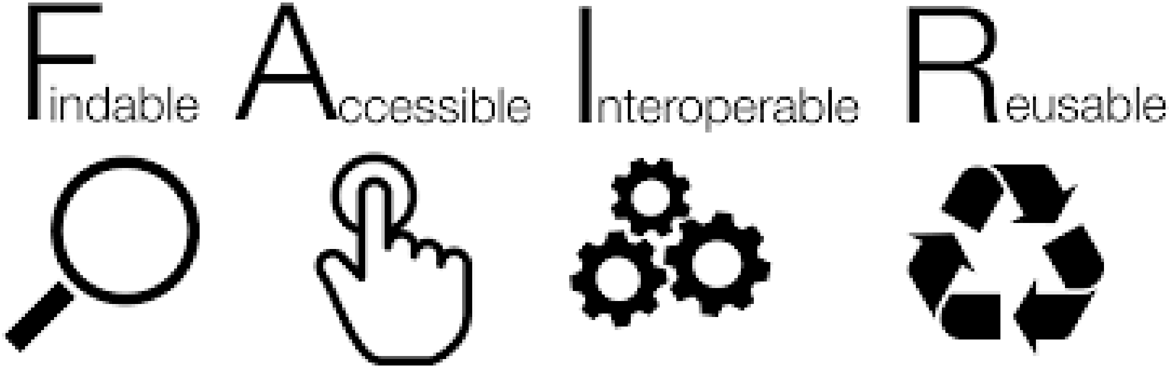


Andy Götz (ESRF)

7 December 2018

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852

FAIR – sounds easy but … is not so easy

The new **FAIR bible**

https://doi.org/10.2777/1524

European Commission

Final Report and Action Plan from the European Commission Expert Group on FAIR Data

TURNING FAIR INTO REALITY

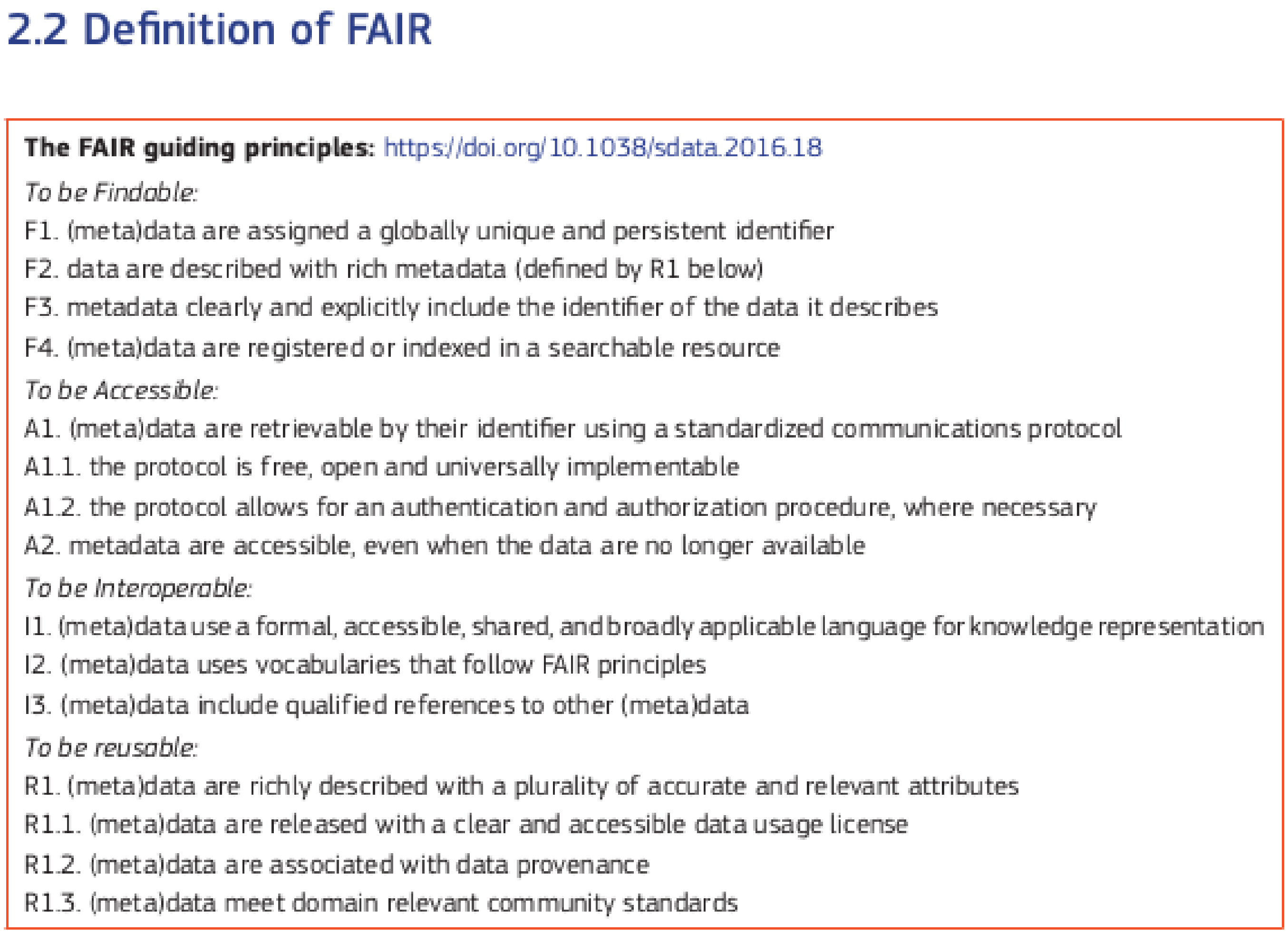
2018

Research and Innovation

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852

FAIR – guiding principles

https://doi.org/10.2777/1524



2.2 Definition of FAIR

The FAIR guiding principles: https://doi.org/10.1038/sdata.2016.18

To be Findable:

F1. (meta)data are assigned a globally unique and persistent identifier

F2. data are described with rich metadata (defined by R1 below)

F3. metadata clearly and explicitly include the identifier of the data it describes

F4. (meta)data are registered or indexed in a searchable resource

To be Accessible:

A1. (meta)data are retrievable by their identifier using a standardized communications protocol

A1.1. the protocol is free, open 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

To be Interoperable:

I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation

I2. (meta)data uses vocabularies that follow FAIR principles

I3. (meta)data include qualified references to other (meta)data

To be reusable:

R1. (meta)data are richly described with 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 data provenance

R1.3. (meta)data meet domain relevant community standards

Figure 2. The FAIR guiding principles

- ➢ **DOI**
- ➢ **Nexus**
- ➢ **E-logbook**
- ➢ **Metadata Catalogue**

- ➢ **http**
- ➢ **ResourceAsync**
- ➢ **AAI**
- ➢ **Metadata Catalogue**

- ➢ **Nexus**

- ➢ **CC BY**
- ➢ **ESRF DOI**

zon 2020 research and

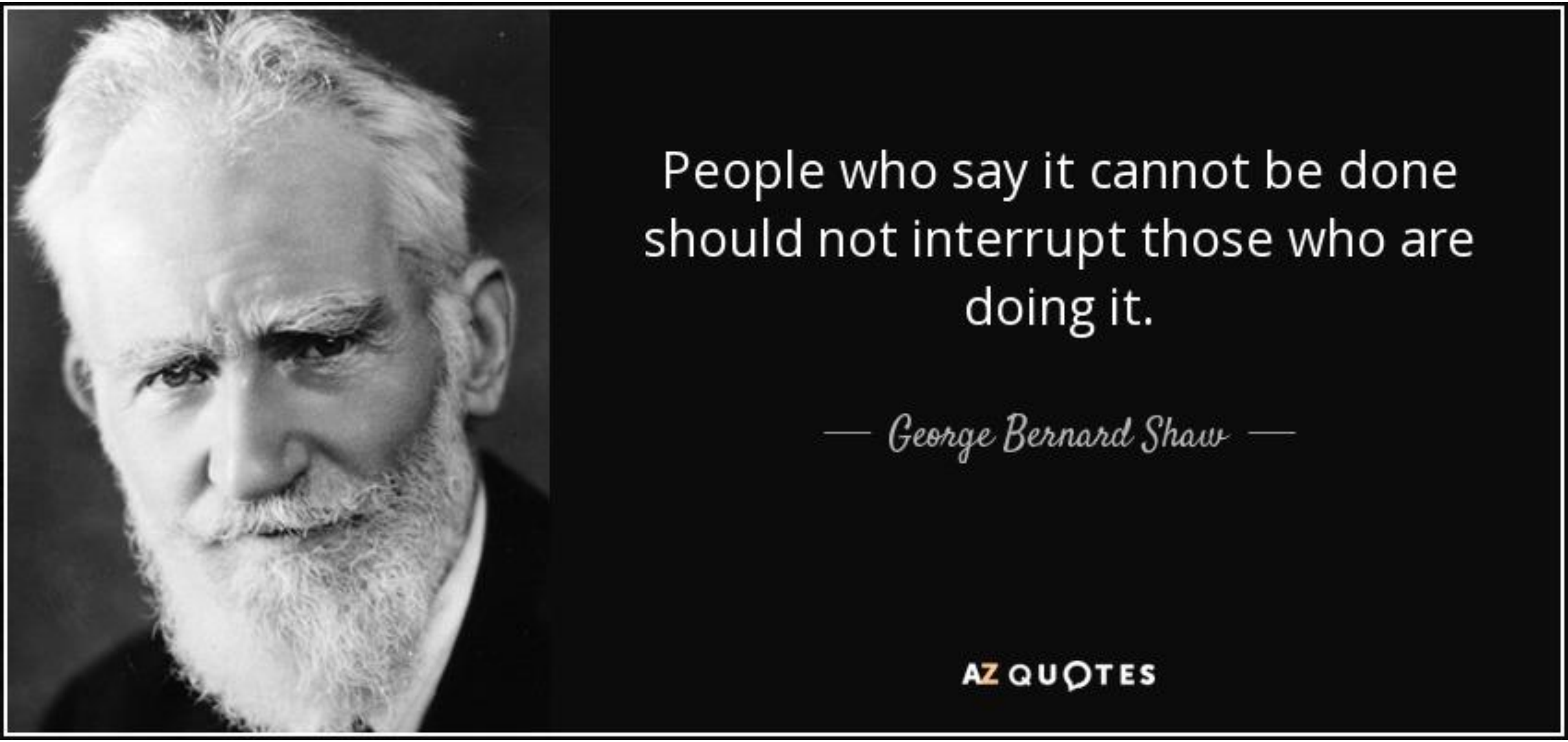
Define | Implement | Embed and sustain

27 recommendations !

Concept for FAIR implementation	FAIR data culture	FAIR data ecosystem	Skills for FAIR	Incentives and metrics for FAIR data and services	Investment in FAIR
Rec. 1: Define FAIR for Implementation	Rec. 4: Develop interoperability frameworks	Rec. 7: Support semantic technologies	Rec. 10: Professionalise data science & stewardship roles	Rec. 12: Develop metrics for FAIR Digital Objects	Rec. 14: Provide strategic and coordinated funding
Rec. 2: Implement a Model for FAIR Digital Objects	Rec. 5: Ensure data management via DMPs	Rec. 8: Facilitate automated processing	Rec. 11: Implement curriculum frameworks and training	Rec. 13: Develop metrics to certify FAIR services	Rec. 15: Provide sustainable funding
Rec. 3: Develop components of a FAIR ecosystem	Rec. 6: Recognize & reward FAIR data & stewardship	Rec. 9: Certify FAIR services	Above line = priority recommendations		
Rec. 16: Apply FAIR broadly	Rec. 18: Cost data management	Rec. 22: Use information held in DMPs	Below line = supporting recommendations	Rec. 25: Implement and monitor metrics	Rec. 27: Open EOSC to all providers but ensure services are FAIR
Rec. 17: Align and harmonise FAIR and Open data policy	Rec. 19: Select and prioritise FAIR digital objects	Rec. 23: Develop components to meet research needs		Rec. 26: Support data citation and next generation metrics	
	Rec. 20: Deposit in Trusted Digital Repositories	Rec. 24: Incentivise research infrastructures to support FAIR data			
	Rec. 21: Encourage/incentivise reuse of FAIR outputs				

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852

FAIR – don’t be put off …



People who say it cannot be done should not interrupt those who are doing it.

— George Bernard Shaw —

AZ QUOTES

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852

FAIR – in simple words

GOAL: Making your data FAIR means managing your data in a professional manner so that it can be found, accessed, understood and re-used in the future (10 years or more) by scientists who were not involved in producing the data.

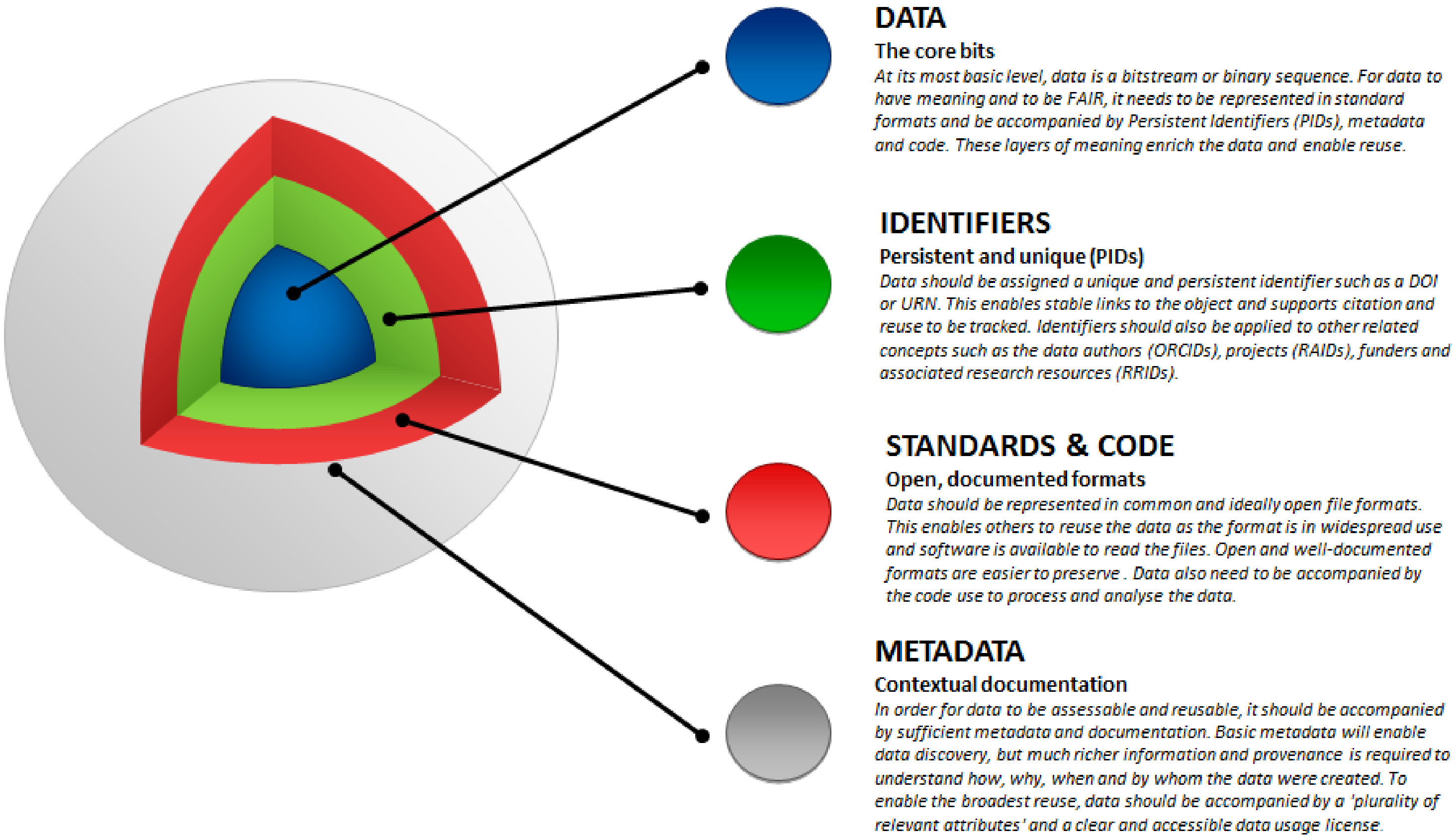
ADVANTAGES: By making data FAIR (1) the quality of the metadata are improved, (2) the data can be referenced by journals and data search machines, (3) the data are better organized, (4) data integrity is ensured, (5) fraud is more difficult, (6) services can be built on top of data repositories, (7) re-use of data is enabled, (8) new algorithms can be developed more easily, (9) the origin of the data are correctly acknowledged, (10) metrics on data can be tracked.

DISADVANTAGES: implementing FAIR requires human and infrastructure resources.

This project has received funding from the European Union's Horizon 2020 research and
innovation programme under grant agreement No 823852

FAIR – Data Objects

DATA: are not simply data in files any more but are now data objects with rich metadata, following standards, linked to software codes and identified by persistent identifiers



DATA

The core bits

At its most basic level, data is a bitstream or binary sequence. For data to have meaning and to be FAIR, it needs to be represented in standard formats and be accompanied by Persistent Identifiers (PIDs), metadata and code. These layers of meaning enrich the data and enable reuse.

IDENTIFIERS

Persistent and unique (PIDs)

Data should be assigned a unique and persistent identifier such as a DOI or URN. This enables stable links to the object and supports citation and reuse to be tracked. Identifiers should also be applied to other related concepts such as the data authors (ORCIDs), projects (RAIDs), funders and associated research resources (RRIDs).

STANDARDS & CODE

Open, documented formats

Data should be represented in common and ideally open file formats. This enables others to reuse the data as the format is in widespread use and software is available to read the files. Open and well-documented formats are easier to preserve . Data also need to be accompanied by the code use to process and analyse the data.

METADATA

Contextual documentation

In order for data to be assessable and reusable, it should be accompanied by sufficient metadata and documentation. Basic metadata will enable data discovery, but much richer information and provenance is required to understand how, why, when and by whom the data were created. To enable the broadest reuse, data should be accompanied by a 'plurality of relevant attributes' and a clear and accessible data usage license.

FAIR – Data Ecosystem

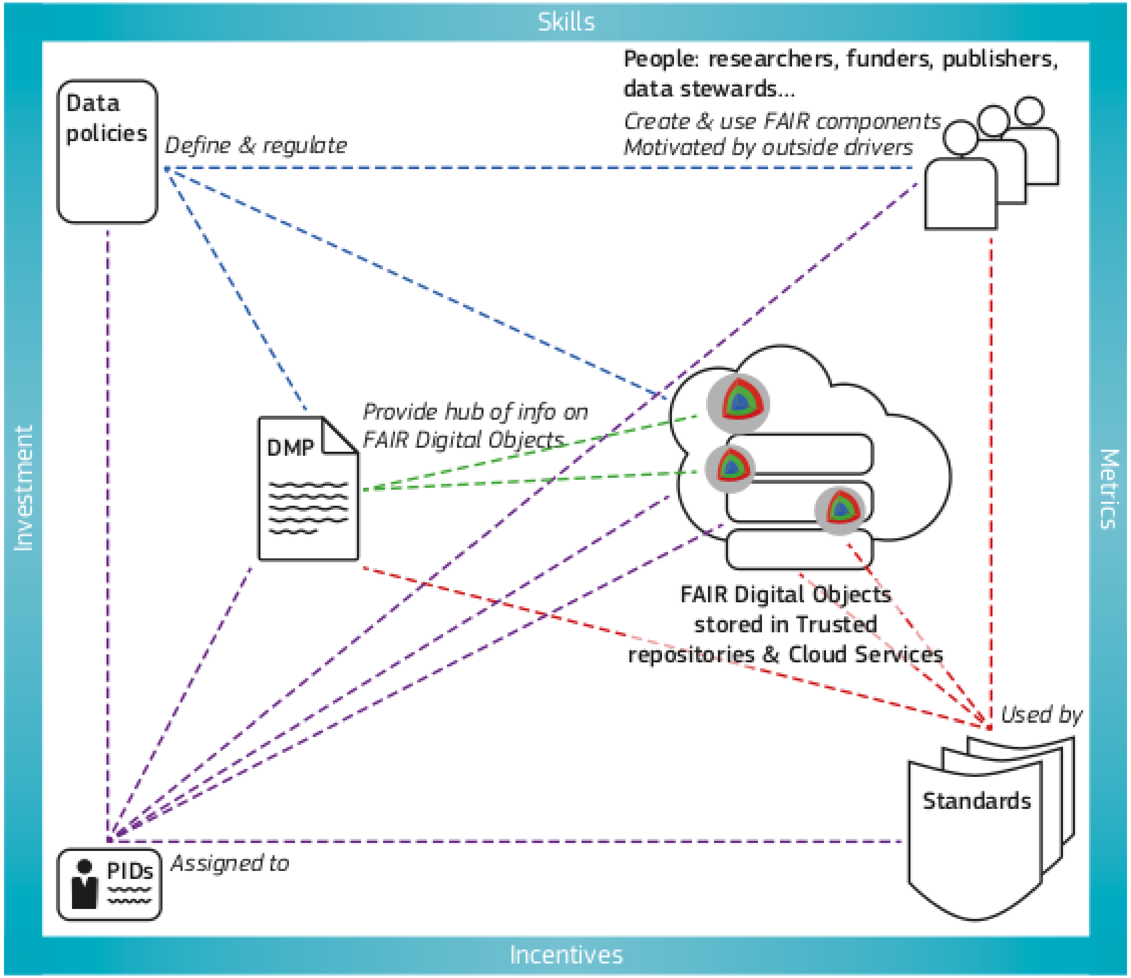

Skills
Investment
Metrics
Incentives
Data policies
Define & regulate
People: researchers, funders, publishers, data stewards...
Create & use FAIR components
Motivated by outside drivers
DMP
Provide hub of info on FAIR Digital Objects
FAIR Digital Objects stored in Trusted repositories & Cloud Services
Used by
Standards
PIDs
Assigned to

Figure 9. The interactions between components in the FAIR data ecosystem. Notes on this figure:

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852

ESRF’s long and winding road to FAIR

2011 - Developed Data Policy framework

2015 - Data Policy endorsed

2014 - Hired Data manager

2014 - Built a prototype

2015 - Tested on beamline

2018 - Implement DOIs

2015-2018 - Rollout on ½ beamlines

2017 - Hired 2nd Data manager

2018 - Developed e-logbook

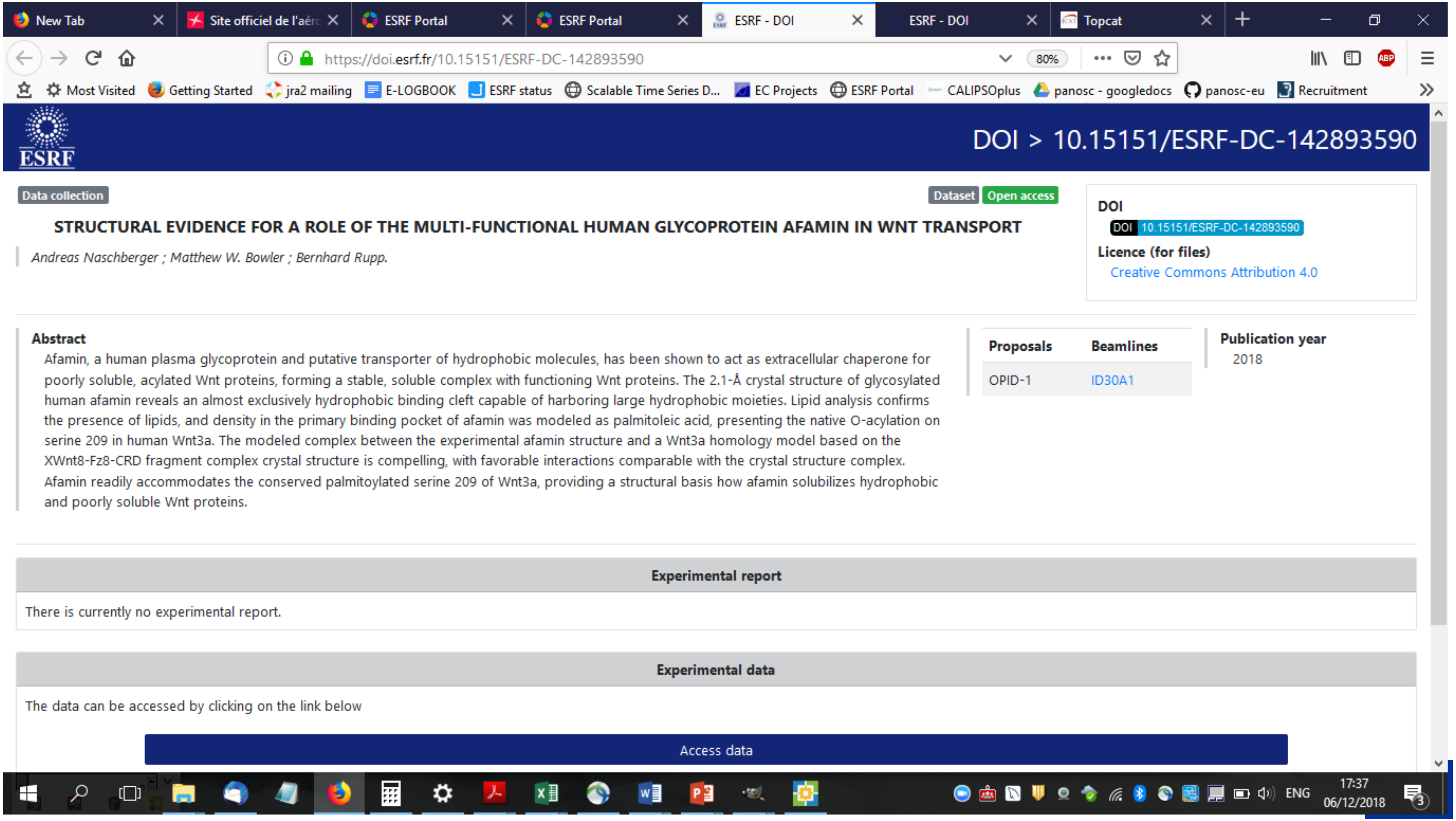
2019 - Train scientists

2021 – Rollout on all beamlines

2021-future – Develop data services

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852

PID example - https://doi.esrf.fr/10.15151/ESRF-DC-142893590



DOI > 10.15151/ESRF-DC-142893590

Data collection

Dataset Open access

STRUCTURAL EVIDENCE FOR A ROLE OF THE MULTI-FUNCTIONAL HUMAN GLYCOPROTEIN AFAMIN IN WNT TRANSPORT

Andreas Naschberger ; Matthew W. Bowler ; Bernhard Rupp.

DOI

DOI 10.15151/ESRF-DC-142893590

Licence (for files)

Creative Commons Attribution 4.0

Abstract

Afamin, a human plasma glycoprotein and putative transporter of hydrophobic molecules, has been shown to act as extracellular chaperone for poorly soluble, acylated Wnt proteins, forming a stable, soluble complex with functioning Wnt proteins. The 2.1-Å crystal structure of glycosylated human afamin reveals an almost exclusively hydrophobic binding cleft capable of harboring large hydrophobic moieties. Lipid analysis confirms the presence of lipids, and density in the primary binding pocket of afamin was modeled as palmitoleic acid, presenting the native O-acylation on serine 209 in human Wnt3a. The modeled complex between the experimental afamin structure and a Wnt3a homology model based on the XWnt8-Fz8-CRD fragment complex crystal structure is compelling, with favorable interactions comparable with the crystal structure complex. Afamin readily accommodates the conserved palmitoylated serine 209 of Wnt3a, providing a structural basis how afamin solubilizes hydrophobic and poorly soluble Wnt proteins.

Proposals	Beamlines
OPID-1	ID30A1

Publication year

2018

Experimental report

There is currently no experimental report.

Experimental data

The data can be accessed by clicking on the link below

Access data

FAIR – Persistent Identifiers

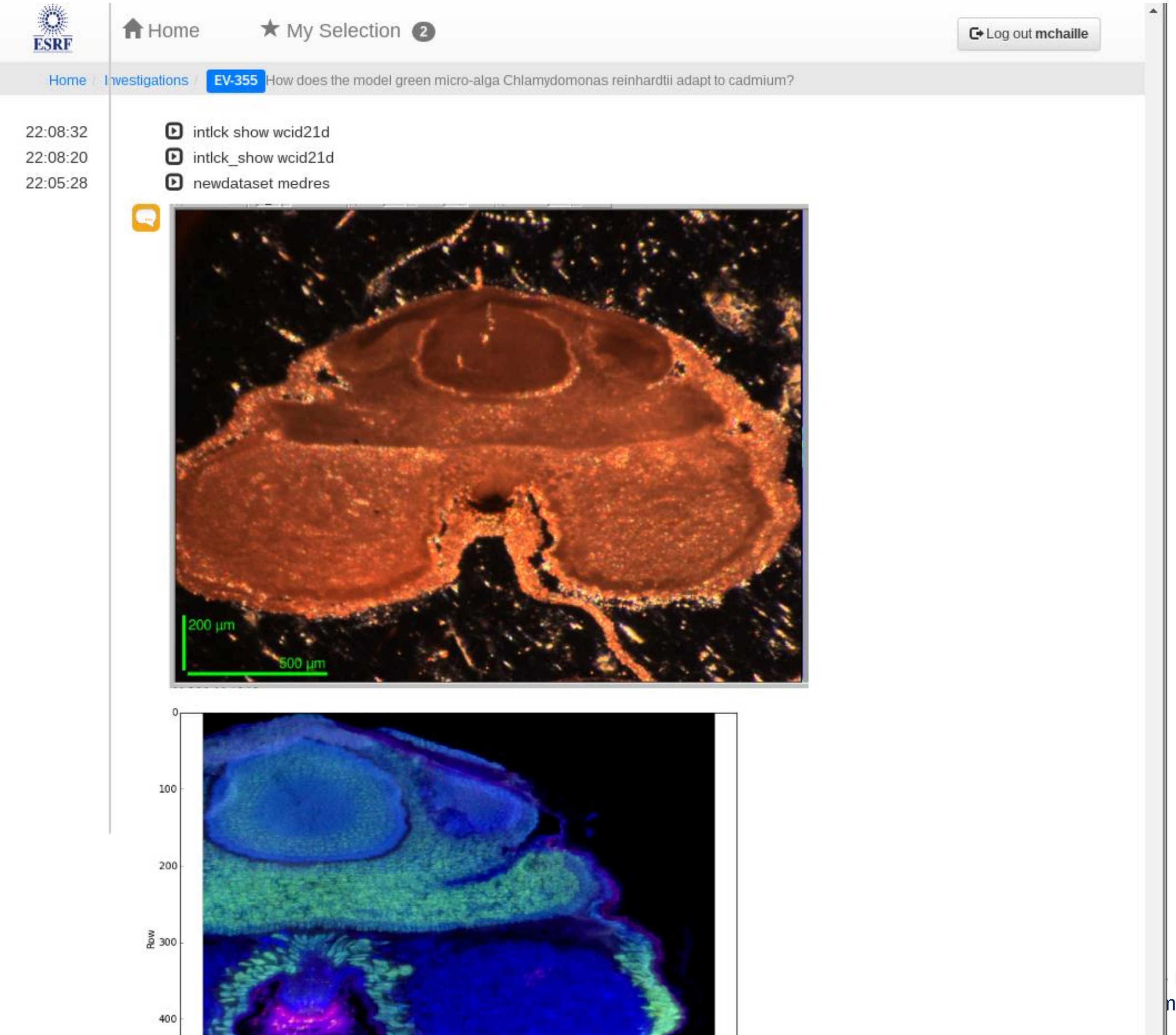
PID : A persistent identifier is required to be able to refer to a data object in a permanent way i.e. independent of changing urls

We did it this way

1. Chose datacite as PID provider (**datacite.org**)
2. Setup a contract with Datacite (a PID cost between 3 – 20 cents / year)
3. Chose and implemented a long term archiving solution (tape library for 90 PBs costs roughly 100 000 euros / yr)
4. Defined and collected metadata for experimental techniques
5. Chose and deployed a metadata catalogue (**icatproject.org**)
6. Setup a workflow to archive data with the correct metadata
7. Setup a web services for creating the landing page for the PIDs (**doi.esrf.fr**)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852

FAIR – eLogbook



ESRF

Home

My Selection 2

Log out mchaille

Home / Investigations / EV-355 How does the model green micro-alga Chlamydomonas reinhardtii adapt to cadmium?

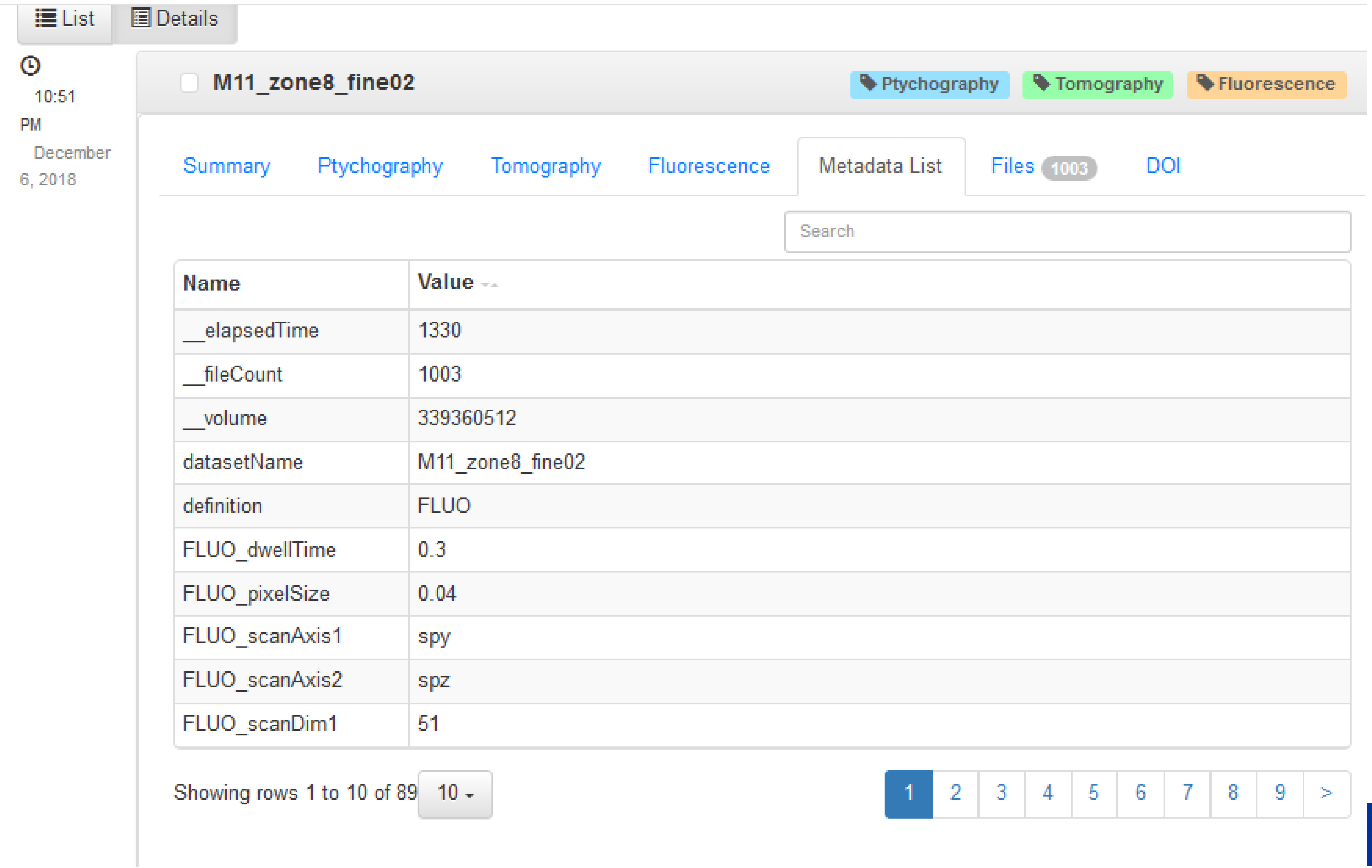
22:08:32 intlck show wcid21d

22:08:20 intlck_show wcid21d

22:05:28 newdataset medres

European Union's Horizon 2020 research anc
nent No 823852

FAIR – Metadata



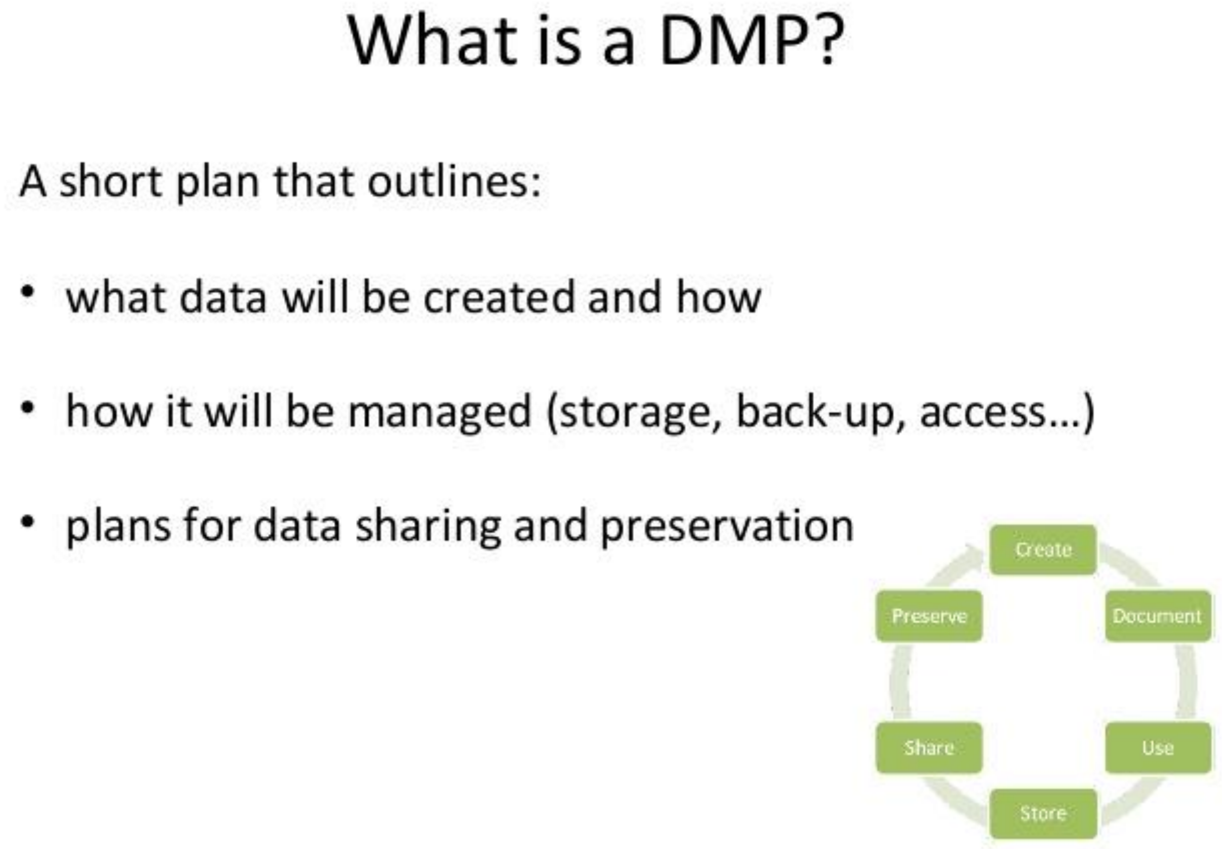
10:51 PM December 6, 2018

□ M11_zone8_fine02

Name	Value
__elapsedTime	1330
__fileCount	1003
__volume	339360512
datasetName	M11_zone8_fine02
definition	FLUO
FLUO_dwellTime	0.3
FLUO_pixelSize	0.04
FLUO_scanAxis1	spy
FLUO_scanAxis2	spz
FLUO_scanDim1	51

Showing rows 1 to 10 of 89

FAIR – Data Management Plans



What is a DMP?

A short plan that outlines:

- what data will be created and how
- how it will be managed (storage, back-up, access…)
- plans for data sharing and preservation

Create
Document
Use
Store
Share
Preserve

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852

FAIR – Data Repositories

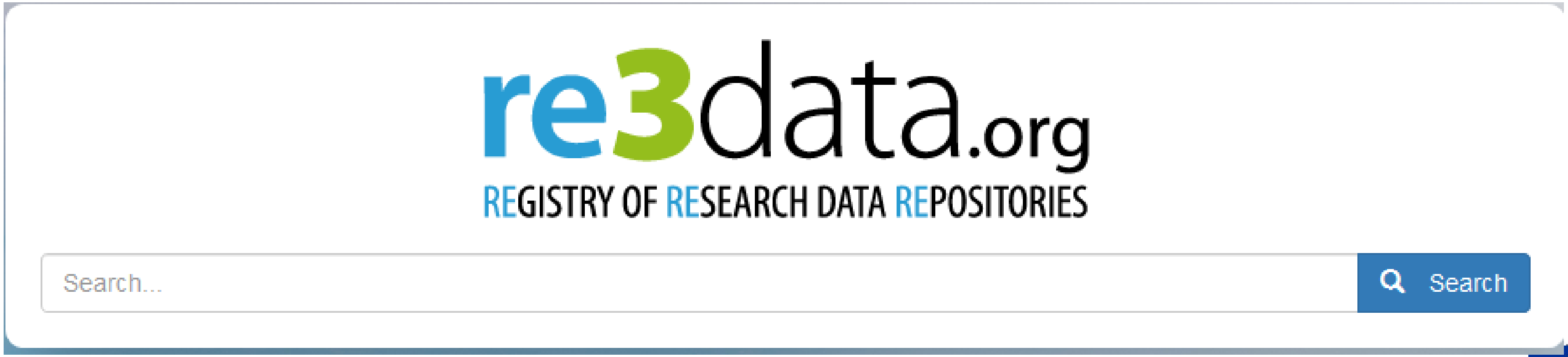
Data Repository: where your data is stored and are FAIR

Public repositories : zenodo, figshare, …

Community repositories: CXIDB, EMPIAR …

Institute repositories: ESRF, ILL, XFEL, …

Private repositories: are not a repository e.g. Dropbox



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852

FAIR – Resources

To implement FAIR Data Management you need at least one or more of the following human resources:

1. **Scientist –** who produces data and does science
2. **Data scientist –** defines metadata for techniques
3. **Data manager –** IT specialist who implements data policy
4. **Data archiver –** DevOps specialist who implements data infrastructure
5. **Management –** who understands the need for Data Policy

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852

FAIR – Open Data

Some scientists misinterpret FAIR data policy to be a way to “steal their data” :

1. **Open Data –** is data which the Scientist has made open
2. **Embargo Data–** data which is under restricted access
3. **Publicly funded –** data which was obtained for free at a publicly funded site
4. **Proprietary Data –** data the user has paid for and is private
5. **Data Services –** only make sense if there is a Data Policy and data will be open at some time

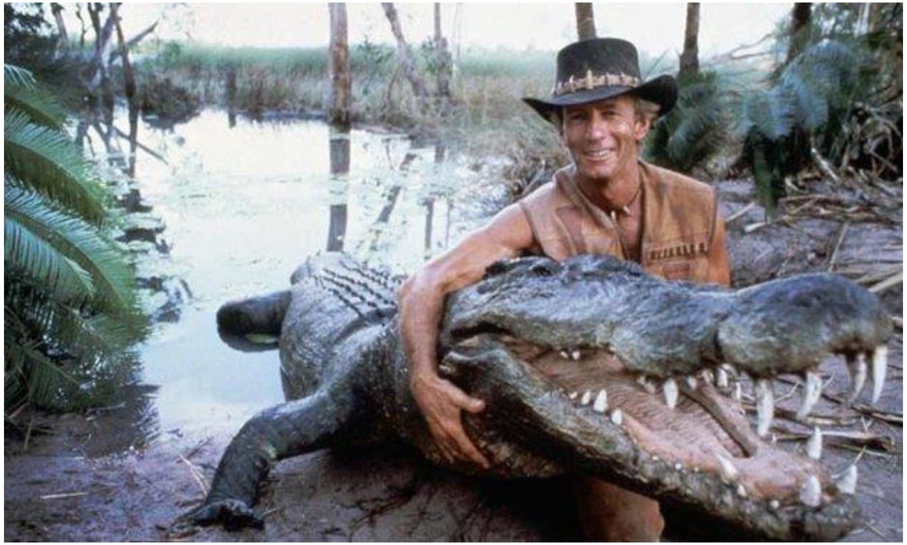
This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852

Conclusion

Implementing FAIR data is a big challenge but is worth it because it improves the quality of the data and enables Open Science

There are many resources out there – use them!

Hire a data manager !



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 823852