



# Open Science

for EURIZON fellowship for Ukrainian researchers

Paul Millar  & Melanie Nentwich   
Hamburg, 2024-07-16

HELMHOLTZ

**eurizon**  
European network  
for developing new horizons for RIs



# What is Open Science?

# Evolution, not revolution



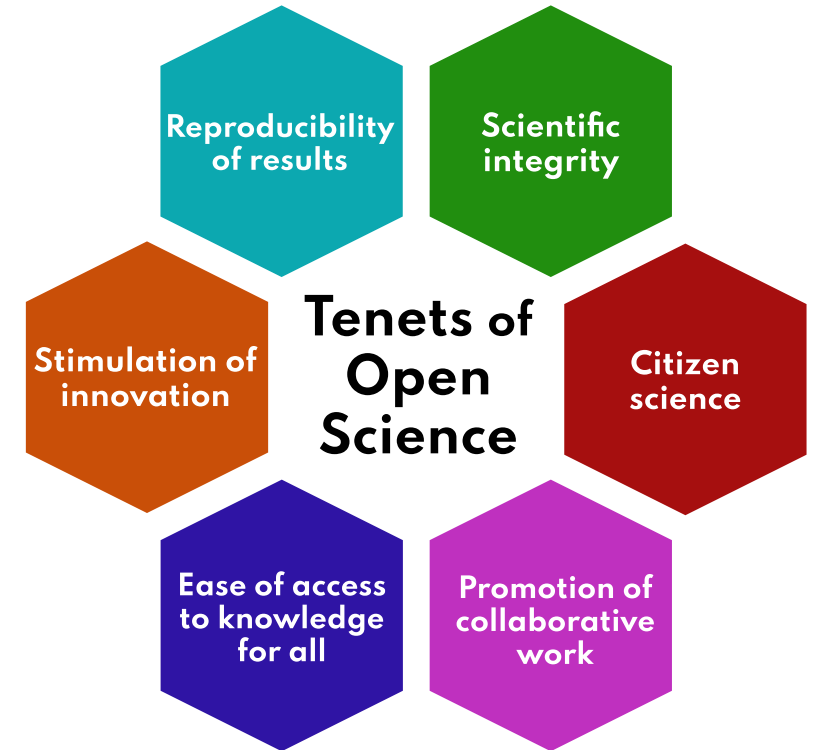
## Universal Declaration of Human Rights

### Article 27

Right to take part in cultural, artistic and scientific life.

You have the right to share the benefits of your community's culture, arts and sciences.

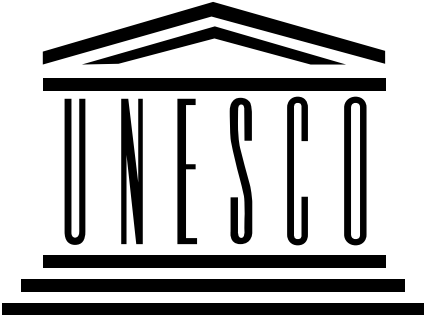
1 Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and to share in scientific advancement and its benefits. [...]



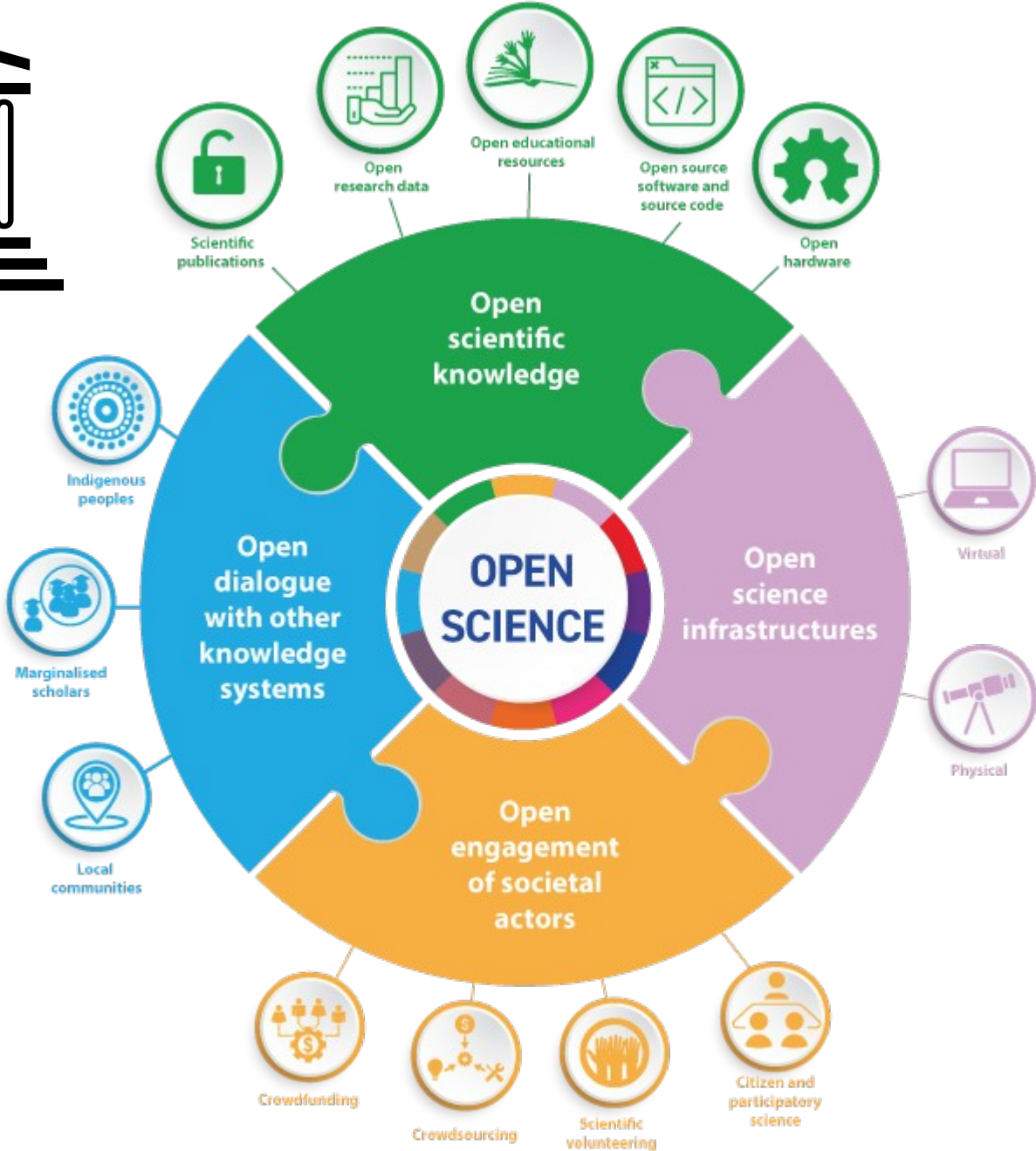
## Open Science

Academic freedom, Research integrity, Scientific excellence

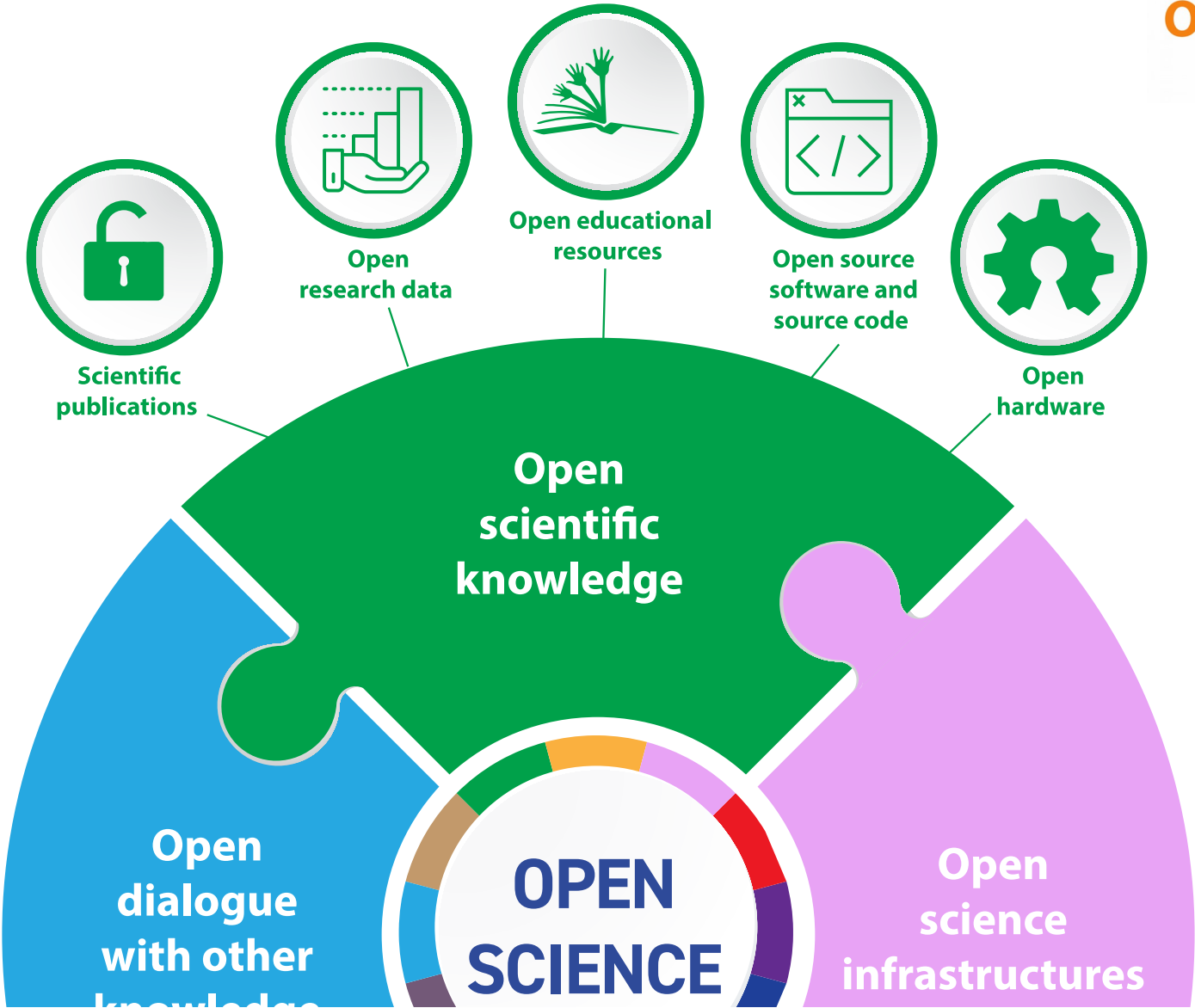
# Open Science pillars



<https://en.unesco.org/science-sustainable-future/open-science/recommendation>



# Open Scientific knowledge



OPEN  ACCESS

RCSB **PDB**  
PROTEIN DATA BANK

*NeXus*

zenodo

Photon and Neutron  
Trainin<sup>o</sup> $\gamma$

  
open source  
initiative<sup>®</sup>

**GPLv3**  
Free Software  
*Free as in Freedom*

  
ARDUINO



# Open Scientific infrastructures



**LEAPS** League of European Accelerator-based Photon Sources



**LENS** LEAGUE OF ADVANCED EUROPEAN NEUTRON SOURCES



**eosc**



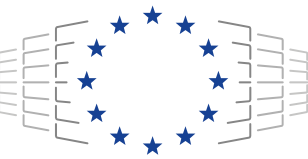
**OpenAIRE**



**ESI**



**PRACE**

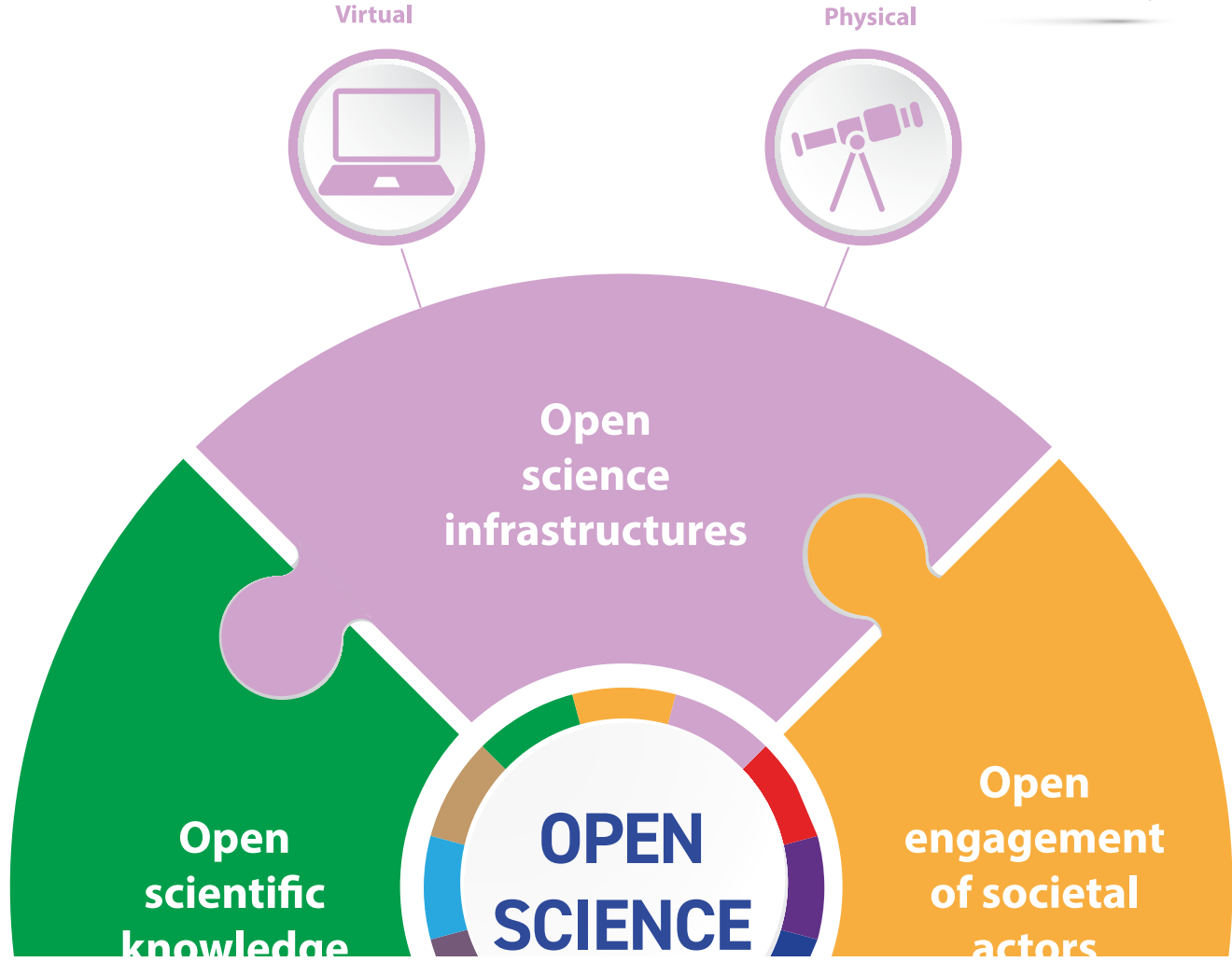


**EuroHPC** Joint Undertaking

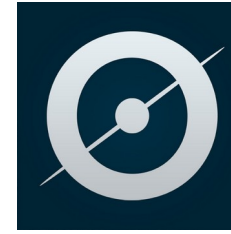
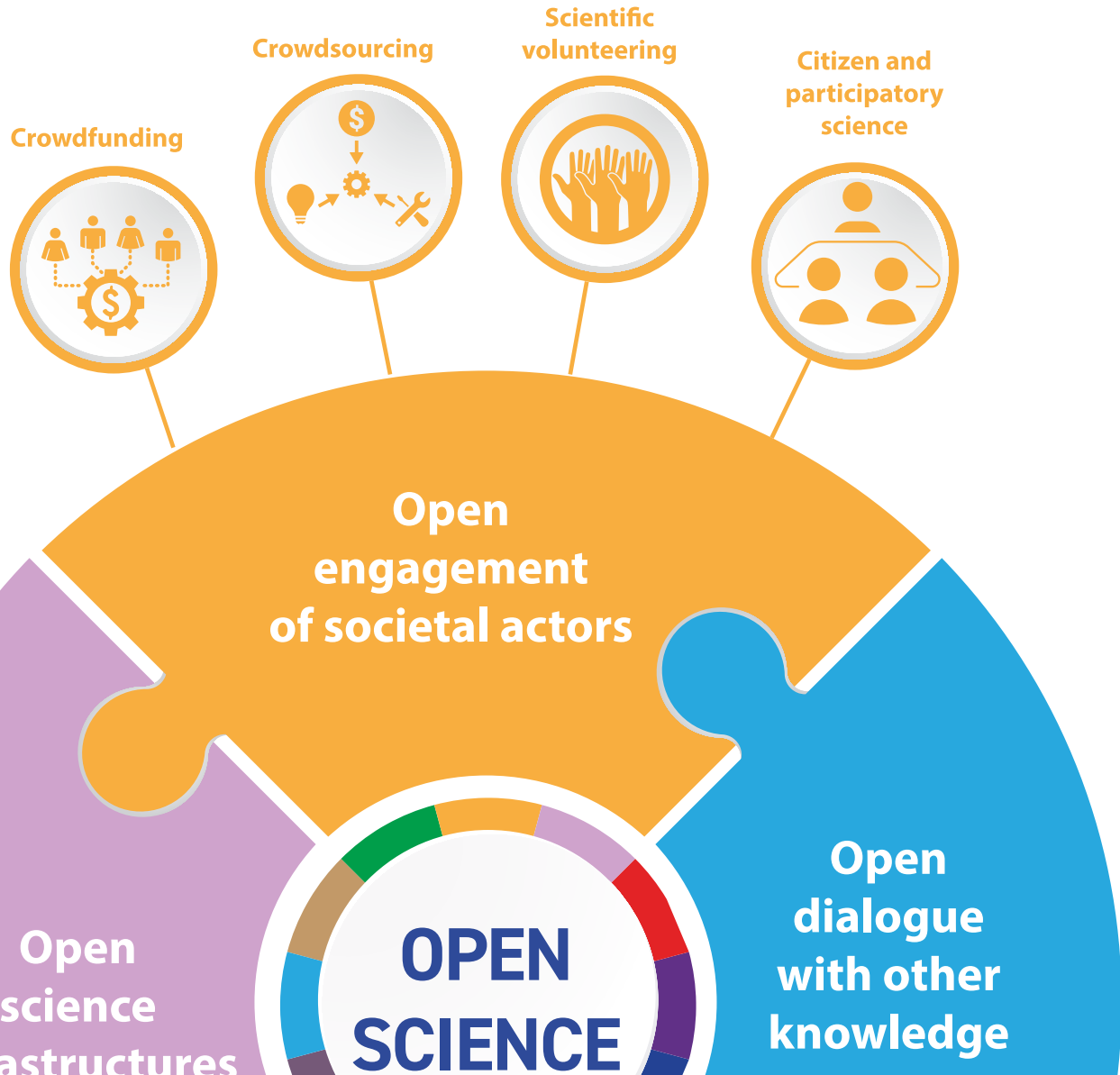
**EUDAT**



**EUDAT**



# Open Engagement of Societal actors



ZOONIVERSE



SETI@HOME

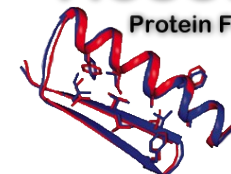
PUBPEER

BOINC



OpenStreetMap

Rosetta@home  
Protein Folding, Design, and Docking



FOLDING@HOME

# Open Dialogue with other knowledge systems

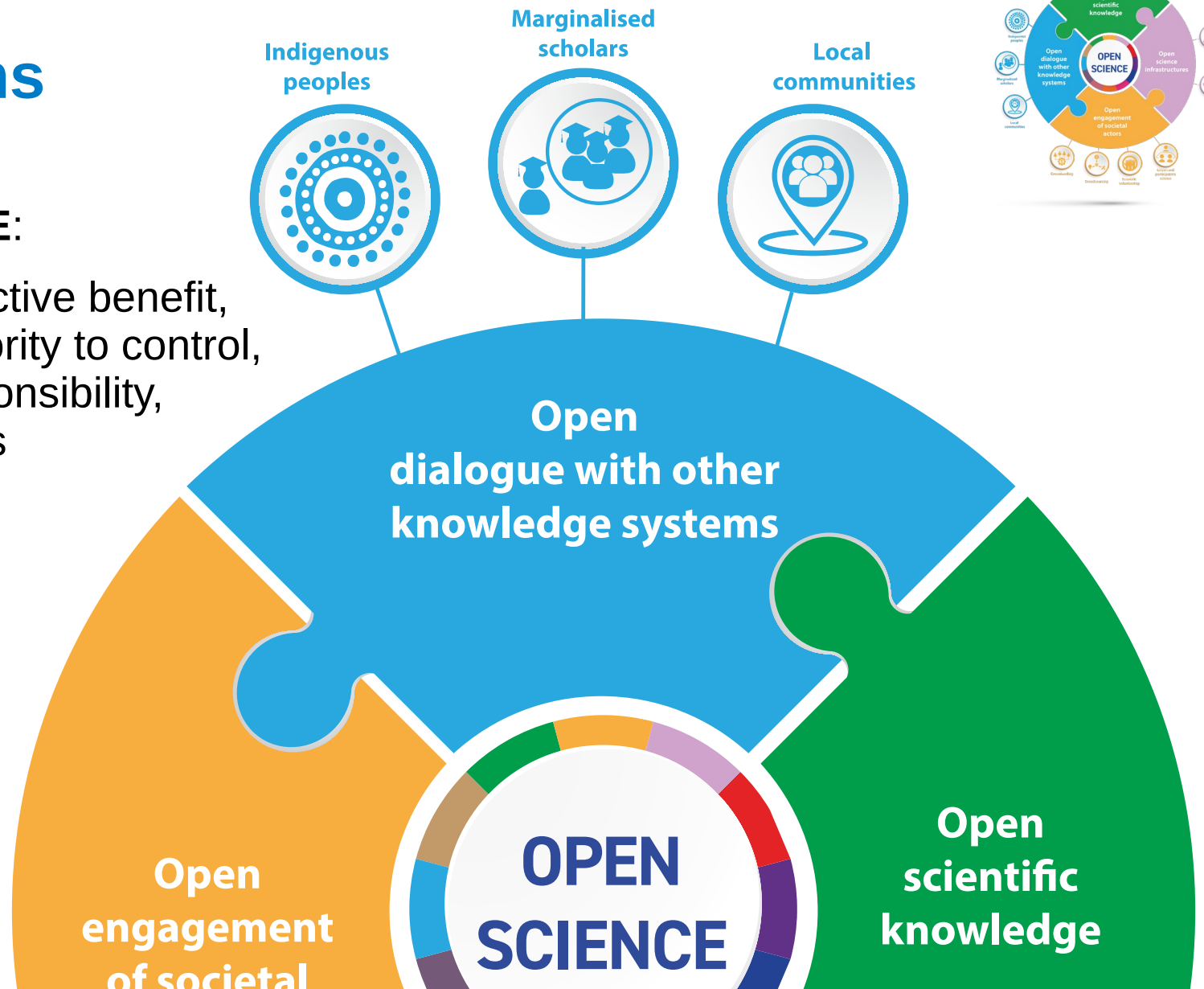
## Indigenous Peoples' Rights in Data

DATA FOR GOVERNANCE	GOVERNANCE OF DATA
<p><b>RIGHT TO SELF-DETERMINATION</b> the ability to organise and control data in relation to a collective identity</p>	<p><b>RIGHT TO GOVERN</b> the right to lead and collaborate in the development and implementation of protocols and in decisions about access to data</p>
<p><b>RIGHT TO POSSESS</b> the ability to exercise jurisdictional control over the ways that data flow/move/are queried</p>	<p><b>RIGHT TO DEFINE</b> the right to define lifeways of knowing and being including how they are represented in data</p>
<p><b>RIGHT TO USE</b> the ability of individuals and collectives to use data for their own purposes</p>	<p><b>RIGHT TO PRIVACY</b> the protection of collective identities and interests from undue attention, also including the possibility of requesting omission and/or erasure</p>
<p><b>RIGHT TO CONSENT</b> the expression of digital autonomy and the ability to assess risks and accept potential harms</p>	<p><b>RIGHT TO KNOW</b> the ability to track the storage, use, and reuse of the data and who has had access to them</p>
<p><b>RIGHT TO REFUSE</b> the right to say "no" to certain uses of data</p>	<p><b>RIGHT TO ASSOCIATION</b> the recognition of provenance and terms of attribution</p>
<p><b>RIGHT TO RECLAIM</b> the right to reclaim, retain, and preserve data, data labels, and data outputs that reflect Indigenous Peoples' identities, cultures, and relationships</p>	<p><b>RIGHT TO BENEFIT</b> the opportunity to benefit from the use of data and equitable benefit sharing from derivatives of data</p>

Global Indigenous Data alliance. (2023). "Indigenous Peoples' Rights in Data." The Global Indigenous Data Alliance. GIDA-global.org.  
DOI: 10.6084/m9.figshare.22138160



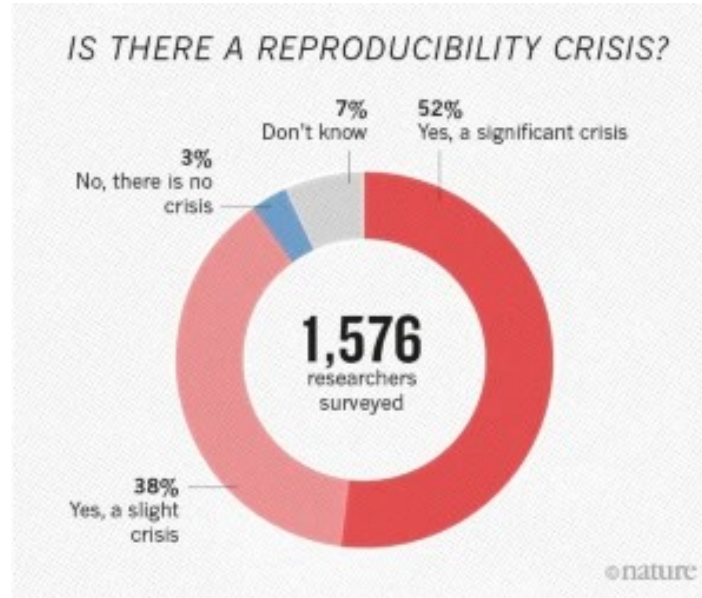
**CARE:**  
Collective benefit,  
Authority to control,  
Responsibility,  
Ethics





# Why is Open Science important?

# Why? Reproducibility & transparency



ELSEVIER



Springer

WILEY



IUCr  
International Union  
of Crystallography

**1,500 scientists lift the lid on reproducibility**

<https://doi.org/10.1038/533452a>



**Raw Data Letters**

“IUCrData has launched a new section for authors to describe their unprocessed or ‘raw’ diffraction images [...]”

# Why? Public funded works should be accessible



- A moral argument: public funded research should be in the hands of the people who paid for it.
  - Science should benefit everyone in society
  - Avoid science silos → Open Science
- Science budgets cannot be assumed:
  - Science targeting Societal Grand Challenges should be clearly identified.
  - Easier to justify money being spent if people can see it is having an effect.

## Article 27

Right to take part in cultural, artistic and **scientific life**.

You have the **right to share the benefits of your community's culture, arts and sciences**.

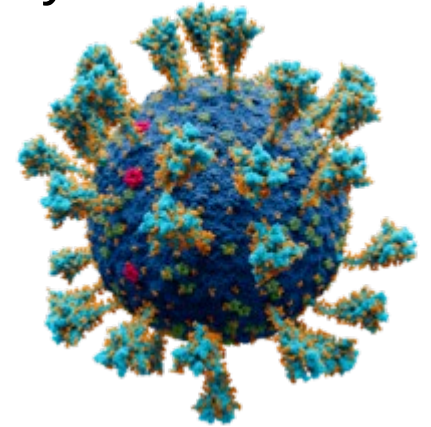
1 Everyone has the right freely to participate in the cultural life of the community, to enjoy the arts and **to share in scientific advancement and its benefits**.  
[...]

# Why? Greater impact of research

- The impact of research is a **hotly debated topic**:
  - Depends on how you measure the “impact”
- Journalistic publications tends to create barriers to access:
  - A sustainable model, but can we do better?
  - Preprint services; social media platforms (X, Instagram, Tiktok, ...)
- Based on traditional measures of impact, Open Access and Open Data have been proven to outperform traditional versions.
- Making research output (journals, data, software, ...) open allows it to reach new people.

# Why? Opening up opportunities for learning and cooperation

- Complexity of science. Areas of science that cannot be investigated by teams working in isolation.
  - A network of experts is needed.
  - How do we enable sharing of research output, without losing the acknowledgment?
  - Communication is key
- Provide learning opportunities.
  - Open Data as reference material
  - Open Data/Teaching as training material
  - Open Data as AI/ML training material
  - ...



# Why? Challenges of open science

- Handling dangerous information
- Avoiding public misunderstanding
- Avoid “low quality science”
- Risk of vendor lock-in.
- WEIRD-focus (Western, Educated, Industrialised, Rich and Democratic)
- Sustainability of research groups, when not derived from information scarcity.

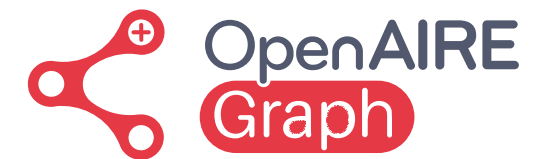
“as open as possible,  
as closed as necessary.”

# How is Open Science achieved in Europe?

# How? Incentives and rewards

## Providing a better “carrot”

- Traditional evaluation of academic output is based on journal publications.
  - Other output not considered
- Allows people to “game” the system:
  - “Paper mill”,
  - Self-citations, Group-citations, ...
- Contributing to a “Publish or perish” cultures
- ARRA agreement, opened for signatures 2022-09-28
- CoARA – Coalition for Advancing Research Assessment
  - Reforming how academic output is evaluated





# How? Legislative and regulatory aspects

- Various legal frameworks impacting Open Science
  - EU data law, Copyright, Software liability ...
- Looking for conflicts between European Research Area (ERA) Policy Agenda
  - Work to resolve problems.
- From another perspective: SciHub

From another perspective:

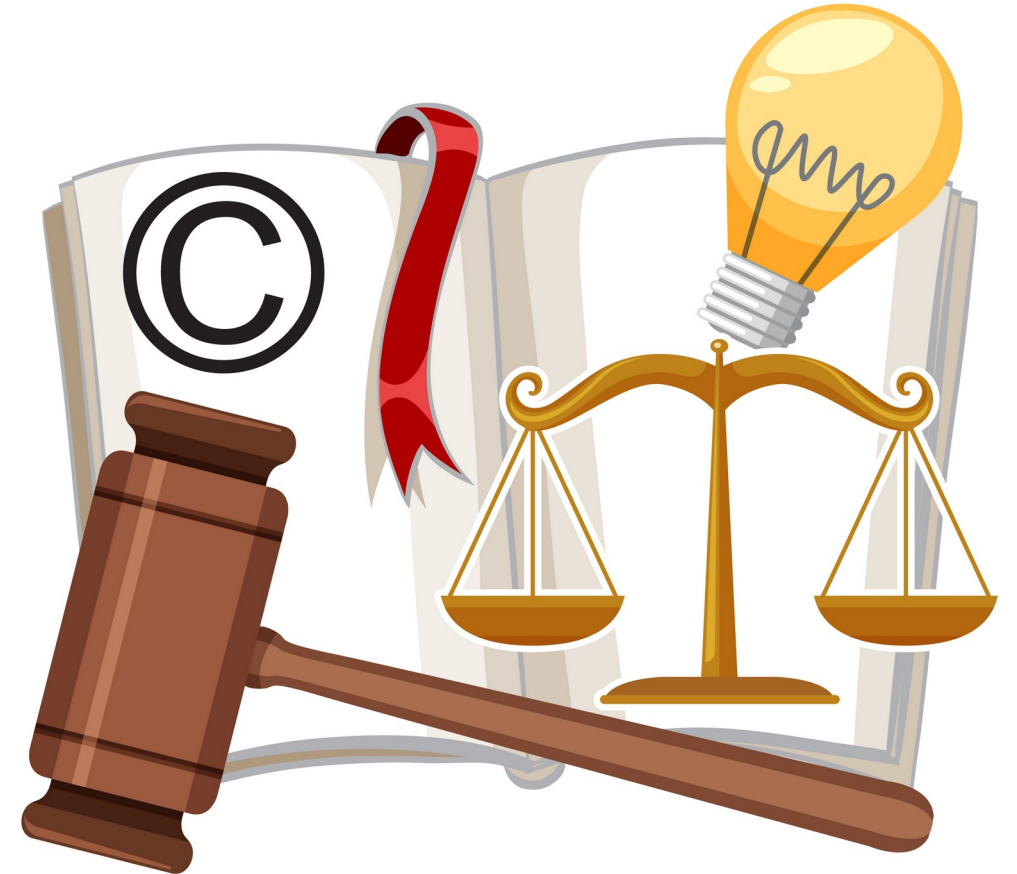


- Aaron Swartz
- Sci-Hub



**SCI-HUB**

...to remove all barriers in the way of science

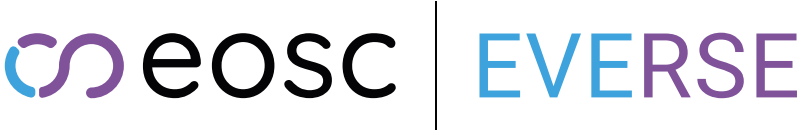


Designed by rawpixel.com / Freepik

# How? Framework Programmes



FP7	2007–2013	€53.5B
FP8 “Horizon 2020”	2014–2020	€77B
FP9 “Horizon Europe”	2021–2027	€95.5B



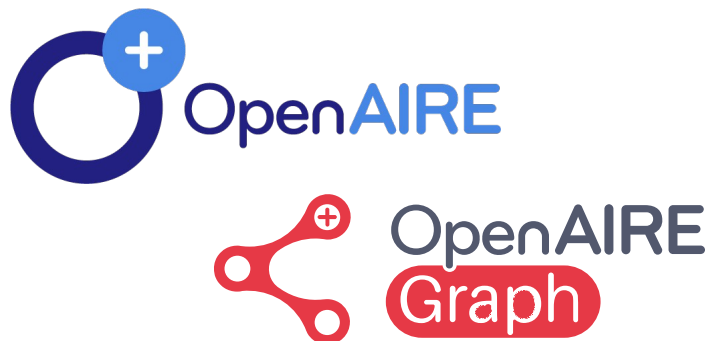
**Horizon Europe provisions on Open Science**  
– see talk

# How? Infrastructure and skills

## Practical support



Designed by rawpixel.com / Freepik



Open Access publishing platform for research funded by all EU Programmes

Support for skills and education for equitably practising Open Science and FAIR research data management

**Some real-world experience**

# FAIR data

... making Open Data actually useful.

For further information, see:

- <https://www.go-fair.org/fair-principles/>
- <https://www.fairsfair.eu/>

# FAIR: Findable

 **F**indable

 **A**ccessible

 **I**nteroperable

 **R**eusable

**F1.** Use globally unique and persistent identifiers.

**F2.** Data described with rich metadata

**F3.** Metadata clearly and explicitly include the identifier of the data they describe.

**F4.** Metadata registered or indexed in a searchable resource.



# FAIR: Accessible

 **F**indable

 **A**ccessible

 **I**nteroperable

 **R**eusable

**A1.** Metadata retrievable by the identifier using a standard communications protocol:

**A1.1.** Protocol is open, free, and universally implementable

**A1.2.** Protocol allows for an authentication and authorisation procedure, where necessary

**A2.** Metadata are accessible, even when the data no longer available.



# FAIR: Interoperable

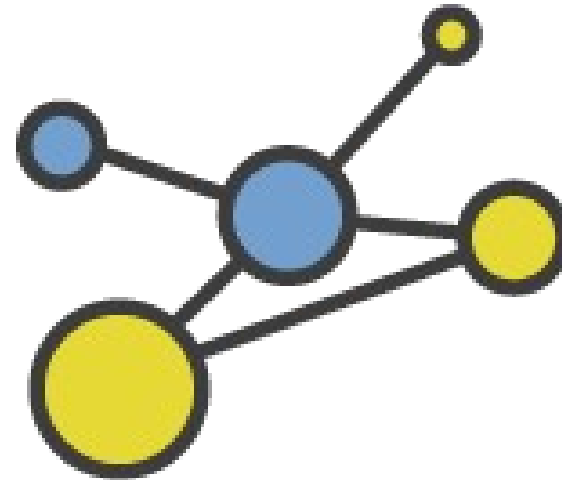
 **F**indable

 **A**ccessible

 **I**nteroperable

 **R**eusable

- I1.** Metadata: a formal, accessible, shared, and broadly applicable language for knowledge representation
- I2.** Metadata: use vocabularies that follow FAIR principals
- I3.** Include qualified references to other metadata.



*NeXus*



# FAIR: Reusable

 **F**indable

 **A**ccessible

 **I**nteroperable

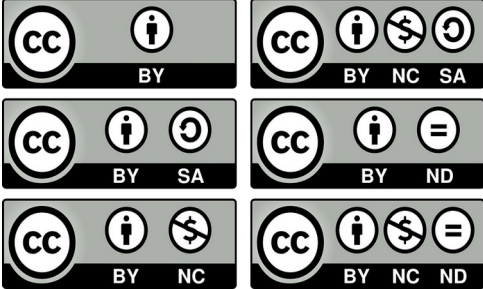
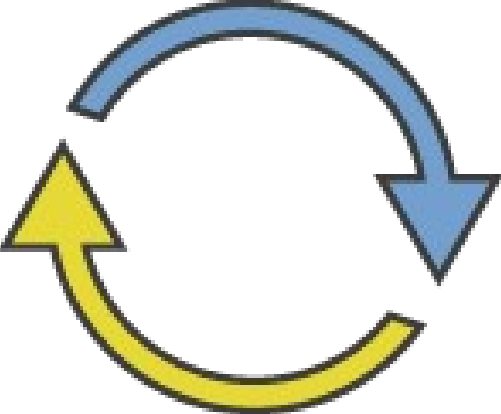
 **R**eusable

**R1.** Metadata are richly described with a plurality of accurate and relevant attributes.

**R1.1.** Metadata are released with a clear and accessible data usage license.

**R1.2.** Metadata are associated with detailed provenance.

**R1.3.** Metadata meet domain-relevant community standards.



# **Digital Object Identifiers (DOIs)**

**(and other PIDs)**

# PIDs: motivation

An identifier is something (usually text-based) that is associated with a specific item, and is not reused for any other item.

ISBN – identify a book

VIN – identify a motor vehicle

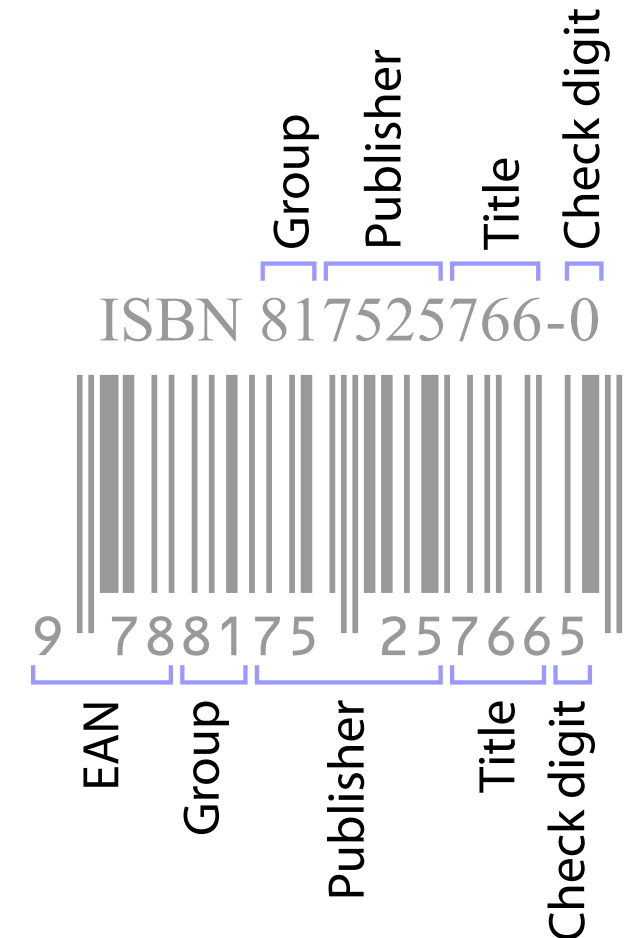
A persistent identifier (PID) is an identifier for a digital object that is persistent.

“**Persistence is purely a matter of service**” – it needs to be someone’s job to keep PIDs alive.

PID → some sustainable registry (e.g., business model)

PID will (usually) have some associated metadata.

**Best practice:** PID is opaque identifier with no embedded metadata.



# Handle / DOIs / Crossref / DataCite

HANDLE.NET

Solving the problem of “link rot”.

Handle: separate infrastructure from web, designed to provide a small amount of metadata (easy to maintain) and with other features to support robustness.

DOIs are examples of Handle IDs,  
Stronger guarantees.

DOI delegate operations to registration agencies  
(currently 12):

**Crossref** – academic journals

**DataCite** – research output (data, ...)

**OP** – EU related material.

...



A typical DOI: **10.5281/zenodo.11109625**

Compact representation: **doi:10.5281/zenodo.11109625**

Recommended version: **https://doi.org/10.5281/zenodo.11109625**

# DOI metadata

<b>Mandatory</b>	<b>Recommended</b>	<b>Optional</b>
Identifier	Subject	Language
Creator	Contributor	Alternate ID
Title	Date	Size
Publisher	Resource Type	Format
Publication year	Related identifier	Version
	Description	Rights
	GeoLocation	

<https://project-thor.readme.io/docs/datacite-metadata-schema>

# Other PIDs



<https://projects.tib.eu/pid-service/en/persistent-identifiers/persistent-identifiers-pids/>

# **Data Management Plans (DMPs)**

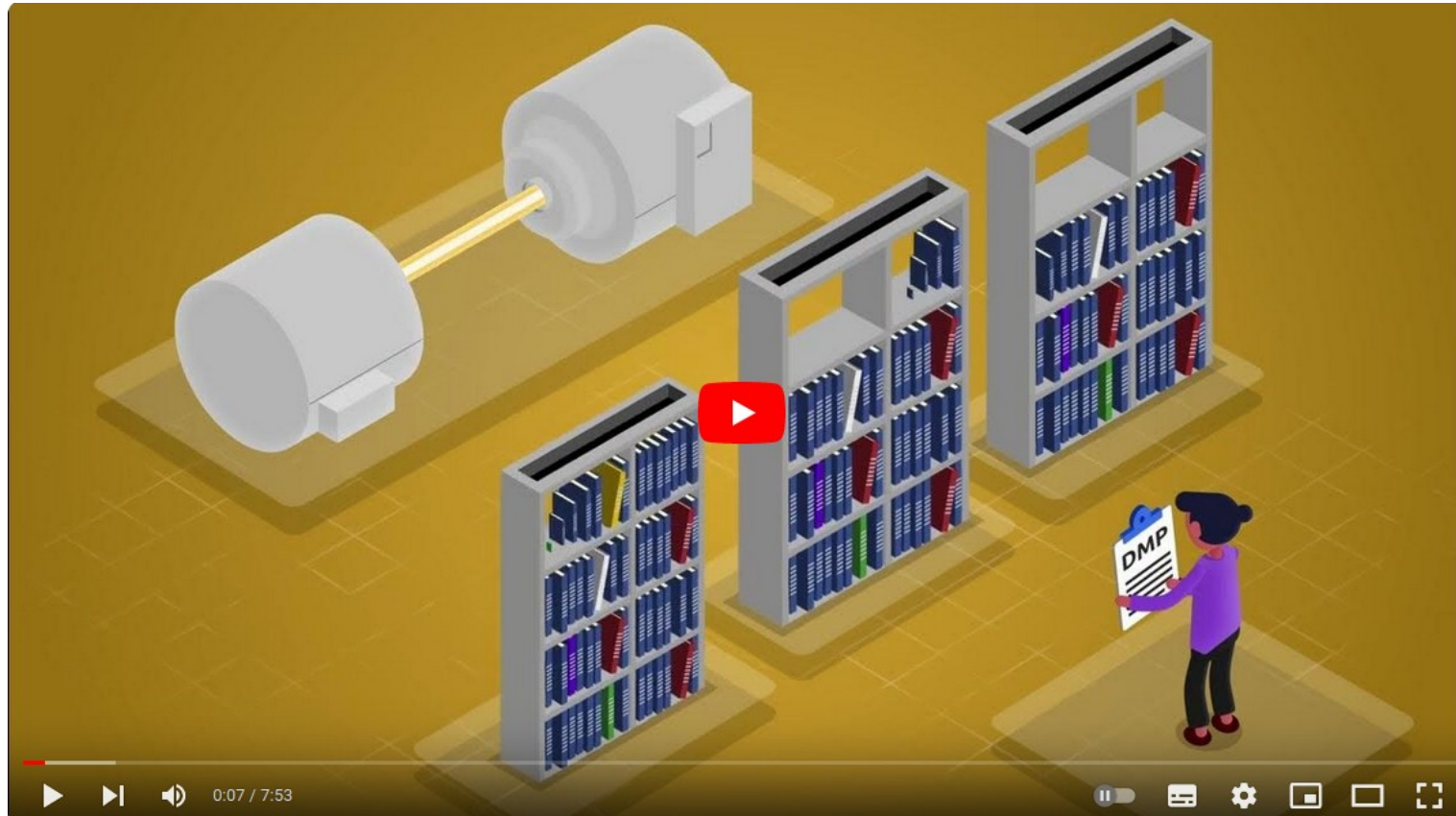
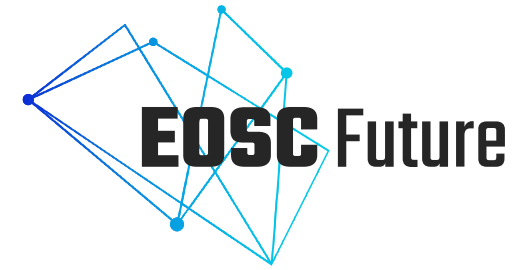
# What problem are we solving?

- Science inevitably requires taking measurements (== “data”)
  - Modern equipment takes ever increasing volumes of data (the “data deluge”)
  - Handling the data is increasingly non-trivial.
- Facilities may have certain expectations on how data will be handled. These should be agreed upon up front.
  - Data availability and locality
  - The process of making data FAIR and Open.
  - Embargo period: expectations.
  - Agreement on how long data will be kept.
  - ...
- A DMP is a (living) document, describing how data will be handled.
- EC has rolled in DMP requirements for EU projects (H2020, Horizon Europe)



# DMP, the movie

Output from EOSC-Future Project



**From beams to bytes: navigating data management for users of PaN facilities**

# Summary

# Key points

- We are currently in a transition towards Open Science:
  - Different fields and different countries are exploring this area.
  - Nobody has a complete solution yet.
- Open Science has many different aspects:
  - different fields have different emphases.
- Open Science has many benefits
  - Give you research more impact.
- Evaluation of research output is also evolving

**Thank you**



Research *and*  
Innovation *in*  
Computing



## Contact

Deutsches Elektronen-  
Synchrotron DESY

[www.desy.de](http://www.desy.de)

Paul Millar  & Melanie Nentwich 

RIC, IT

[Paul.Millar@desy.de](mailto:Paul.Millar@desy.de) [Melanie.Nentwich@desy.de](mailto:Melanie.Nentwich@desy.de)