naming, documenting data and software, …)

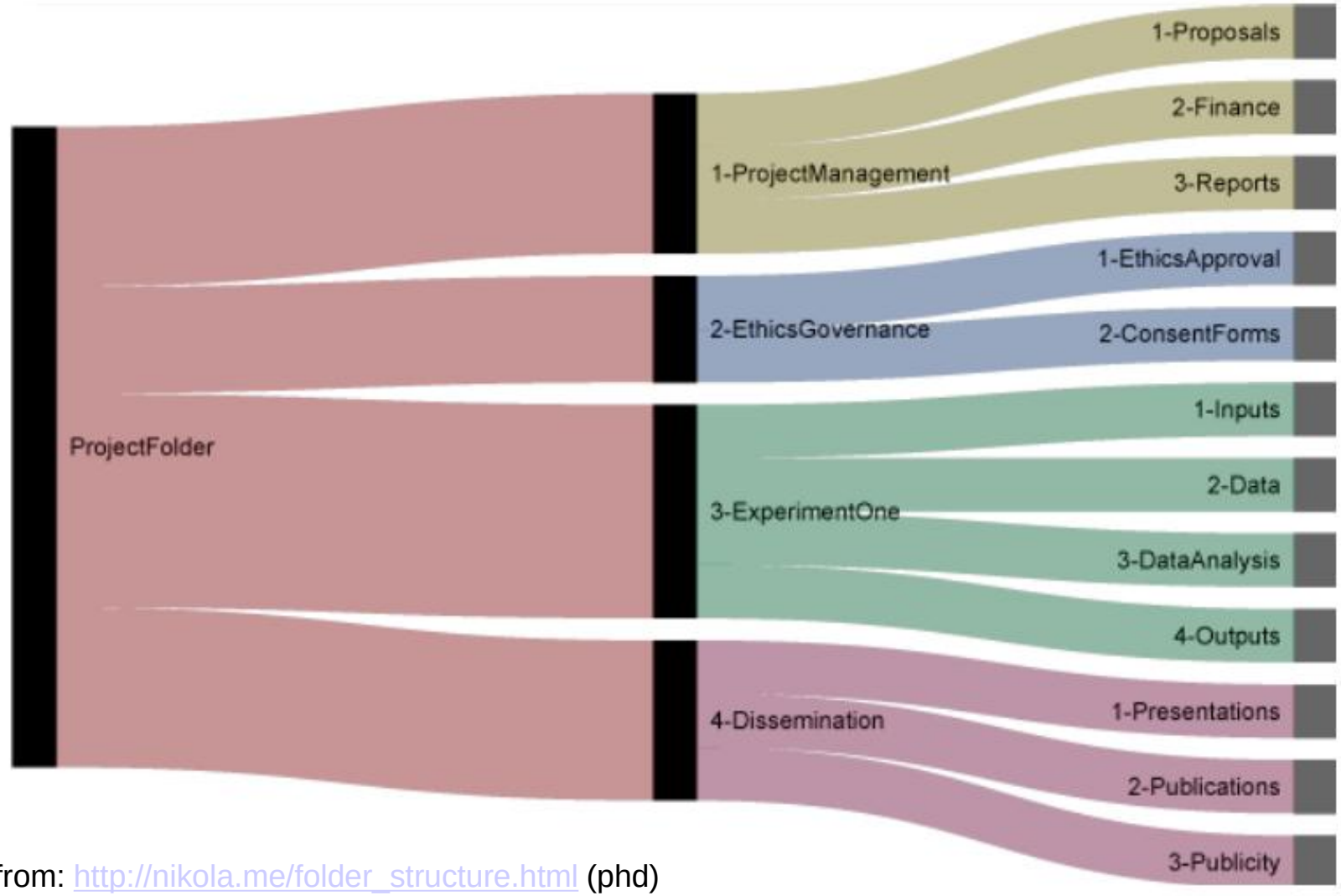
Folder structure: Determine the tree levels and naming criteria for the folders. Choose a consistent organizational structure for all your project folders. Although it may seem obvious, thinking about the structure of your folders and planning effectively makes navigation much easier. Minimize the number of clicks necessary to reach files. In conjunction with a consistent file naming convention, an efficient structure saves a lot of time.

Best practices:

- Be consistent.
- Structure your hierarchy logically. Follow the logic that makes the most sense for your project.
- Keep folders and subfolders separate to reduce overlap. However, don't make an excessive number of subfolders.
- Keep subfolder categories narrow to restrict the number of files in each.
- Your Desktop is meant to be temporary storage. Never keep files there for longer than necessary.
- When naming folders, think about information you might want when looking up files, and use numerals to keep the desired order (the system sorts automatically by names).
- Consider a final project folder with all the relevant final documents and data, once closed.

Data management workshop - 4. The FAIR principles

IBEC
Institute for Bioengineering of Catalonia



Extracted from: http://nikola.me/folder_structure.html (phd)

IBEC
Institute for Bioengineering of Catalonia

File naming: avoid special characters and give a meaningful name (audit trail)

- File name e.g. IP02R0120180731.docx
- File name components I P02 R01 20180731
- I = interview (type of data)
- P[n] = participant ID - participant 02
- R[n] = researcher ID – researcher 01
- Date of interview in form YYYYMMDD – 20180731
  - ✓ Compare with the file name interview02.docx

Recommended file formats to allow interoperability

Use:

Common formats (may be de facto standards)

Standard formats accepted in your field

Interchangeable or open (published) formats for long-term preservation

Avoid : Dependency on proprietary software to render your data

Data management workshop - 4. The FAIR principles

Institute for Bioengineering of Catalonia

- Documentation & metadata

Documentation

The human-readable stuff that contextualises the research outputs and processes so your future self or others can understand how you got your findings and/or how the data can be repurposed.

- Data dictionary
- Readme file
- Study protocol
- Methodology statement
- Sampling frame description

Format: often text or PDF

Metadata

Machine-readable, standardised fields that allow discovery through search engines, or mark up the structure of a database, or show relationships between different digital objects.

- Dublin Core (DCMI)
- DataCite
- MI, Minimum Information Standards in Biosciences, e.g. MIAME, gene expression microarray.
- Metadata Standards Catalog: https://rdamsc.bath.ac.uk/

Format: often XML or JSON

- Use of data management tools

such as Electronic Laboratory Notebooks (ELN)

Institute for Bioengineering of Catalonia

When open access data is not possible?

Because potential risk / harm to research subjects is too great.

Information that can be used to discriminate requires extra protection.

When it is not permitted by the data producer, funder, health authority etc.

Sometimes precautions are required even for anonymized data.

Because anonymization is either not feasible or would negate the value of a dataset.

Population too small to be anonymous, e.g. those with a rare genetic condition.

5. IBEC's Research Data Management Policy, procedures and tools

Institute for Bioengineering of Catalonia

IBEC's Research Data Management Policy:
https://ibecbarcelona.eu/wp-content/uploads/2021/11/IBEC-Data-Management-Policy.pdf

As a result of its activity, IBEC generates research data, defined as all information (independent of form or presentation) needed to support or validate the development, results, observations or findings of a research project, including contextual information. Research data include all materials which are created in the course of academic work, including digitization, records, source research, experiments, measurements, surveys and interviews. This includes software and code. Research data can take on several forms: during the lifespan of a research project, data can exist as gradations of raw data, processed data (including negative and inconclusive results), shared data, published data and Open Access published data, and with varying levels of access, including open data, restricted data and closed data.

Data management workshop - 5. IBEC's Research Data Management Policy, procedures and tools

Institute for Bioengineering of Catalonia

Procedures, recommendations and guides:

Guide: *Managing research data in IBEC research groups*: available at the Intranet and published openly at: https://ibecbarcelona.eu/wp-content/uploads/2023/06/RDM_IBEC_Guide_eng_d1.pdf

Research Data Management Working Group (with representatives of each lab).

Tools:

Institutional ELN: eLabFTW.

IT Services (storage, etc)

Institutional repository, CORA.RDR: https://dataverse.csuc.cat/dataverse/IBEC

Service: knowledge manager from Strategic Initiatives Unit

Advice on DMP elaboration in collaboration with Projects Office

Assistance in data curation and publishing

Resources at IBECNet: Knowledge > Resources: https://ibecnet.ibecbarcelona.eu/Resources

OS section at new IBEC website: https://ibecbarcelona.eu/open-science/

OS Policy: https://ibecbarcelona.eu/wp-content/uploads/2023/07/IBEC_Policy_OpenScience_v20230627.pdf

6. The Data Management Plan

A data management plan or DMP is a formal document that outlines how data are to be handled both during a research project, and after the project is completed. The goal of a data management plan is to consider the many aspects of data management, metadata generation, data preservation, and analysis before the project begins; this may lead to data being well-managed in the present and prepared for preservation in the future.

Wikipedia: https://en.wikipedia.org/wiki/Data_management_plan

Data management workshop - 6. The Data Management Plan

Institute for Bioengineering of Catalonia

What is data management planning?

The act of planning how you will manage all aspects of your data before you begin collecting or creating it. It may include:

- Where and how you will store and back-up your data;
- How and when it will be shared with collaborators;
- The steps required to anonymise sensitive data;
- When, how, and where your data will be preserved and shared.

Writing DMPs can help to:

- Make informed decisions to anticipate & avoid problems.
- Avoid duplication, data loss and security breaches.
- Develop procedures early on for consistency.
- Ensure data are accurate, complete, reliable and secure.
- Save time and effort to make your lives easier.
- Plan to share data early on and increase impact.

Elements of data management

Dataset

A set of files containing both research data - usually numeric or encoded - and documentation sufficient to make the data re-usable.

Documentation

Any digital files such as a codebook, technical or methodology report or user guide, which explain the research data's production, provenance, processing or interpretation.

Metadata

Information about a data item in the repository, including descriptive metadata such as title and other fields used in a citation, and administrative metadata such as date of submission. Usually conforms to a standard to allow computer-to-computer interoperability.

Digital repository

Differs from other digital collections in that: content is deposited in a repository, whether by the content creator, owner or third party; the repository architecture manages content as well as metadata; the repository offers a minimum set of basic services e.g. put, get, search, access control; the repository must be sustainable and trusted, well-supported and well-managed.

DMP Contents at Horizon 2020 template

1. Data summary (description, origin, total size, formats, usefulness…)
2. FAIR data:
   - Making data findable, including provisions for metadata
   - Making data openly accessible
   - Making data interoperable
   - Increase data re-use (through clarifying licenses)
3. Allocation of resources
4. Data security
5. Ethical aspects
6. Other relevant aspects

7. Choosing a data repository

Data repositories

Most repositories will expect data to be deposited in preferred preservation formats (to enable reuse also in the long term) and for accompanying high-quality documentation to enable correct use of the data.
Good repositories will assign a **unique permanent identifier**, display a clear **reuse license** and data **citation format**.

Trusted cross-disciplinary repositories:

- Zenodo
- Figshare
- Open Science Framework
- Harvard Dataverse
- EUDAT's B2Share

Discipline-specific repositories can be found via lists advised by publishers:

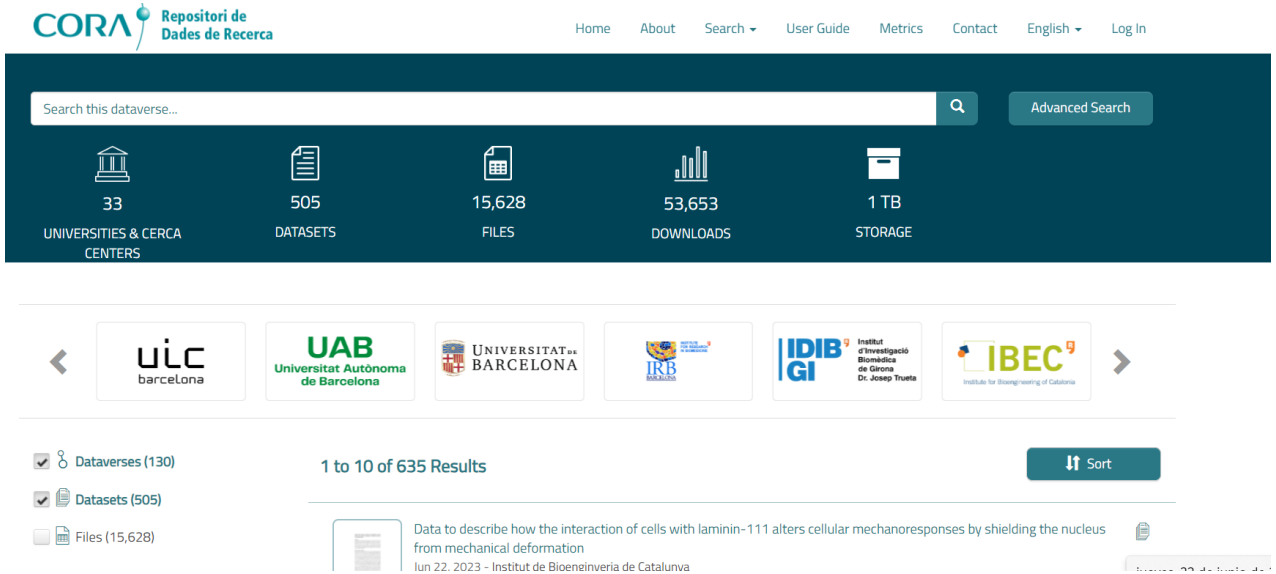
- Nature recommended data repositories
- PLoS One recommended repositories
- Springer Nature recommended repositories
- F1000Research approved repositories

or via re3data.org, a registry of over 2500 research data repositories.

Data management workshop - 7. Choosing a data repository

Institute for Bioengineering of Catalonia

CSUC Data repository: CORA. Repositori de Dades de Recerca https://dataverse.csuc.cat/ with IBEC's section.



The CORA. Repositori de dades de Recerca is a repository of open, curated and FAIR data that covers all academic disciplines. CORA. Repositori de dades de Recerca is a shared service provided by participating Catalan institutions (Universities and CERCA Research Centers). The repository is managed by the CSUC and technical infrastructure is based on the Dataverse application, developed by international developers and users led by Harvard University (https://dataverse.org).

8. Licenses and copyright

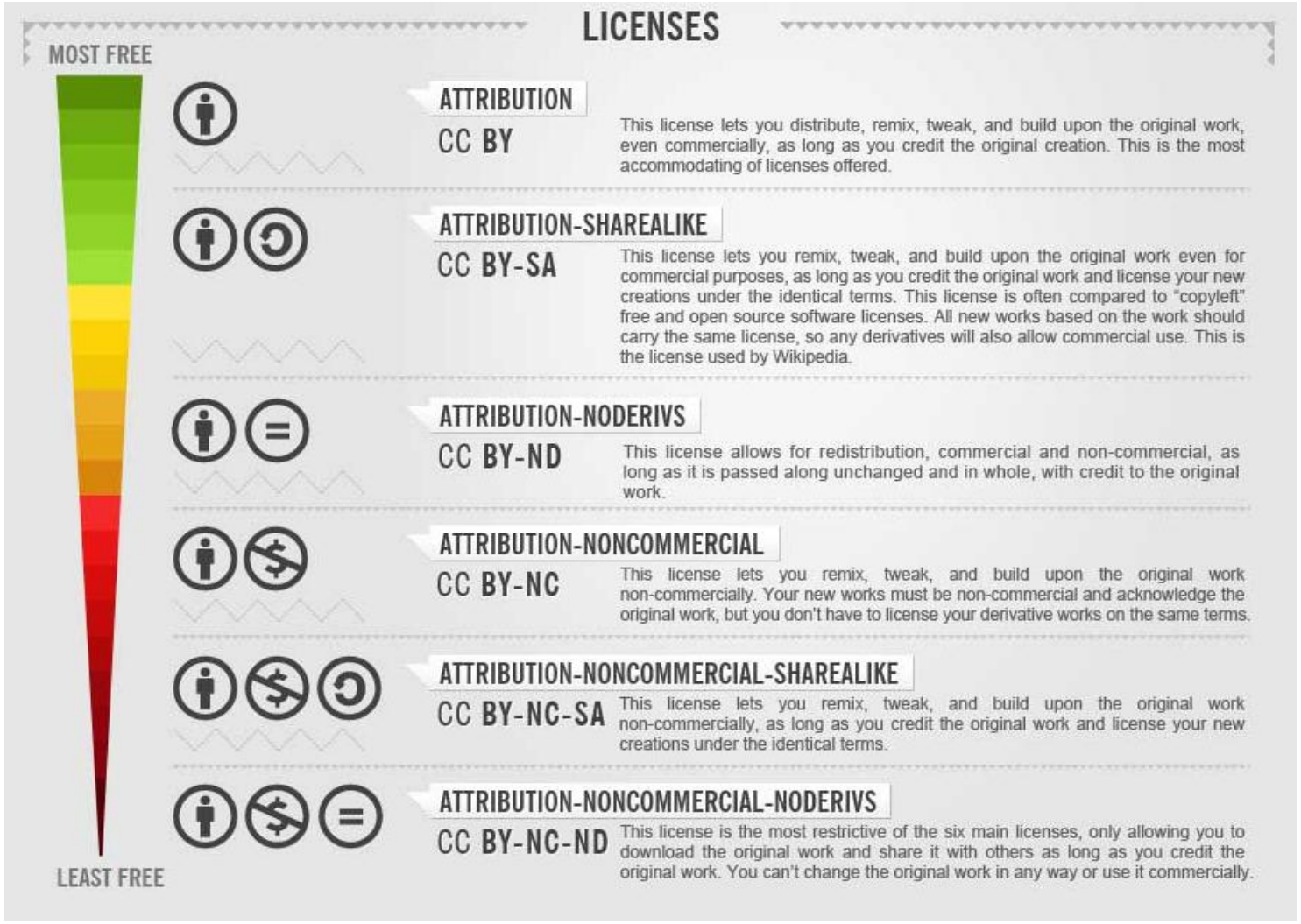
A piece of content or data is open if anyone is free to use, reuse, and redistribute it — subject only, at most, to the requirement to attribute and share-alike: conformant licenses: https://opendefinition.org/licenses/

- Open Data Commons is the home of a set of legal tools and licenses to help you publish, provide and use open data. https://opendatacommons.org/

- Copyright law gives creators certain kinds of control over their creative work. If people want to use copyrighted work, they often have to ask for permission from the creator. Creative Commons works within copyright law. It allows creators to grant permission to everyone in the world to use their work in certain ways. https://creativecommons.org/about/cclicenses/

Data management workshop - 8. Licenses and copyright

Institute for Bioengineering of Catalonia



(Creative Commons Infographic from: Technology Enhanced Learning Blog)

Wrap up

Institute for Bioengineering of Catalonia

Manage Data: data lifecycle

Plan: policies (institutional, funders' requirements)

Criteria: FAIR Principles

Storage: access, permissions, folders and files organisation

Share: licenses, repositories

https://ibecbarcelona.eu/