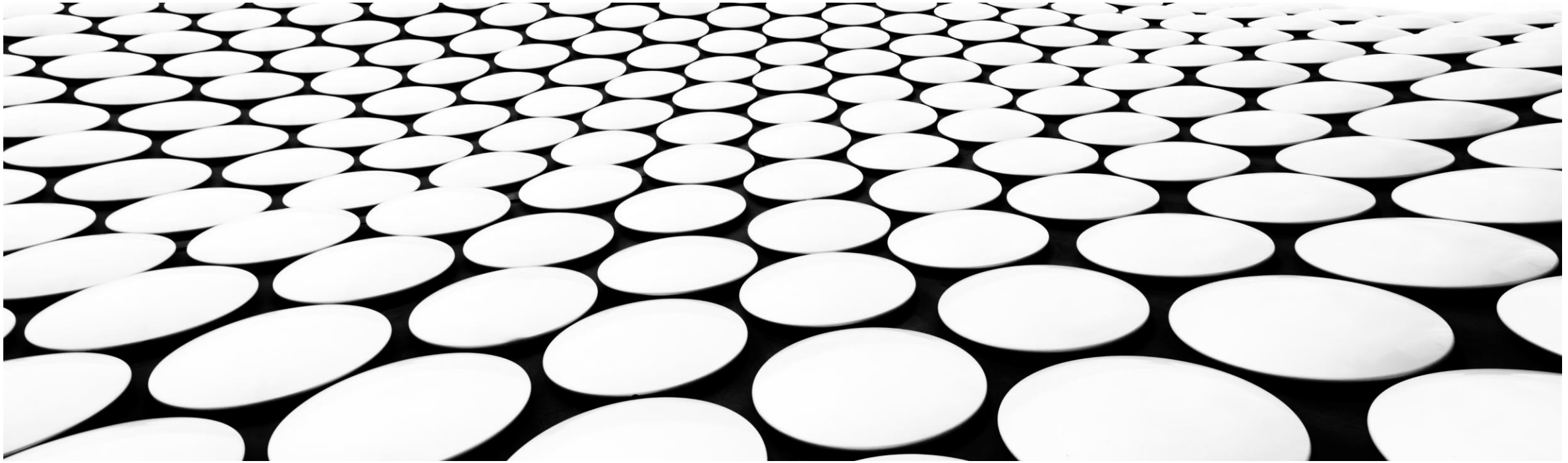
ELI FAIR DATA POLICY. GUIDELINES AND IMPLEMENTATION CHALLENGES



TEODOR IVANOAICA 07.07.2021

ELI ERIC SCIENTIFIC COMPUTING DATA MANAGEMENT

In-depth analysis of the community/ELI Sites tools and services

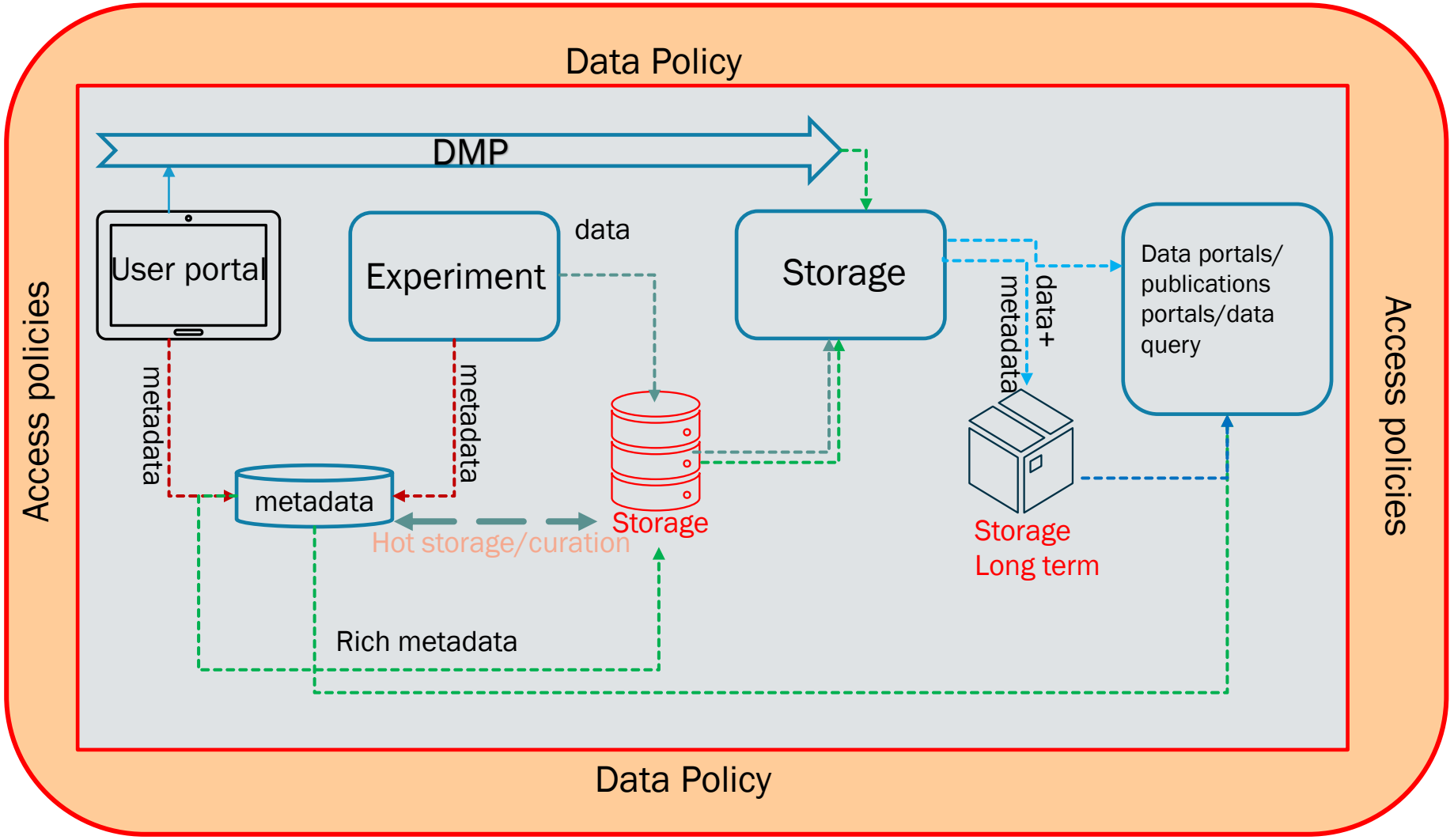
The core components of the Scientific Computing and Data Management System are developed for testing and fine-tuning

Initiate and define ELI Policies-> Data Policy

ELI, will be, by definition/design a FAIR Data Facility with a leading role in defining, integrating and developing standards for the HPLS community

ELI ERIC SCIENTIFIC COMPUTING DATA MANAGEMENT

In the process of designing systems and services for a User Facility, Users' Journey process **the process** driving the change.



Vision, Mission – the key management support:

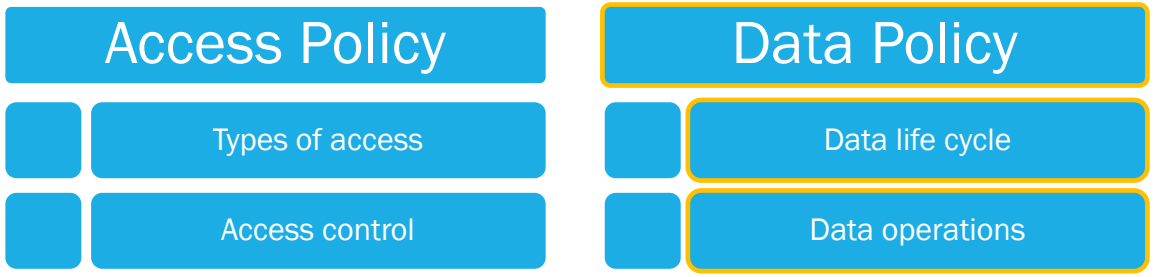
- ❑ *Access Policies defining the types of access and guidelines;*
- ❑ *Data Policies defining the concepts supporting the definition of the data lifecycle*

Next steps:

- Coordination to integrate tools and services;
- Design of the unified missing layers for the Data Management/Data Operations.

1. FAIR policies!

Policies supporting the strategies!



"Policies were identified as guides to thinking in decision-making. They assume that when decisions are made, these will fall within certain boundaries."

Koontz & O 'Donnel.

What they provide? What they will support?

- **What the policies are:** Access and Data Policies – the key set of guidelines and strategies driving ELI ERIC IT and Computing Strategy;
- **What the policies provide:** Vision and Mission for which particular tools/systems and services are developed;
- **What the policies will require for the implementation:** procedures, tools, services and integrated systems.

Scientific communities are adopting OPEN DATA and Open SCIENCE:

- It started with the OECD recommendation back in 2006
- The recommendation was updated in 2021(https://www.oecd.org/sti/recommendation-access-to-research-data-from-public-funding.htm)

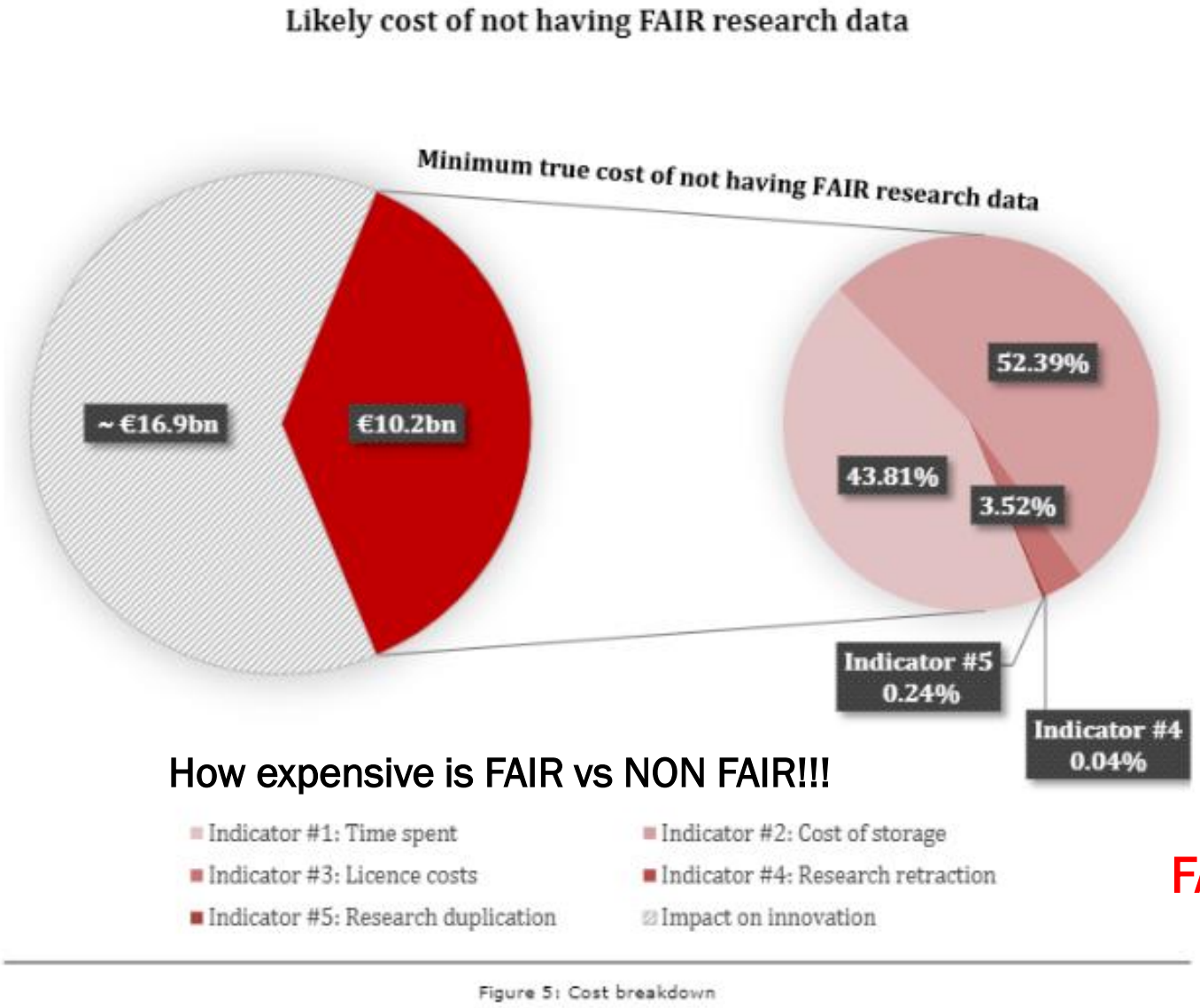


Figure 5: Cost breakdown

WHY FAIR?

FAIR Origins? FAIR originated from the current patchy data management practices in EU, which are not optimal yet.

The report estimates a total cost for not having FAIR data is calculate to a minimum of 10.2bn EURO/year!!!

FAIR is more than funds! FAIR makes good use of the funds to maximize scientific output!

https://op.europa.eu/en/publication-detail/-/publication/d375368c-1a0a-11e9-8d04-01aa75ed71a1

ELI ERIC SCIENTIFIC COMPUTING DATA MANAGEMENT

eli

Findable **A**ccesible **I**nteroperable **R**eusable

Data quality

Accountability

Efficiency

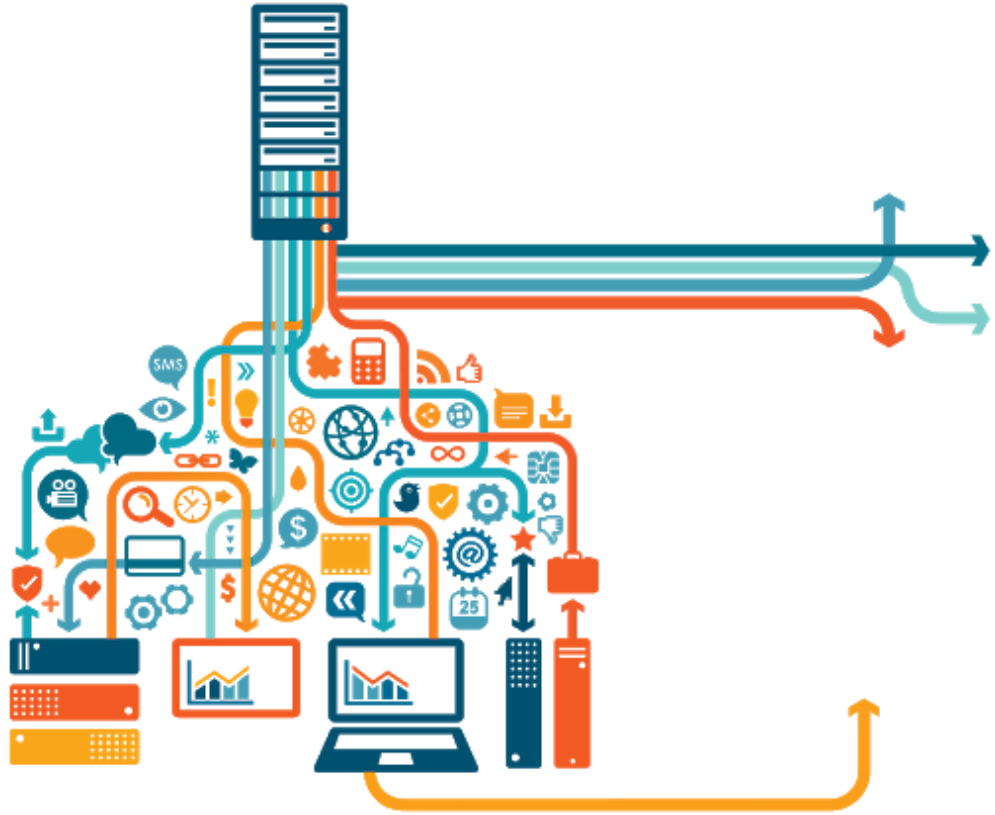
Collaboration

Visible and transparent

Reputation

Connects you to a scientific community!

FAIR Benefits



New Research!!
New tools and services!
Traceability/visibility!

How to start?!

ELI FAIR Facility & Open Data challenge! Basic Roadmap!

1. Defined a (FAIR) Data Policy – ELI Data Policy is currently submitted for approval

2. Based on the Data Policy, adopt a strategy, develop a Research Data Management plan and develop necessary procedures to support the implementation/adoption of the Data Policy and RDM

3. STANDARDIZATION:
   - Standardize Data Outputs
   - Standardize File types
   - Standardize File Formats
   - Standardize Software
   - Define Workflows
   - Adopt e-Logbooks
   - Develop Data Storage Strategy
   - Implement Data Archiving
   - Assign Data DOI

FOCUS ON USERS!!

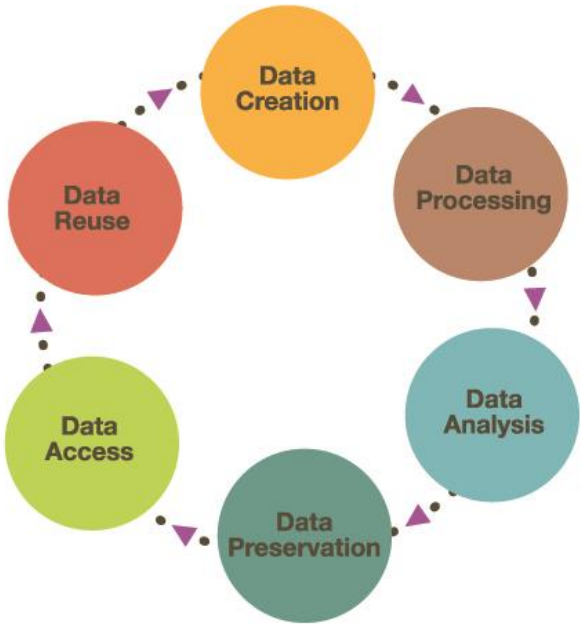
•CERN–open data policy for LHC (since 2020)
•EMBL–open access policy (since 2015)
•ESA–open data policy for most data (since 2010)
•ESO–open data policy (updated in 2016)
•ESRF–open data policy (since 2015)
•CERIC-ERIC–open data policy (since 2021)...
…..

This means users are trained/familiar with the standards, they have expectations and requirements!

4. Make your Data FAIR – benchmark your data and your tools, make sure you're not the only one understanding that data
5. There is more science inside your data, make it findable!

ELI ERIC SCIENTIFIC COMPUTING DATA MANAGEMENT

Research Data Management plan!



It all starts with a Data Policy!

What is next!?

Research Data Management plan! RDM

What RDM is?!

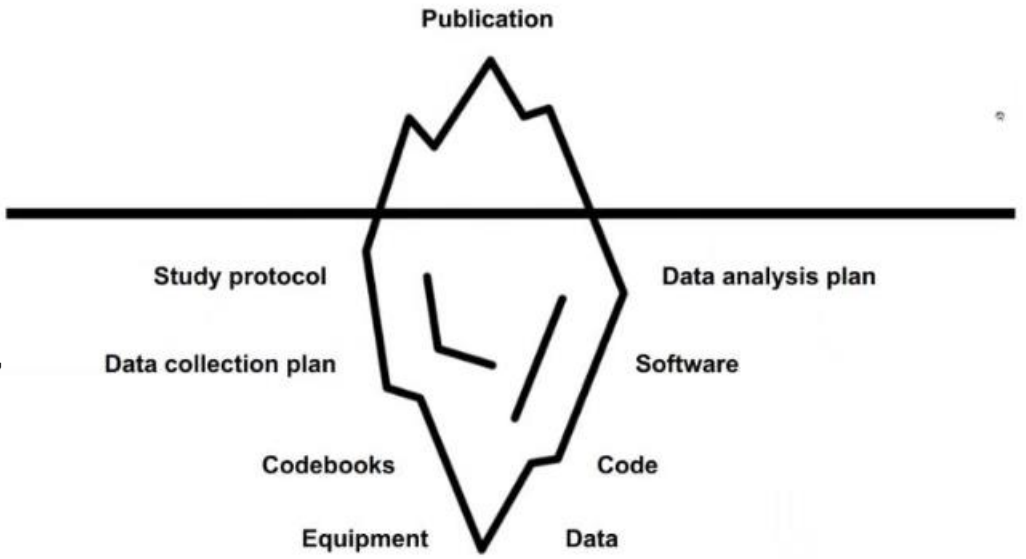
A collection of practices required to plan, collect, process, analyse, preserve, share, make data re-usable.
Further details: https://www.nature.com/articles/d41586-018-03071-1
ELI ERIC RDM will cover the full data lifecycle from the creation to Open Data! Better Data Management Plan results in BETTER SCIENCE!

Why Scientific communities are adopting OPEN DATA and Open SCIENCE?

- It started with the OECD recommendation back in 2006
- The recommendation was updated in 2021(https://www.oecd.org/sti/recommendation-access-to-research-data-from-public-funding.htm)

Implementation challenges!

1. **Use FAIR Guidelines (e.g. PaNOSC FAIR DP Guidelines)**
2. **Identify user requirements**, we need to know our users! We need to identify/create profiles!
   - What are the current identity used?
   - What is missing?
3. **Guide the users**, collect their requirements, guide them to provide a DMP, the DMP allows us to help our users.
   - DMP – Data Management Plan, the living document collecting all details/modifications of a data set during the entire experiment life-cycle
   - Collect expectations, reflect them in tools.
4. **Give fast access** to the data, allow users to see/interact/filter – users are **the facility QA team!**
   - data portal presenting the data to the users….when to get access? Can it be online/real-time?
5. **Give searchable data**, use FAIR, connect to the scientific community!
   - File catalogues, use metadata! Add meaningful metadata to accelerate science!
6. **Store the data/reuse the data/open data.**
   - FAIR standards and Open Data = reproducible results=> **more scientific outputs**
   - **Availability and Access-** It is easily available at a minimal cost. It should also be available in a usable form.
   - **Re-use and redistribution-** It is made available without any restriction on re-use and redistribution.
   - **Universal participation-** Anyone and everyone can access and/or re-use it.
7. **As a custodians of the data the facilities should prepare and train Data Stewards/Data Managers!**

Conclusions!

1. **FAIR is not just a standard, it becomes the community standard!**
   - Allows users to get more science from the same Data!
   - Allows the facility to preserve the time-to-compute unchanged, even if data rates are higher!
   - *Speeds up the research and adds EXTRA CITATIONS! Extra visibility!*
2. Agree on a checklist and schedule the development of the **Research Data Management** plan.
3. COLLEC THE DMP!!! Experiment DMPs are extremely useful!
4. Using the community know-how, tools and experience allows the *"new players"* to access the shared know-how and avoid blocking points!
5. **Standardize and integrate existing/build missing. The digital tools exist, treat the (meta)data seriously!**
6. ***"There is a lot more science than just text publications!"***