

Research Data Management

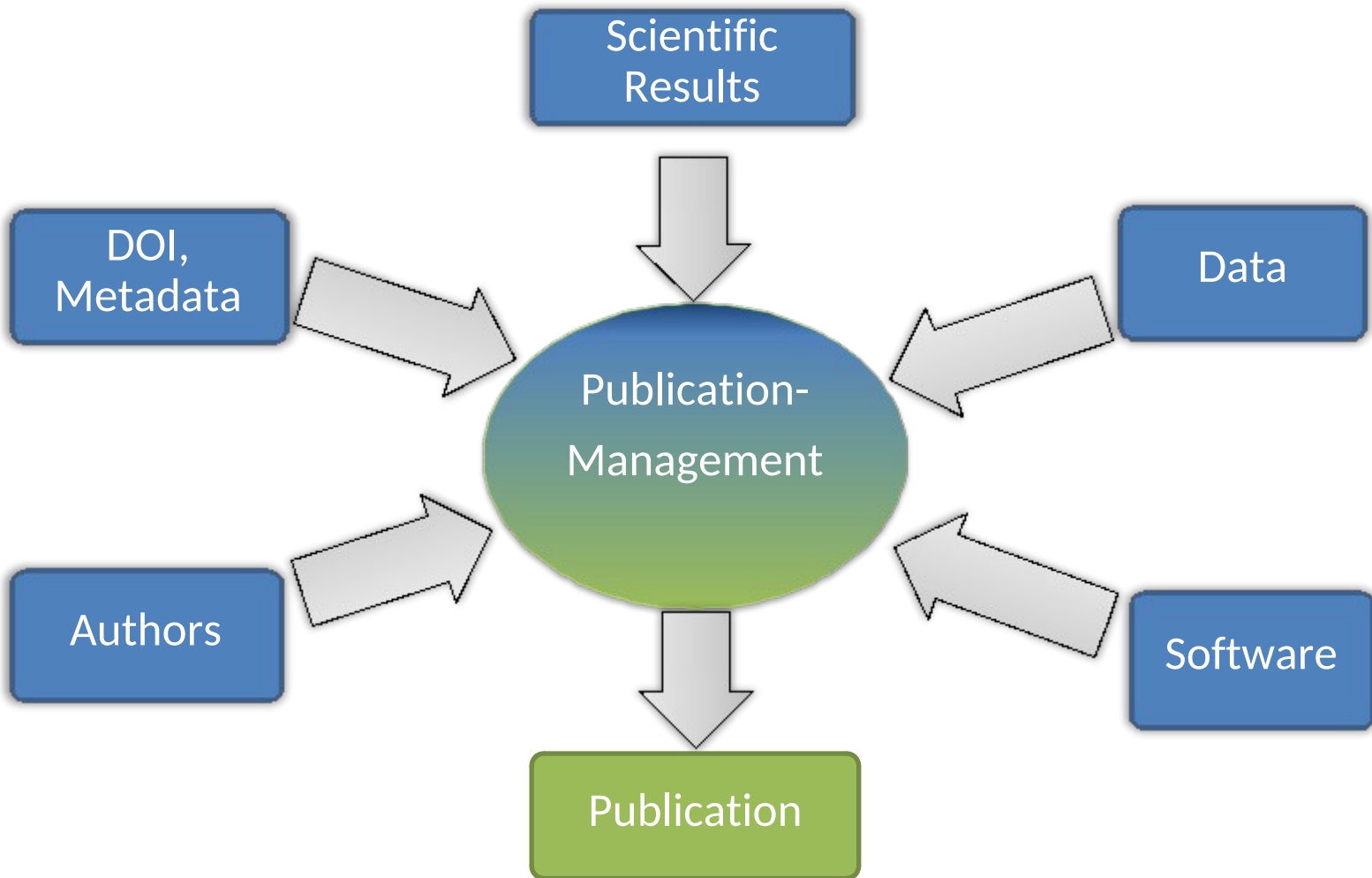
Towards a Data Management Plan
@ LaserLab Europe Workshop, Lisbon

Dec 7, 2018



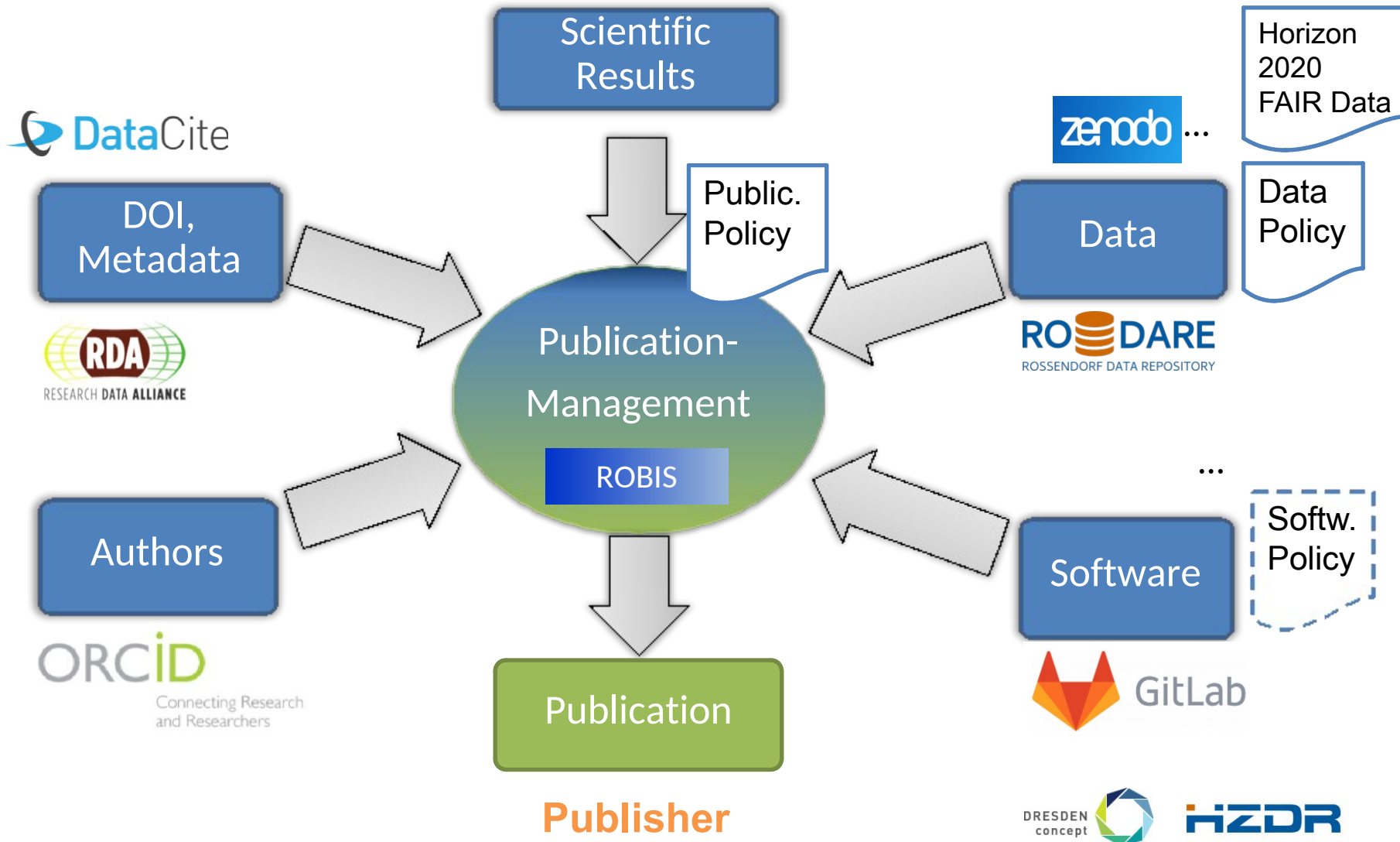
MOTIVATION

There is More to Your Paper/Thesis Than You Think!



MOTIVATION

There is More to Your Paper/Thesis Than You Think!



AGENDA

- Research Data and Research Data Management
- Political Implications or What funders want
- Interactive work on a DMP for LaserLab Europe

Slide from: Ralph Müller-Pfefferkorn, “Open Access and Research Data”, presented at EPC-Workshop on Sep 19, 2018 in Dresden under CC-BY 4.0



AGENDA

- **Research Data and Research Data Management**
- Political Implications or What funders want
- Interactive work on a DMP for LaserLab Europe

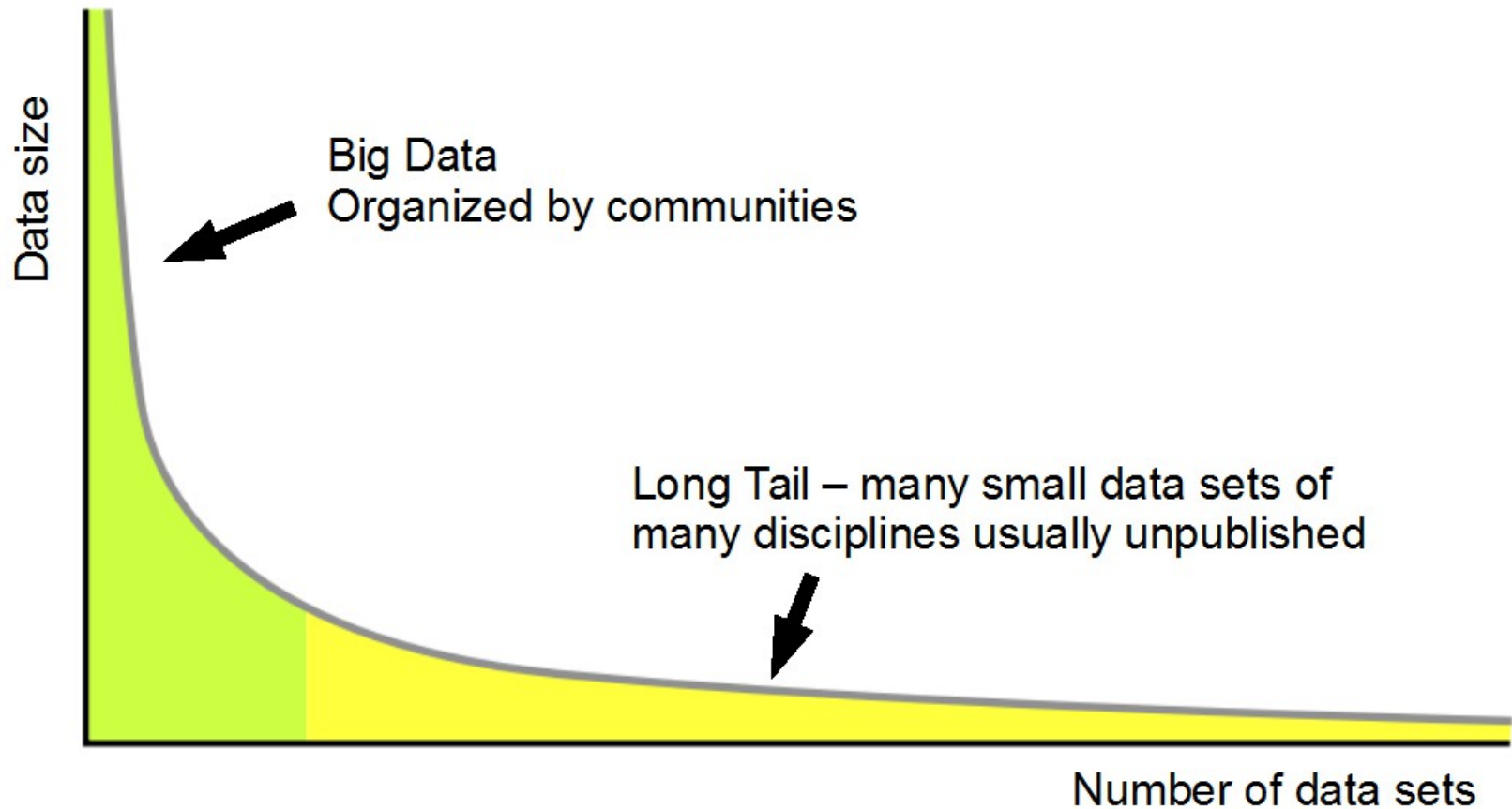
Slide from: Ralph Müller-Pfefferkorn, “Open Access and Research Data”, presented at EPC-Workshop on Sep 19, 2018 in Dresden under CC-BY 4.0

WHAT ARE RESEARCH DATA?

Type	Characteristics	Example
Observations	Data are collected in realtime, Usually irreplaceable	Sensor data Surveys
Experiments	Mostly produces in a laboratory environment Reproducible but expensive	Gene sequences Measurements
Simulations	Generated with models Models and metadata more important than output	Climate models Economic models
Derived data	Derived or compiled from other data Reproducible	text mining 3D models
References	Collections of usually smaller data sets Mostly published	Gene sequence database Primary text sources
Digitized data	Digitized versions of analog objects Reproducible as long as the original object is still available	Scriptures

Slide from: Ralph Müller-Pfefferkorn, "Open Access and Research Data",
presented at EPC-Workshop on Sep 19, 2018 in Dresden under CC-BY 4.0

BIG AND SMALL DATA



Slide from: Ralph Müller-Pfefferkorn, "Open Access and Research Data",
presented at EPC-Workshop on Sep 19, 2018 in Dresden under CC-BY 4.0

OPEN ACCESS TO RESEARCH DATA

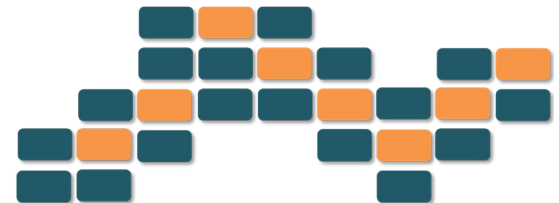
Open Data – Provide others access to research data!

- Validation and verification
- Transparency
- Prevent doubling work – use the data of other scientists
- Transdisciplinary use – new insights?
- Increased number of citations

FAIR Data – Make Data

- Findable
- Accessible
- Interoperable
- Re-usable

Find
Access
Interoperate
Re-use
Data

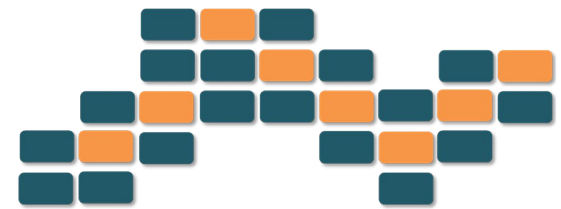


datafairport.org

Slide from: Ralph Müller-Pfefferkorn, “Open Access and Research Data”, presented at EPC-Workshop on Sep 19, 2018 in Dresden under CC-BY 4.0

FAIR DATA

Find
Access
Interoperate
Re-use
Data



datafairport.org

Findable

- F1. (meta)data are assigned a globally unique and eternally persistent identifier.
- F2. data are described with rich metadata.
- F3. (meta)data are registered or indexed in a searchable resource.
- F4. metadata specify the data identifier.

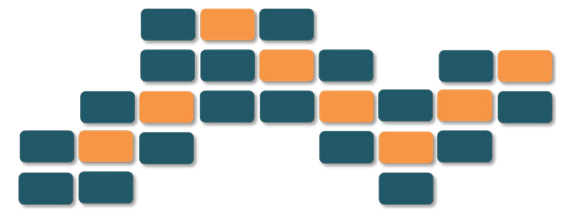
Accessible

- A1 (meta)data are retrievable by their identifier using a standardized communications protocol.
 - A1.1 the protocol is open, free, and universally implementable.
 - A1.2 the protocol allows for an authentication and authorization procedure, where necessary.
- A2 metadata are accessible, even when the data are no longer available.

Slide from: Ralph Müller-Pfefferkorn, “Open Access and Research Data”, presented at EPC-Workshop on Sep 19, 2018 in Dresden under CC-BY 4.0

FAIR DATA

Find
Access
Interoperate
Re-use
Data



datafairport.org

Interoperable

- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles.
- I3. (meta)data include qualified references to other (meta)data.

Re-usable

- R1. meta(data) have a plurality of accurate and relevant attributes.
 - R1.1. (meta)data are released with a clear and accessible data usage license.
 - R1.2. (meta)data are associated with their provenance.
 - R1.3. (meta)data meet domain-relevant community standards.

Prerequisite for Open Access to research data:

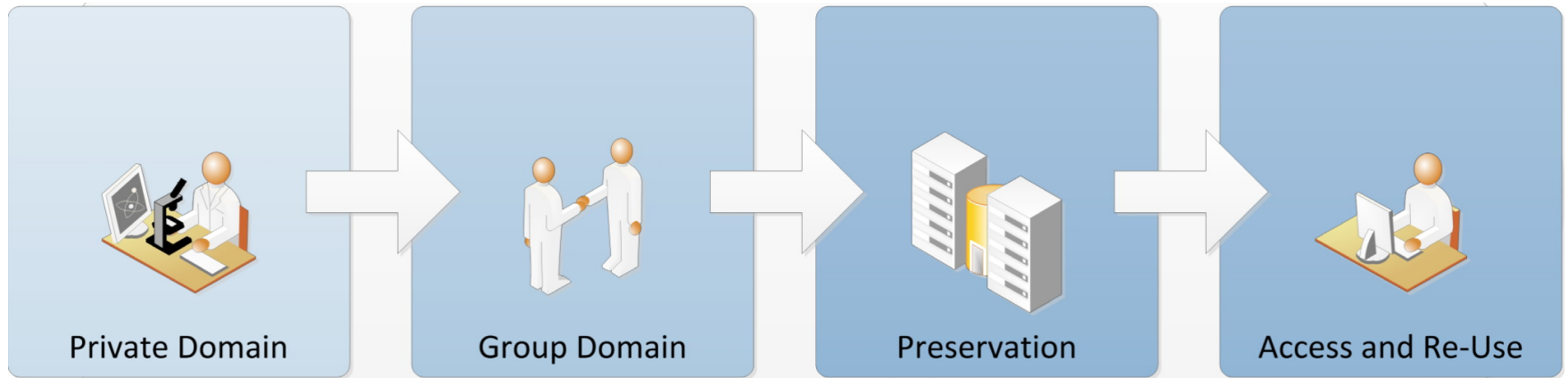
(Systematic) Research Data Management



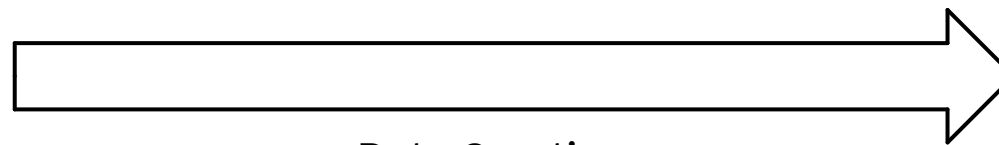
www.digitalbevaring.dk

Slide from: Ralph Müller-Pfefferkorn, "Open Access and Research Data",
presented at EPC-Workshop on Sep 19, 2018 in Dresden under CC-BY 4.0

DATA CONTINUUM



Simple
Documentation
(Metadata)



Data Curation

Rich
Documentation
(Metadata)

SUMMARY RESEARCH DATA (MANAGEMENT)

- Almost everything in research is research data as they come in many flavours.
- Research data management should cover the whole research data life cycle.
- FAIR data is for the benefit of all.

Agenda

- Research Data and Research Data Management
- **Political Implications or What funders want**
- Interactive work on a DMP for LaserLab Europe

Slide from: Ralph Müller-Pfefferkorn, “Open Access and Research Data”,
presented at EPC-Workshop on Sep 19, 2018 in Dresden under CC-BY 4.0

EUROPEAN UNION



“The European Commission’s vision is that information already paid for by the public purse should not be paid for again each time it is accessed or used, and that it should benefit European companies and citizens to the full.”

A vision for 2030: "Researchers and practitioners from any discipline are able to find, access and process the data they need. They can be confident in their ability to use and understand data, and they can evaluate the degree to which that data can be trusted." (EU 2010)

Starting 2017 data management plan requested for all H2020 projects.

Slide from: Ralph Müller-Pfeifferkorn, “Open Access and Research Data”, presented at EPC-Workshop on Sep 19, 2018 in Dresden under CC-BY 4.0

OPEN RESEARCH DATA IN HORIZON 2020

CHALLENGE
Wide access to scientific facts and knowledge helps researchers, innovation and the public. But our research data, and their research results:
offers better value for EU research funds

SOLUTION
Horizon 2020 strongly promotes open access to all scientific publications
Research data is open by default with opportunities to opt out

RESEARCH DATA - OPEN BY DEFAULT
FAIR DATA: Findable, Accessible, Interoperable, Reusable

HORIZON 2020 GRANTEES ARE REQUIRED
take measures to ensure open access to the data underlying their scientific publications
provide open access to any other research data of their choice
Horizon 2020 grants are awarded to those who agreed to share their datasets beyond submission

PROJECTS MUST HAVE
Data management plans fully aligned for funding
No repository imposed, deposit data where you want
The data of research will generate
how to ensure its curatorial, preservation and accessibility
what parts of that data will be open (and how)

AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY
Grantees have the right to opt-out, but need to take into account:
Top three reasons for opt-out:
Intellectual property rights
Privacy
might jeopardise project's main objective

The approach has been tested during a Horizon 2020 pilot action:
From 2015: of 431 H2020 projects, 105.4% opted to share data
From 2017: FAIR compliant Open Research Data Pilot expects to cover all areas of Horizon 2020 with the same rules

HOW IT WORKS
Let funding with open access
Let data stay open
FAIR results
Data is open by default
FAIR research
Data is open by default
FAIR results
Data is open by default

BE PART OF THE NEW ERA OF OPEN SCIENCE
reach more people, boost innovation
avoid duplicated efforts
generate data by better experiments
share your data by open access
improve your research by sharing knowledge
improve your research by sharing knowledge

Bioinformatics Institute
Benefits identified by the European Commission include: to improve scientific information from data available to the global life science community
€1.3 billion per year
improvements: more than 200,000 research papers published per year

ec.europa.eu/research/openscience
openaire.eu



DATA MANAGEMENT PLAN (DMP)

Clarifies all questions regarding data management

- Documents data management during and after a project
- Formal Document

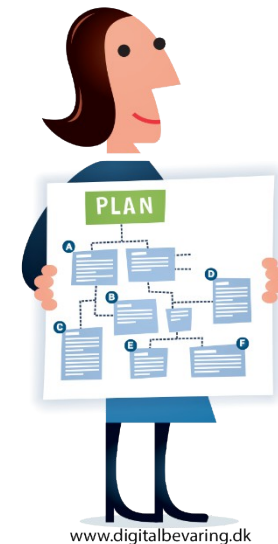
Is created before project start.

- Start early! Write the plan collaboratively!
- Be realistic!

Is regularly updated during project runtime.

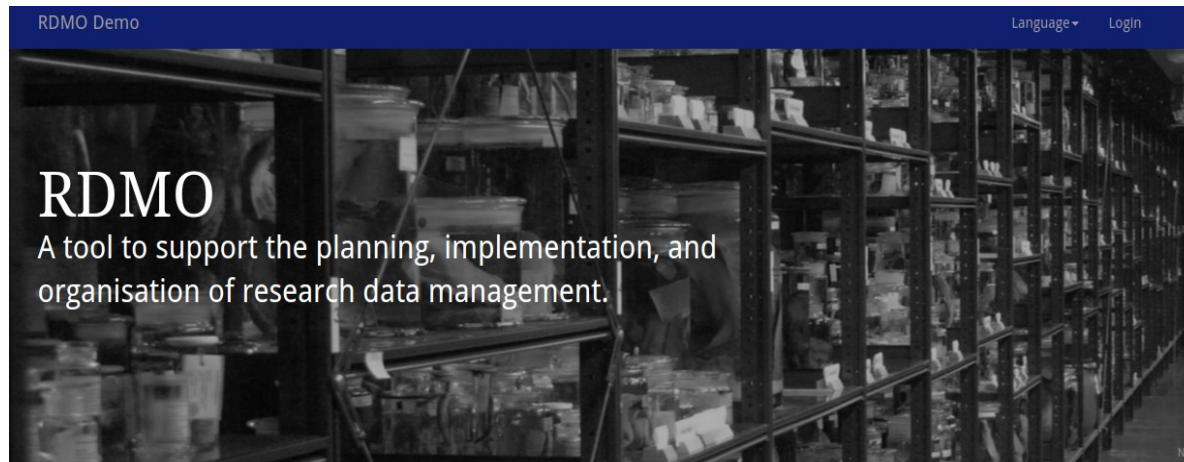
DMP - WHAT SHOULD BE ADDRESSED?

- Administrative information (project, data originators, other contributors ...)
- Data set description
 - Nature of data, sizes, formats, quality control, data privacy, anonymisation ...
- Metadata and metadata standards information
- Plans for data sharing and access
- Archiving and backup of data
- Data set description
 - Which data need to be kept after the project ends?
 - Data formats, sizes ...
- Responsibilities
- Costs – some funders provide funding for data management effort!



DMP - WHERE TO START?

- RDMO: (<https://rdmo.aip.de>)



Welcome to RDMO

The aim of the RDMO project is to deliver a web application to assist structured planning, implementation and administration of the data in a scientific project. Additionally, the gathered information can be cast into textual forms suitable for funding agencies requirements or for reports.

This is a prototype of the software, for demonstration purposes.

For more information about the project visit rdmorganiser.github.io.

Login

Username

Password

Remember Me

SUMMARY POLITICS AND FUNDERS

- The researcher is responsible for research data management.
- More and more funders require research data management.
- A data management plan is a living document in a project, that clarifies and documents all questions regarding research data management.
- DMPs are required by funding agencies.

AGENDA

- Research Data and Research Data Management
- Political Implications or What funders want
- **Interactive work on a DMP for LaserLab Europe**

Slide from: Ralph Müller-Pfefferkorn, “Open Access and Research Data”,
presented at EPC-Workshop on Sep 19, 2018 in Dresden under CC-BY 4.0