Research Data Management
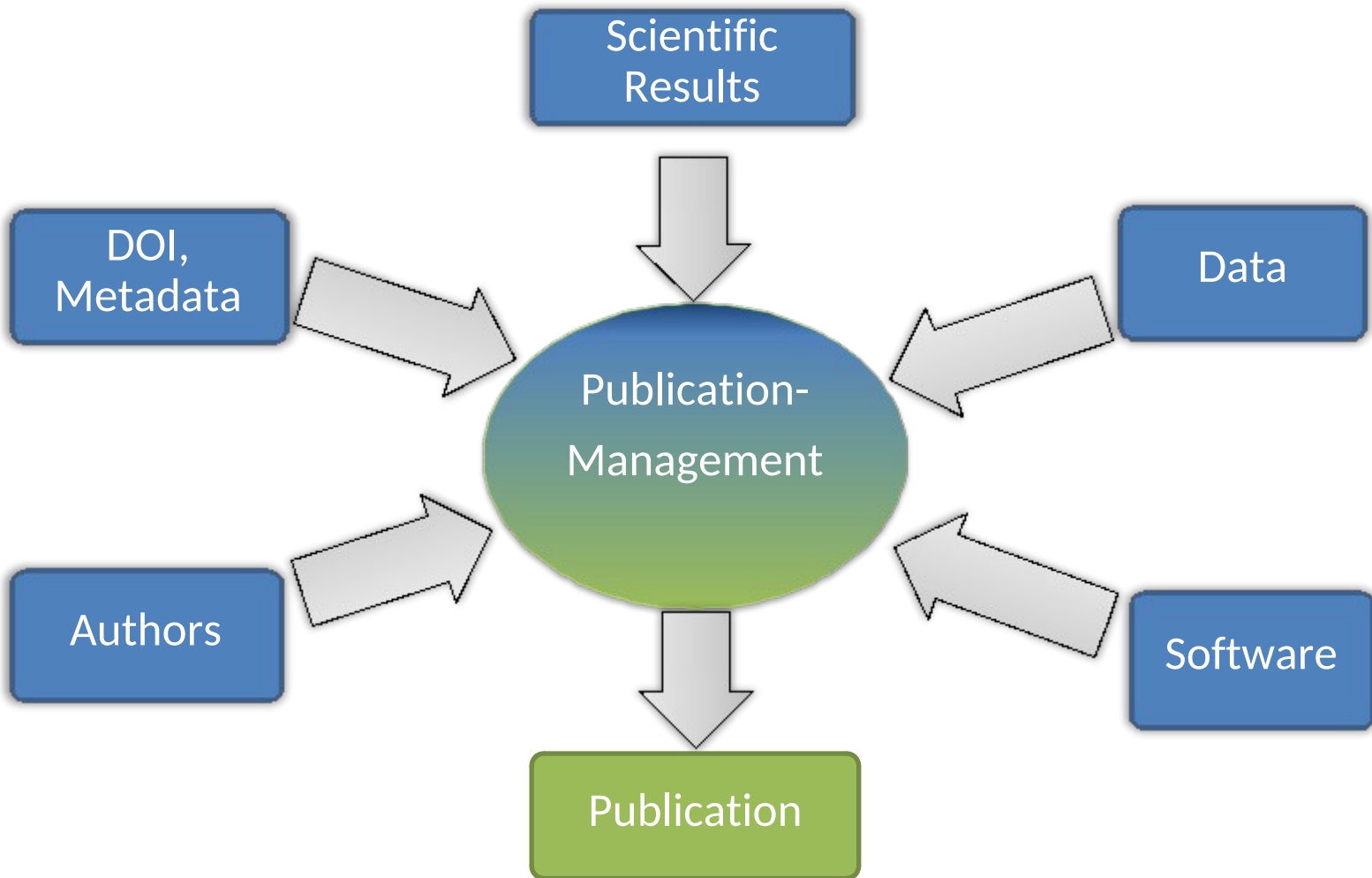
Towards a Data Management Plan
@ LaserLab Europe Workshop, Lisbon

Dec 7, 2018
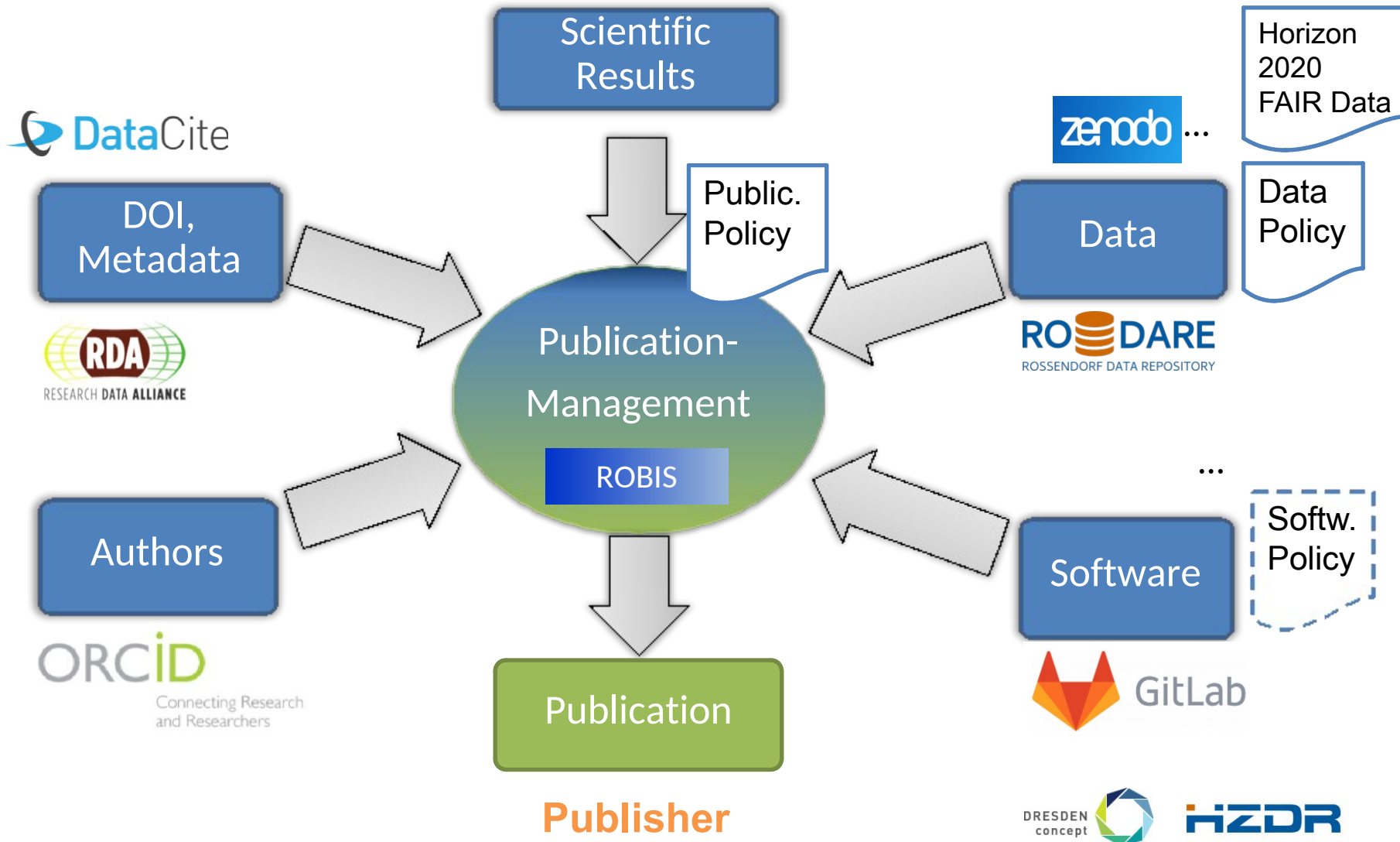
MOTIVATION
There is More to Your Paper/Thesis Than You Think!

Scientific Results

DOI, Metadata

Data

Publication-Management

Authors

Software

Publication

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

Member of the Helmholtz Association

Dr. Guido Juckeland I Computational Science Group (FWCC) I www.hzdr.de/fwcc

AGENDA

- **Research Data and Research Data Management**

- Political Implications or What funders want

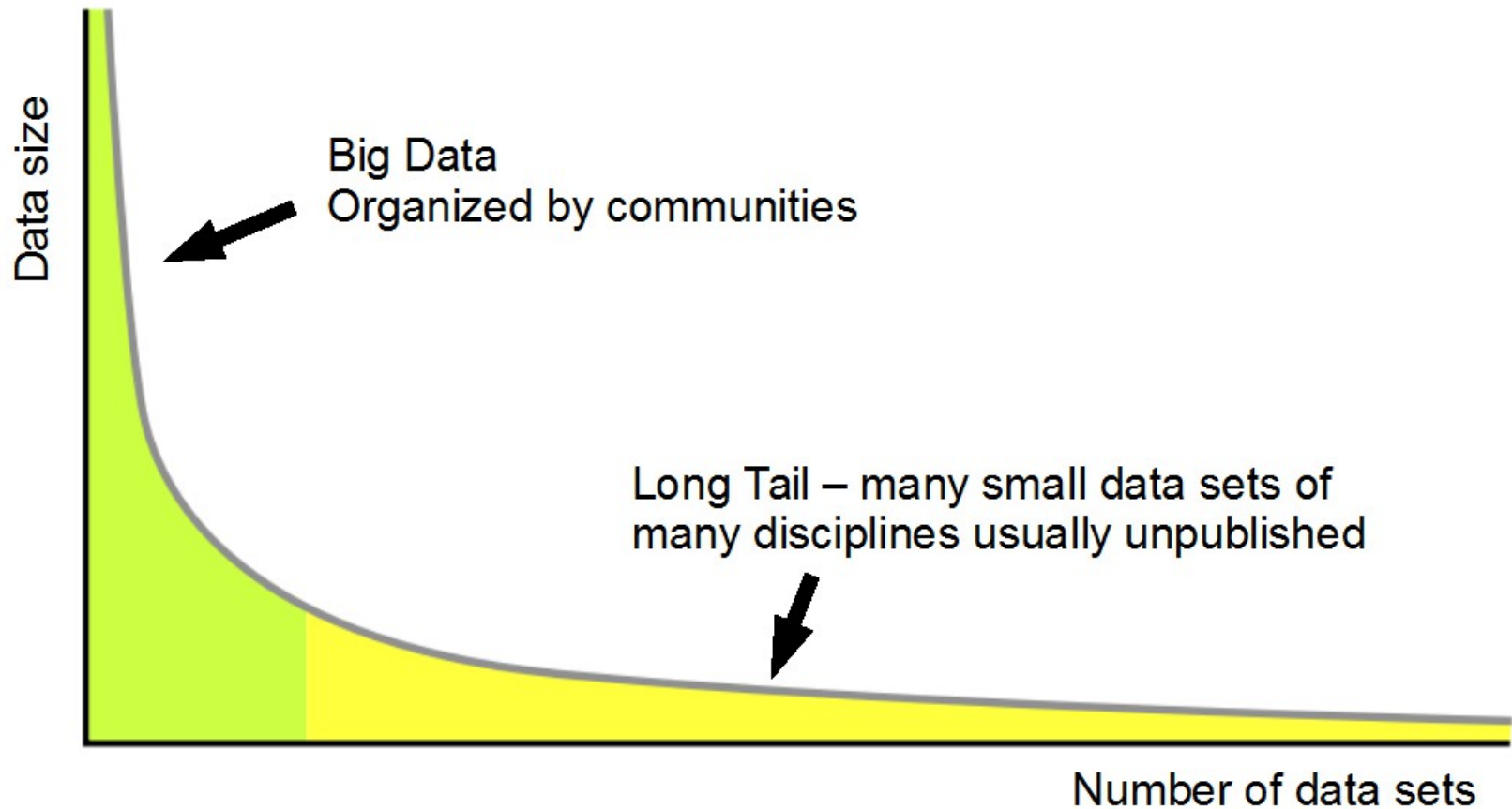- Interactive work on a DMP for LaserLab Europe

Member of the Helmholtz Association

Dr. Guido Juckeland I Computational Science Group (FWCC) I www.hzdr.de/fwcc

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 Reproducable 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 Reproducable	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 Reproducable 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

Dr. Guido Juckeland I Computational Science Group (FWCC) I www.hzdr.de/fwcc

BIG AND SMALL DATA

Member of the Helmholtz Association

Dr. Guido Juckeland I Computational Science Group (FWCC) I www.hzdr.de/fwcc
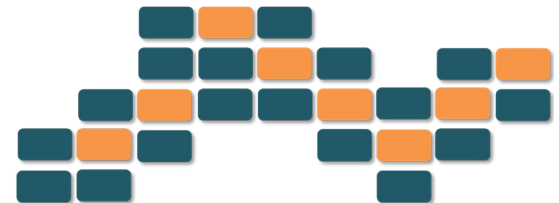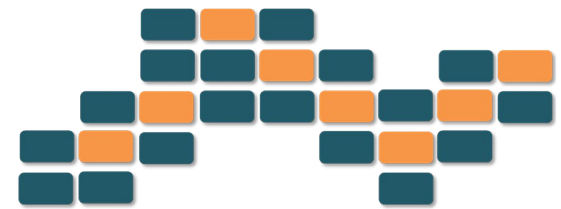
OPEN ACCESS TO RESEARCH DATA

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

- **F**indable
- **A**ccessible
- **I**nteroperable
- **R**e-usable

Find
Access
Interoperate
Re-use
Data

datafairport.org

DRESDEN concept

HZDR

Member of the Helmholtz Association

Dr. Guido Juckeland I Computational Science Group (FWCC) I www.hzdr.de/fwcc

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.
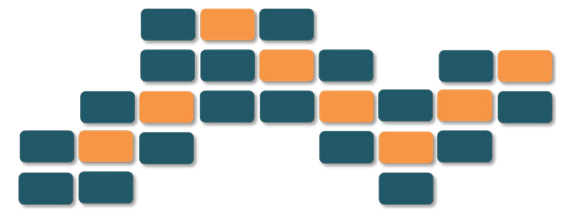
F4. metadata specify the data identifier.

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.

Member of the Helmholtz Association

Dr. Guido Juckeland I Computational Science Group (FWCC) I www.hzdr.de/fwcc

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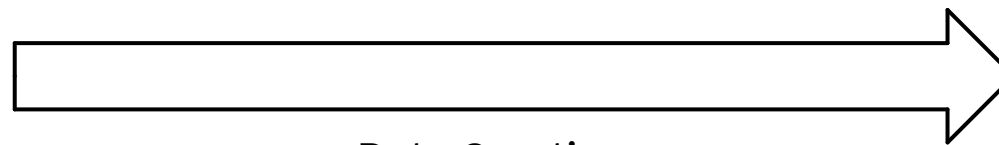
Slide from: Ralph Müller-Pfefferkorn, "Open Access and Research Data", presented at EPC-Workshop on Sep 19, 2018 in Dresden under CC-BY 4.0

Member of the Helmholtz Association

Dr. Guido Juckeland I Computational Science Group (FWCC) I www.hzdr.de/fwcc

Prerequisite for Open Access to research data:

(Systematic) Research Data Management

www.digitalbevaring.dk

Member of the Helmholtz Association

Dr. Guido Juckeland I Computational Science Group (FWCC) I www.hzdr.de/fwcc

DATA CONTINUUM

| Private Domain | Group Domain | Preservation | Access and Re-Use |
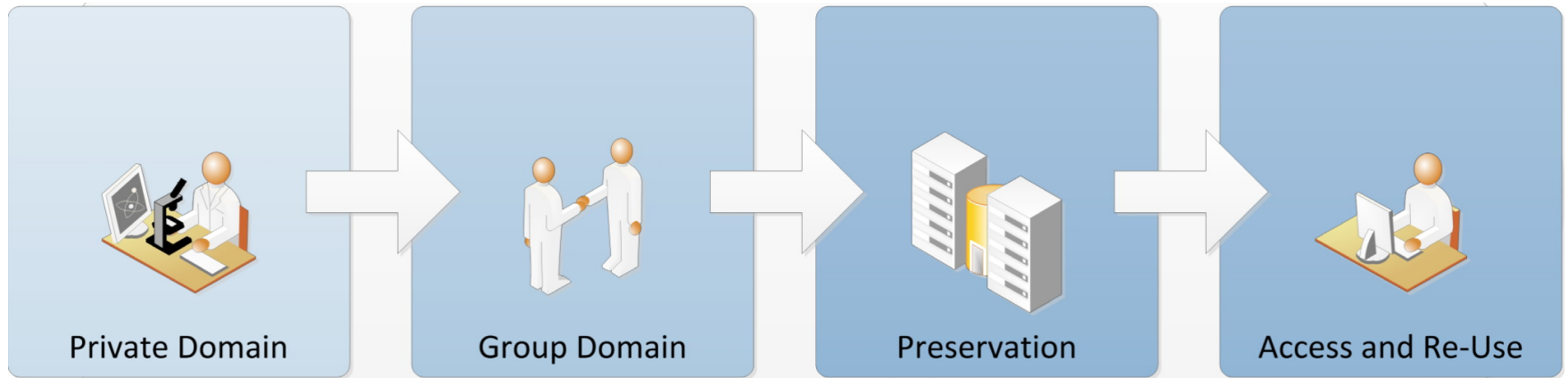
Simple Documentation (Metadata) → Data Curation → Rich Documentation (Metadata)

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.

DRESDEN concept

HZDR

Member of the Helmholtz Association

Dr. Guido Juckeland I Computational Science Group (FWCC) I www.hzdr.de/fwcc

Agenda

- Research Data and Research Data Management

- **Political Implications or What funders want**

- Interactive work on a DMP for LaserLab Europe

Member of the Helmholtz Association

Dr. Guido Juckeland I Computational Science Group (FWCC) I www.hzdr.de/fwcc

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 andprocess 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.**

Data Management Plan (DMP)

Clarifies all questions regarding data management

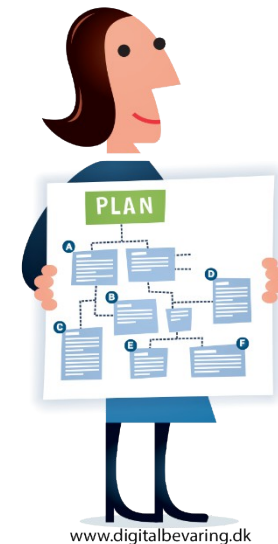- Documents data management during and after a project

- Formal Document

Is created before project start.

- Start early! Write the plan collaboratively!

- Be realistic!

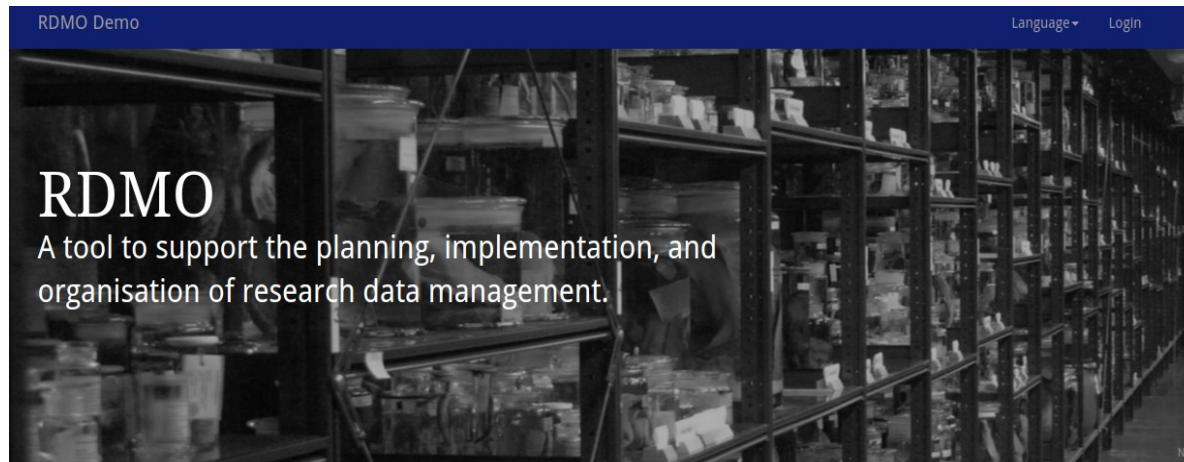Is regularly updated during project runtime.

Member of the Helmholtz Association

Dr. Guido Juckeland I Computational Science Group (FWCC) I www.hzdr.de/fwcc

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!

www.digitalbevaring.dk

Member of the Helmholtz Association

Dr. Guido Juckeland I Computational Science Group (FWCC) I www.hzdr.de/fwcc

DMP - WHERE TO START?

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

Dr. Guido Juckeland I Computational Science Group (FWCC) I www.hzdr.de/fwcc

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.

DRESDEN concept

HZDR

Dr. Guido Juckeland I Computational Science Group (FWCC) I www.hzdr.de/fwcc

AGENDA

- Research Data and Research Data Management

- Political Implications or What funders want

- **Interactive work on a DMP for LaserLab Europe**

Member of the Helmholtz Association

Dr. Guido Juckeland I Computational Science Group (FWCC) I www.hzdr.de/fwcc