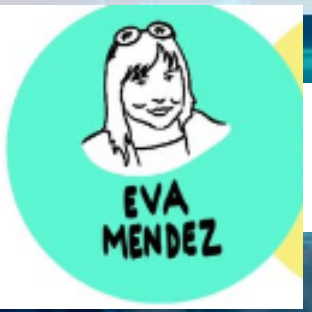
FAIRSFAIR
Fostering Fair Data Practices in Europe

Overview of FAIRsFAIR metadata catalogue integration

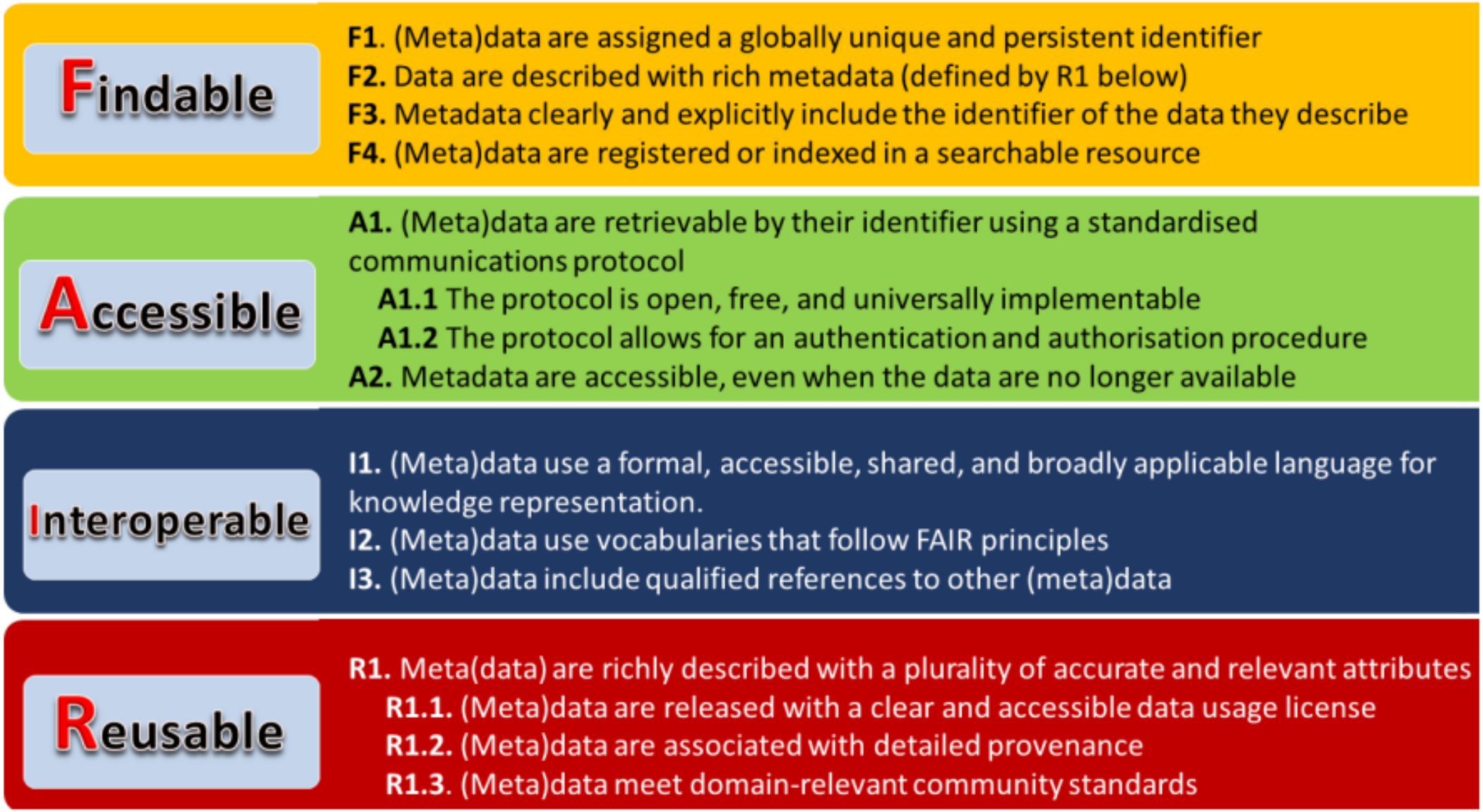
Eva Mendez, UC3M, and Simon Lambert, STFC



emendez@bib.uc3m.es
@evamen

February 17, 2022

FAIRsFAIR "Fostering FAIR Data Practices in Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558



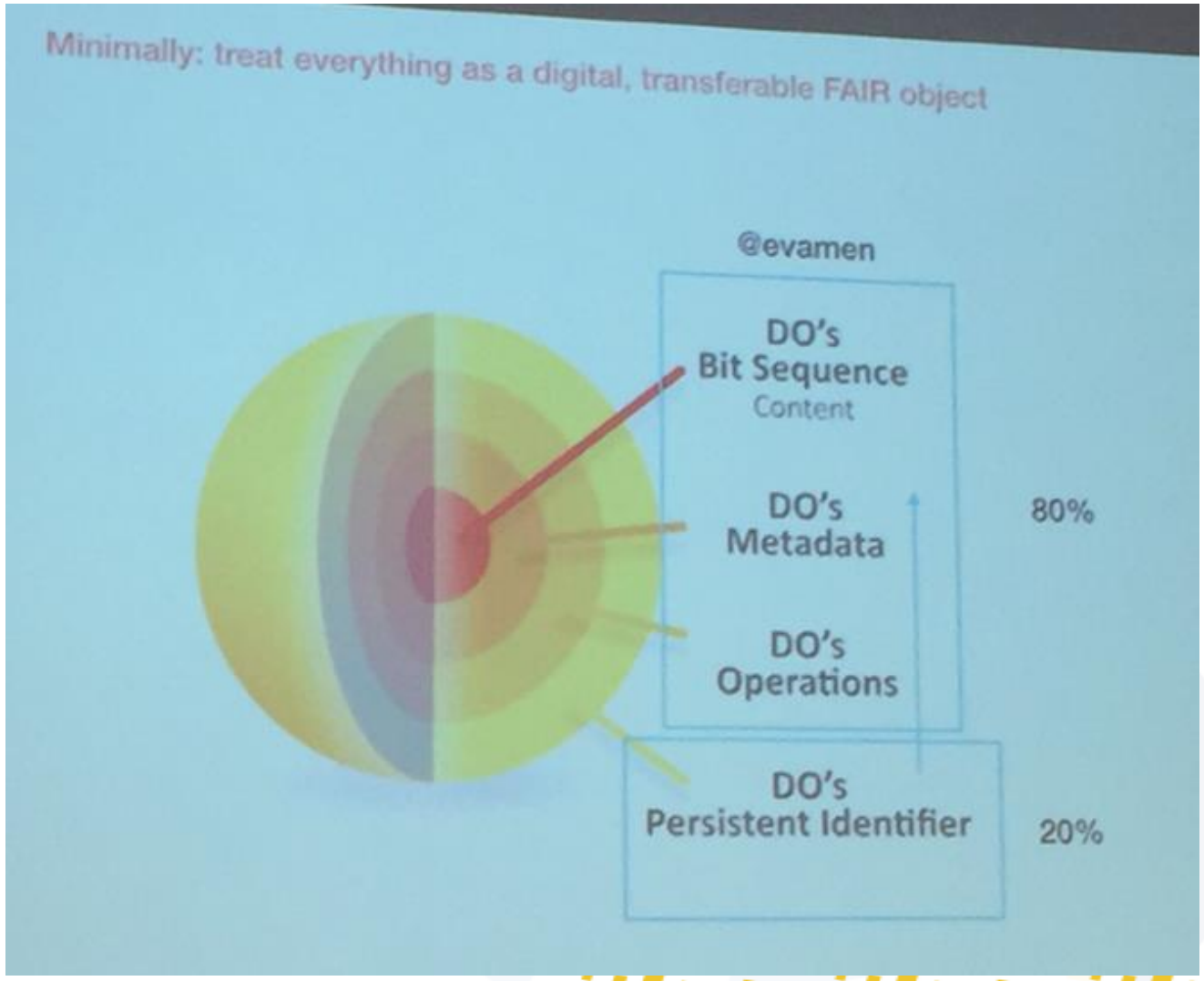
FORCE11 - FAIR Data Guiding Principles
https://www.force11.org/group/fairgroup/fairprinciples

FAIR Guiding Principles: all about metadata

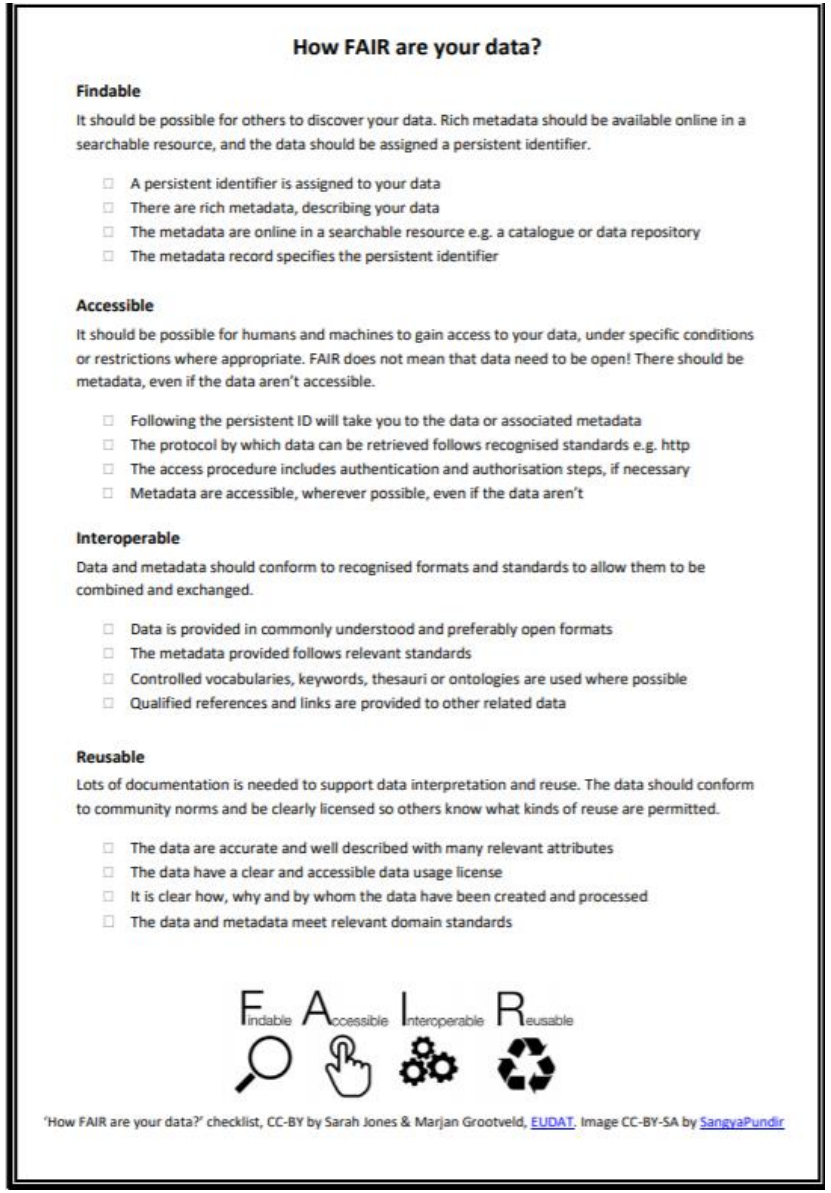
80% metadata

20% PIDs

@



Metadata at the core of FAIR data



How FAIR are your data?

Findable

It should be possible for others to discover your data. Rich metadata should be available online in a searchable resource, and the data should be assigned a persistent identifier.

- ☐ A persistent identifier is assigned to your data
- ☐ There are rich metadata, describing your data
- ☐ The metadata are online in a searchable resource e.g. a catalogue or data repository
- ☐ The metadata record specifies the persistent identifier

Accessible

It should be possible for humans and machines to gain access to your data, under specific conditions or restrictions where appropriate. FAIR does not mean that data need to be open! There should be metadata, even if the data aren't accessible.

- ☐ Following the persistent ID will take you to the data or associated metadata
- ☐ The protocol by which data can be retrieved follows recognised standards e.g. http
- ☐ The access procedure includes authentication and authorisation steps, if necessary
- ☐ Metadata are accessible, wherever possible, even if the data aren't

Interoperable

Data and metadata should conform to recognised formats and standards to allow them to be combined and exchanged.

- ☐ Data is provided in commonly understood and preferably open formats
- ☐ The metadata provided follows relevant standards
- ☐ Controlled vocabularies, keywords, thesauri or ontologies are used where possible
- ☐ Qualified references and links are provided to other related data

Reusable

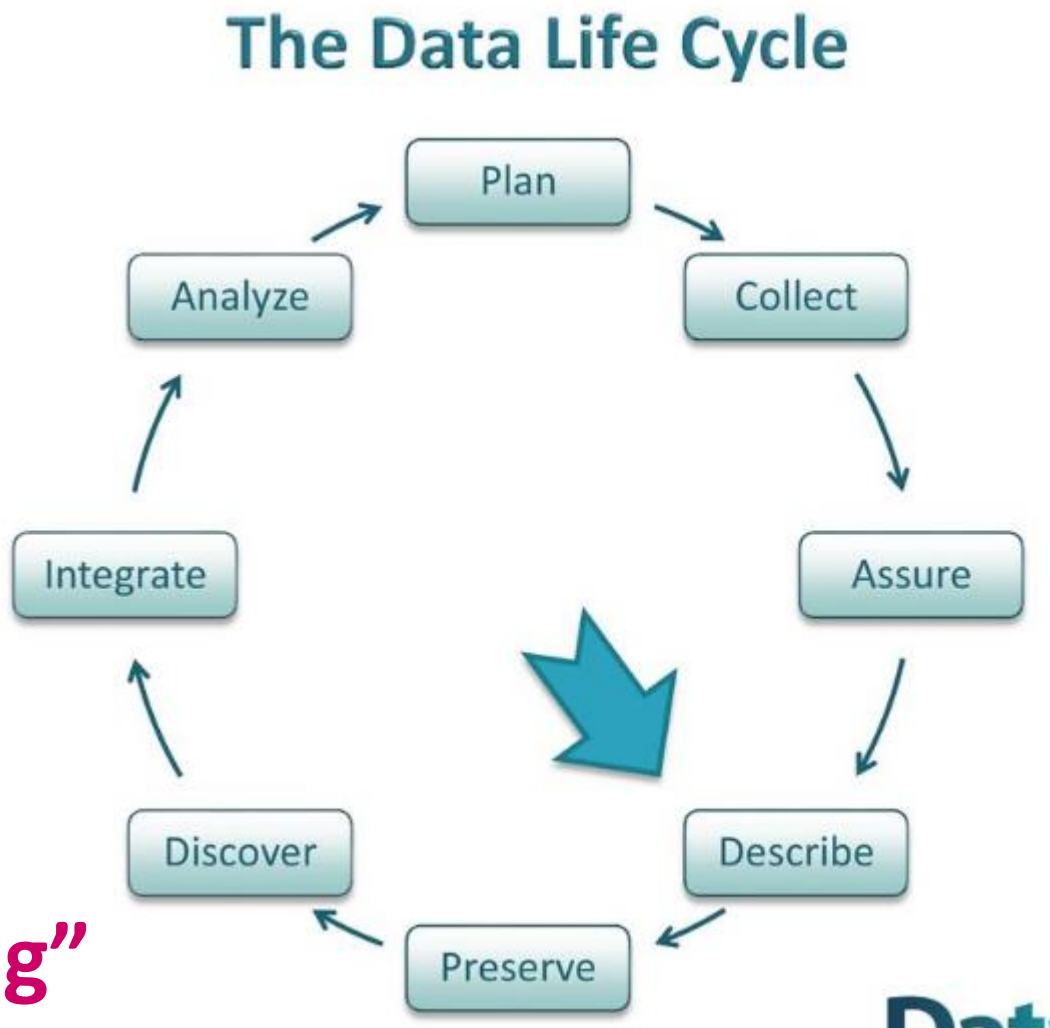
Lots of documentation is needed to support data interpretation and reuse. The data should conform to community norms and be clearly licensed so others know what kinds of reuse are permitted.

- ☐ The data are accurate and well described with many relevant attributes
- ☐ The data have a clear and accessible data usage license
- ☐ It is clear how, why and by whom the data have been created and processed
- ☐ The data and metadata meet relevant domain standards

'How FAIR are your data?' checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT. Image CC-BY-SA by SangyaPundir

- *Making data findable, including provisions for* ***metadata***
- *What* ***metadata*** *will be created? In case* ***metadata standards*** *do not exist in your discipline, please outline what type of metadata will be created and how.*
- *Where will the data and associated* ***metadata****, … be deposited?*
- *Interoperability of your data… What data and* ***metadata vocabularies,*** *standards or methodologies will you follow to make your data interoperable?*
- The **Research Data Alliance** provides a ***Metadata Standards Directory*** that can be searched for discipline-specific standards and associated tools.

- **WHO** created the data?
- **WHAT** is the content of the data?
- **WHEN** were the data created?
- **WHERE** is it geographically?
- **HOW** were the data developed?
- **WHY** were the data developed?



Metadata: data “reporting”

DataONE

Metadata in the DATA communication…

- When you **provide** data to someone else, what types of information would you want to include with the data to make them useful?

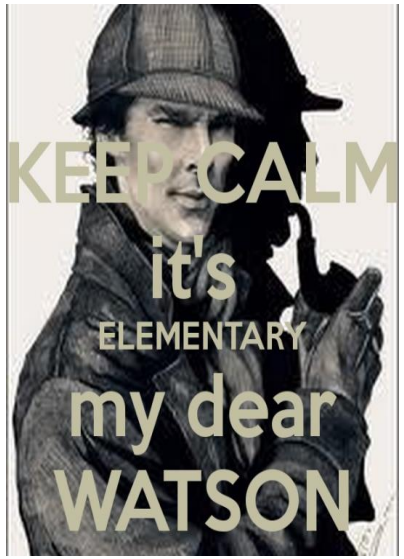
- When you **receive** a dataset from an external project/researcher, what types of details do you want to know about the data?

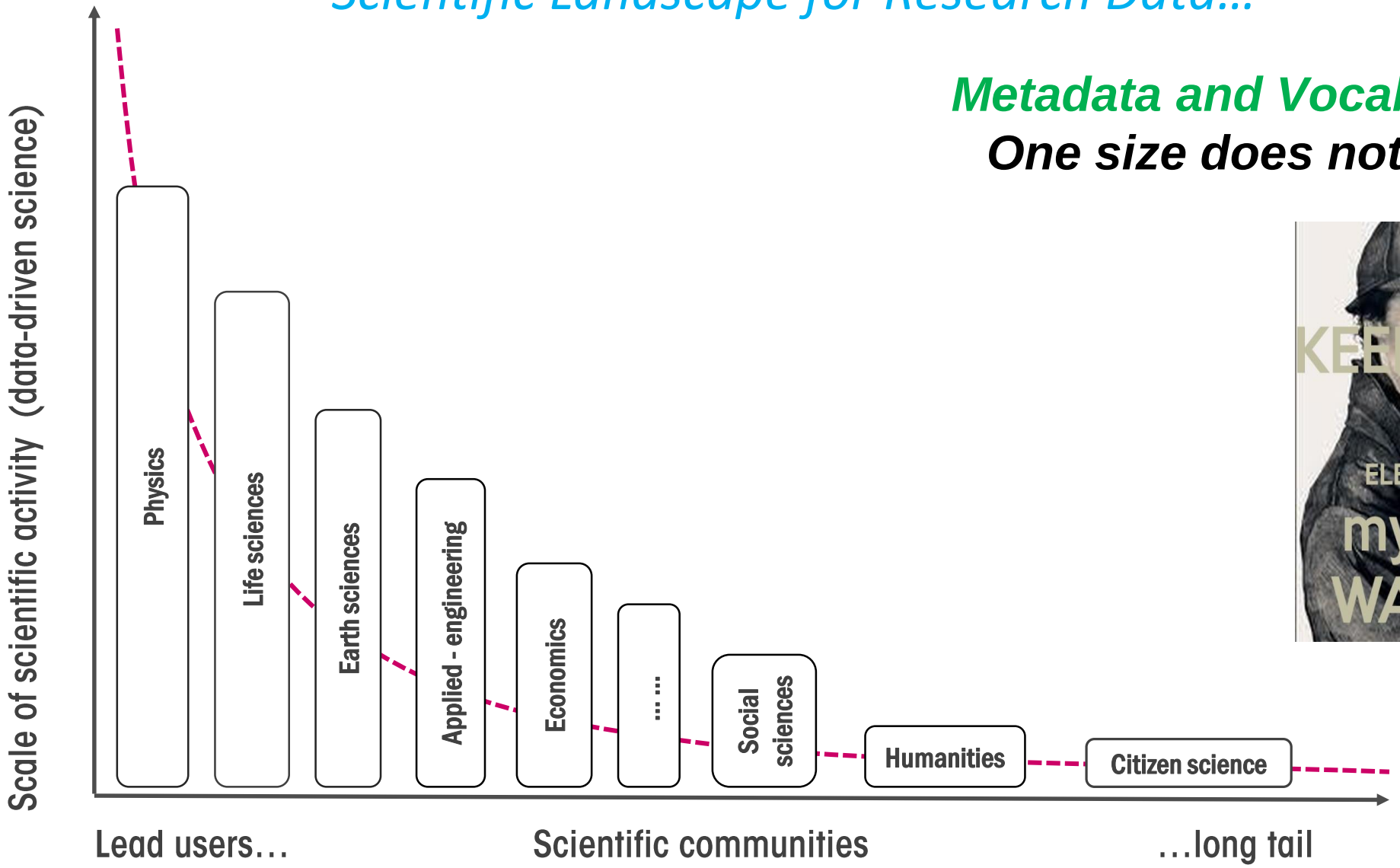
EOSC & FAIR DATA

Scientific Landscape for Research Data…

Metadata and Vocabularies

One size does not fit all!!!





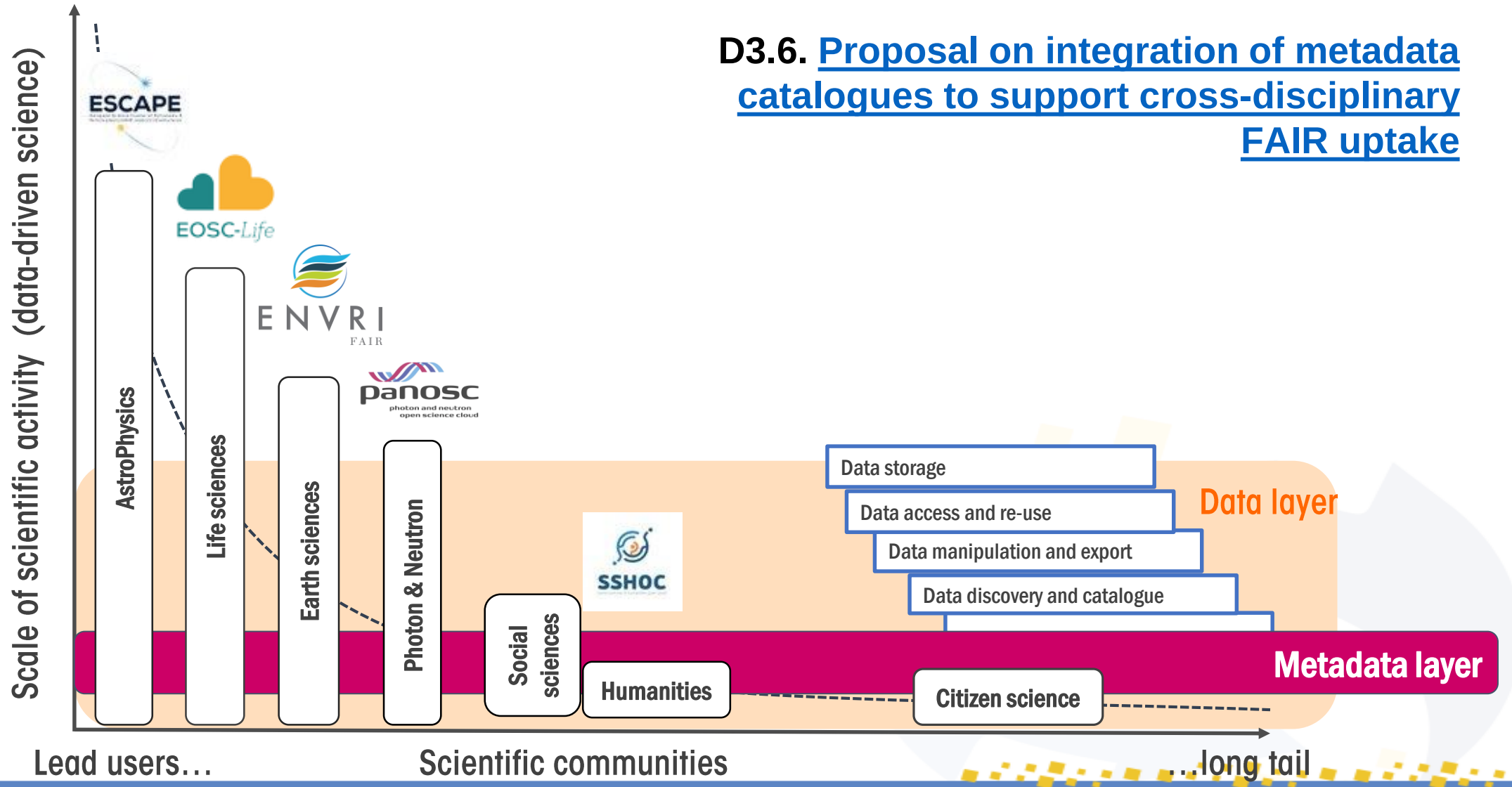
Back to the problem

Metadata are like toothbrushes...

…Everyone thinks that it is a good idea,
but nobody wants to use someone else's.

EOSC & FAIR DATA

D3.6. [Proposal on integration of metadata catalogues to support cross-disciplinary FAIR uptake]



EOSC-Life (Meta)data catalogue

panosc photon and neutron open science cloud (Meta)data catalogue

SSHOC (Meta)data catalogue

ENVRI FAIR (Meta)data catalogue

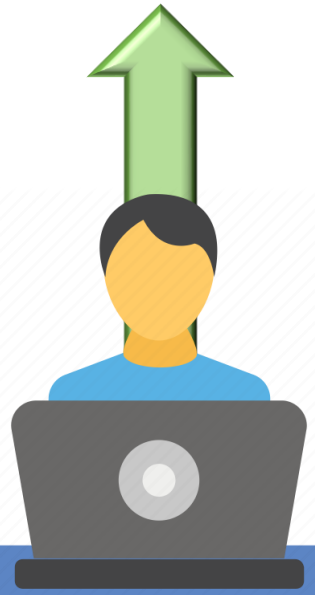
ESCAPE (Meta)data catalogue

Repositories? (Meta)data catalogue

B2FIND Find Research Data

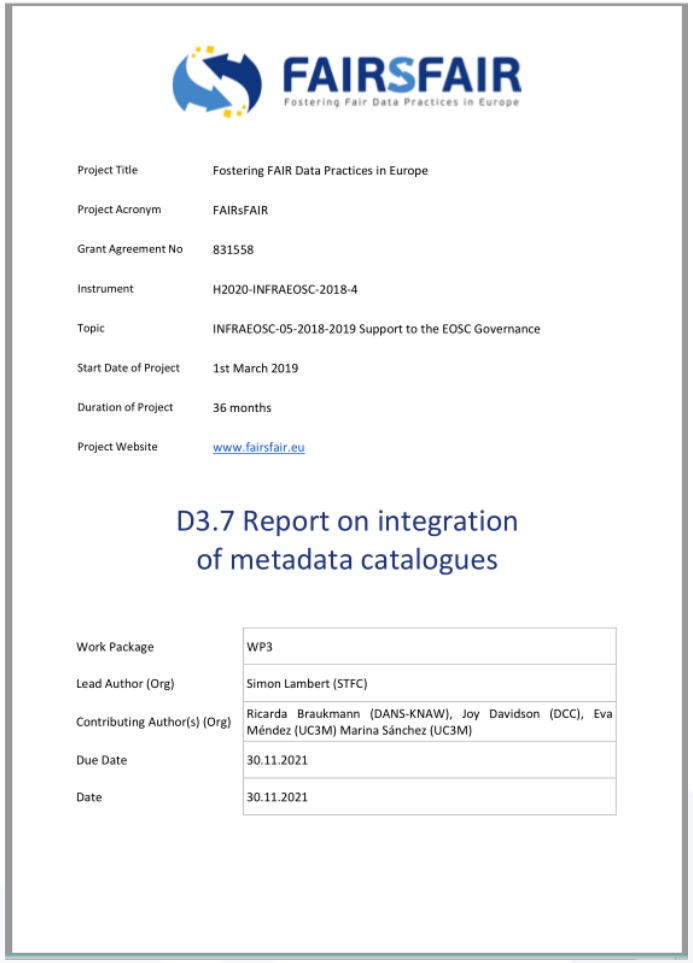
Integrated (Meta)data catalogue

B2FIND Find Research Data



Piloting integration of metadata catalogues

- The goal: **facilitating cross-disciplinary data discovery**
- FAIRsFAIR explored some practical implementation challenges that hinder reuse of existing metadata mappings
  - D3.7: Report on integration of metadata catalogues



FAIRSFAIR
Fostering Fair Data Practices in Europe

Project Title	Fostering FAIR Data Practices in Europe
Project Acronym	FAIRsFAIR
Grant Agreement No	831558
Instrument	H2020-INFRAEOSC-2018-4
Topic	INFRAEOSC-05-2018-2019 Support to the EOSC Governance
Start Date of Project	1st March 2019
Duration of Project	36 months
Project Website	www.fairsfair.eu

D3.7 Report on integration of metadata catalogues

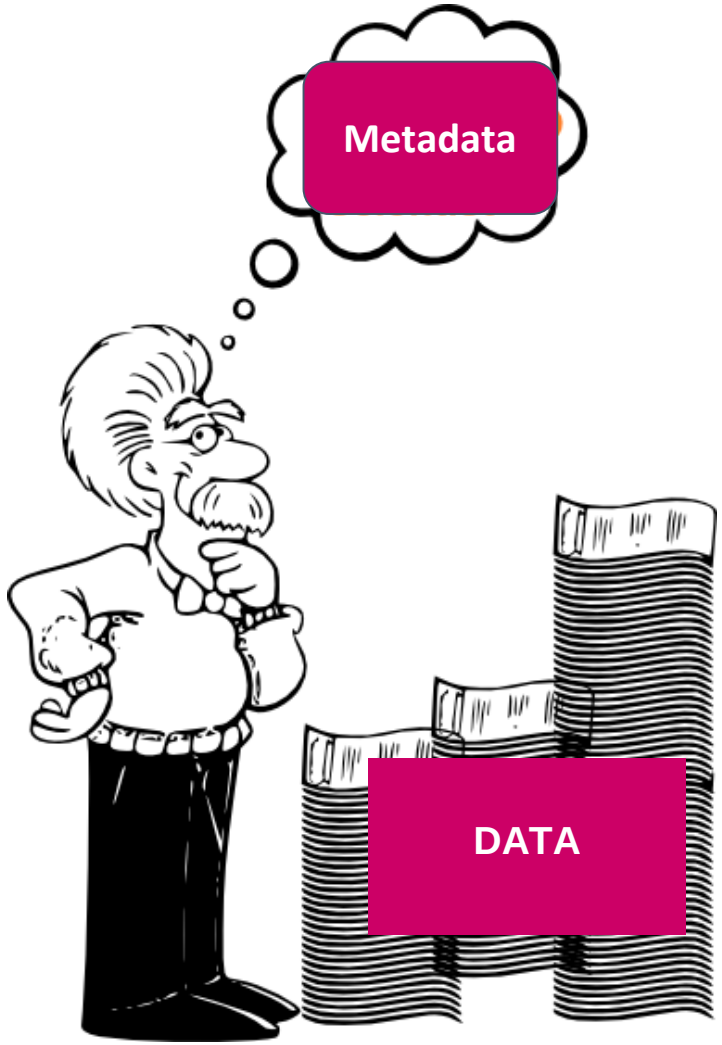
Work Package	WP3
Lead Author (Org)	Simon Lambert (STFC)
Contributing Author(s) (Org)	Ricarda Braukmann (DANS-KNAW), Joy Davidson (DCC), Eva Méndez (UC3M) Marina Sánchez (UC3M)
Due Date	30.11.2021
Date	30.11.2021

The potential of DCAT

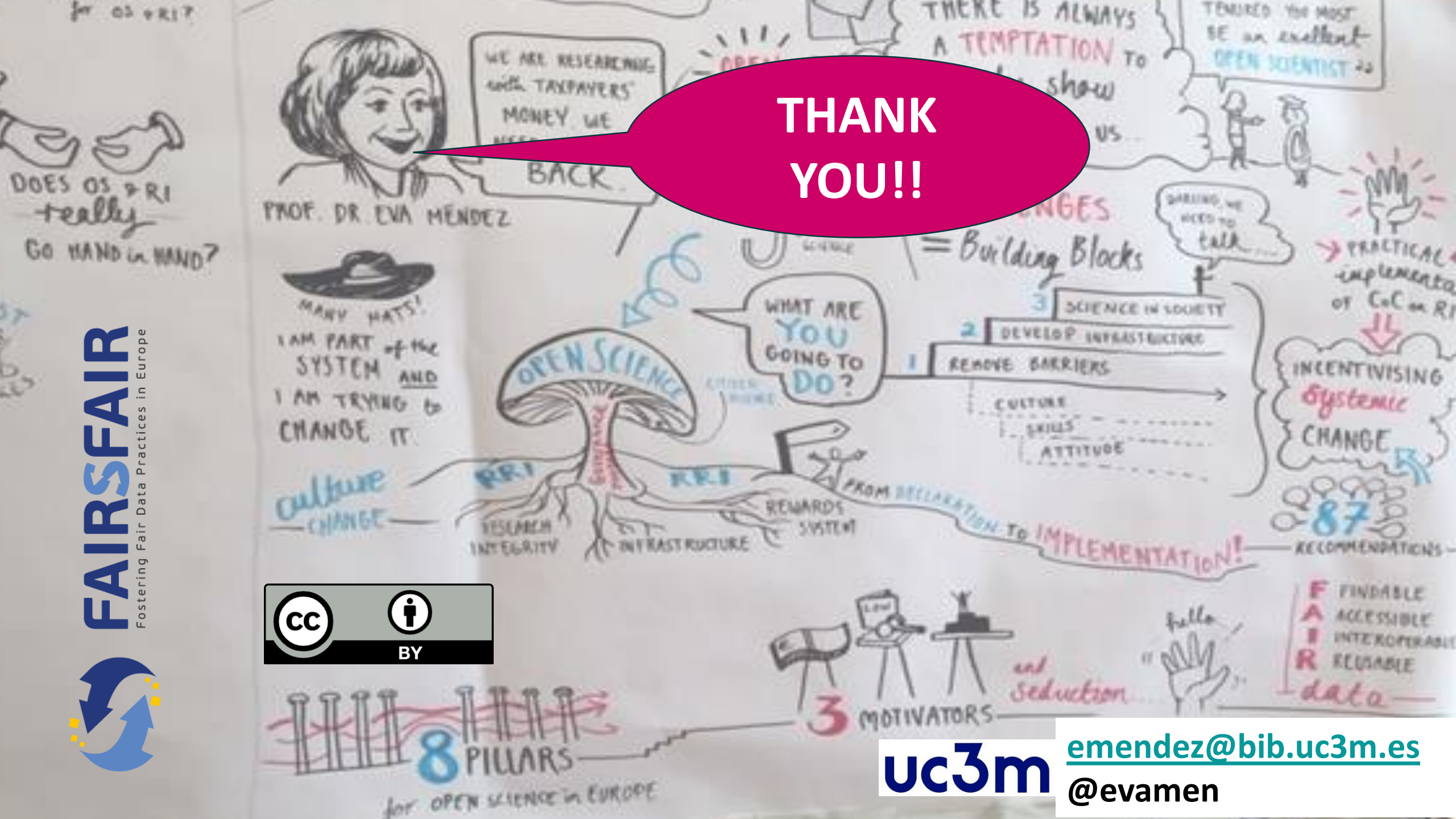
- We worked with two of the thematic cluster projects (SSHOC and PaNOSC) and two service providers (B2FIND and OpenAIRE) to assess the feasibility of DCAT from both the domain specific and aggregator perspectives
- A positive attitude to **DCAT**, but there must be demand
- Potential role for **DCAT** to support aggregation and findability of metadata catalogues
- Combined approach of **DCAT2** at the generic level and **DDI-CDI** at the domain specific level could provide a solution for supporting both the findability and interoperability of heterogeneous research data

(DCAT and DDI-CDI as complementary)

Other findings/Challenges



- Implications of the differences between research domains [**Inter-cross disciplinary** research]
- Metadata **standards**: avoid the *"toothbrush effect"*
- Emergent desire for a central documented collection of metadata mappings
- **Metadata quality** is not only at schema level (metadata formats), but also at scheme (semantic artifacts) and content
- Good research outcomes (publications and data) include **good metadata**, but difficult to balance ***rich/cool/good*** metadata.
- Requirements for **sustainability** (in a broad sense)



emendez@bib.uc3m.es
@evamen