FAIRSFAIR
Fostering Fair Data Practices in Europe

Loving DATA… loving 'cool' metadata

Metadata for FAIR data

emendez@bib.uc3m.es
@evamen

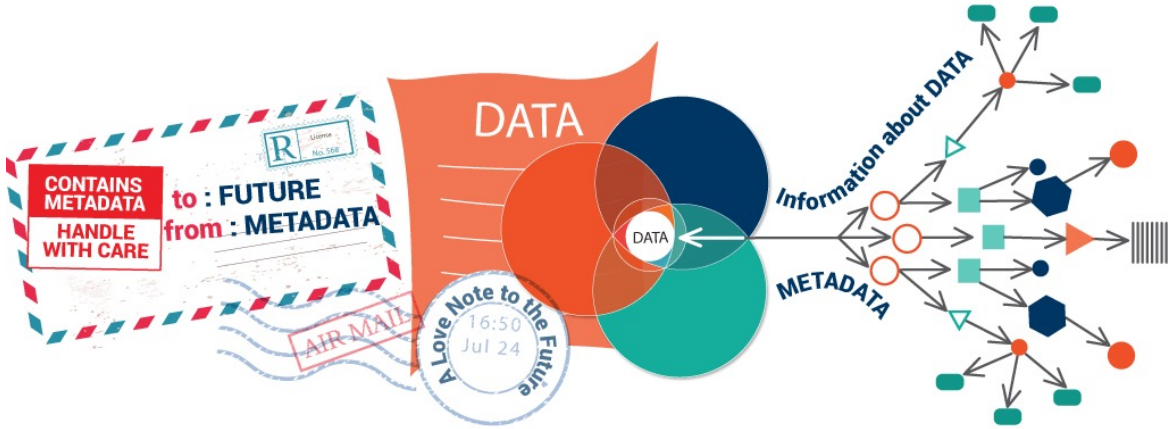
Dra. Eva Méndez
Universidad Carlos III de Madrid
Chair of the EU Open Science Policy Platform

Metadata in the DATA communication…

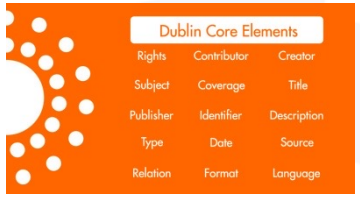
- When you **provide** data to someone else, what types of information would you want to include with the data to make them useful?

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

A simple one…



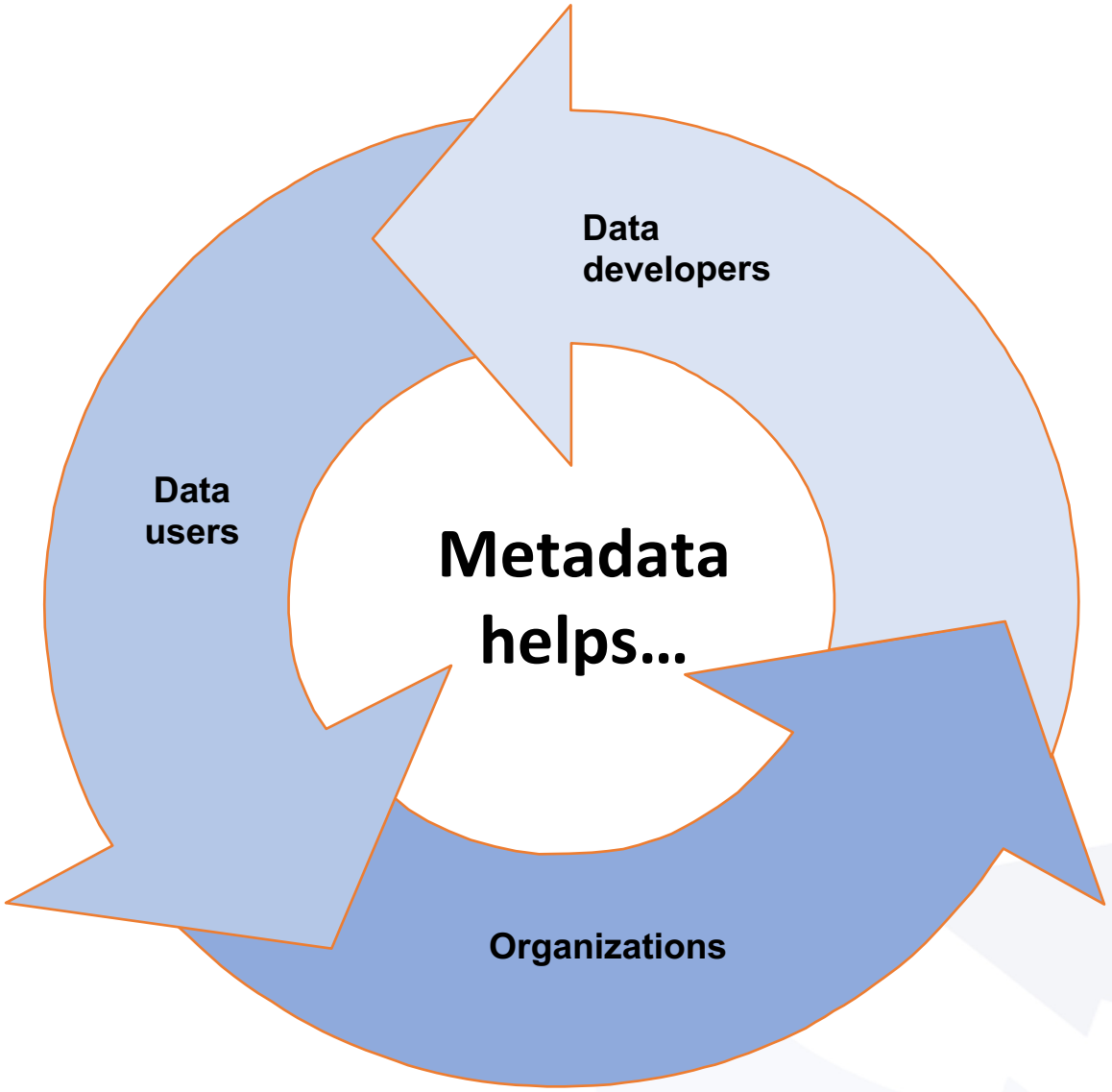
- **Metadata by definition:**
  - Data
  - Data about data
  - Data which provides information about a resource
- **Metadata by example:**
  - Title, author, subject, date, type, coordinates, (metadata elements - schemas) …
  - “physics”, “2004-01-23” (metadata values-schemEs)
  - Digital format, terms and conditions, location & PID



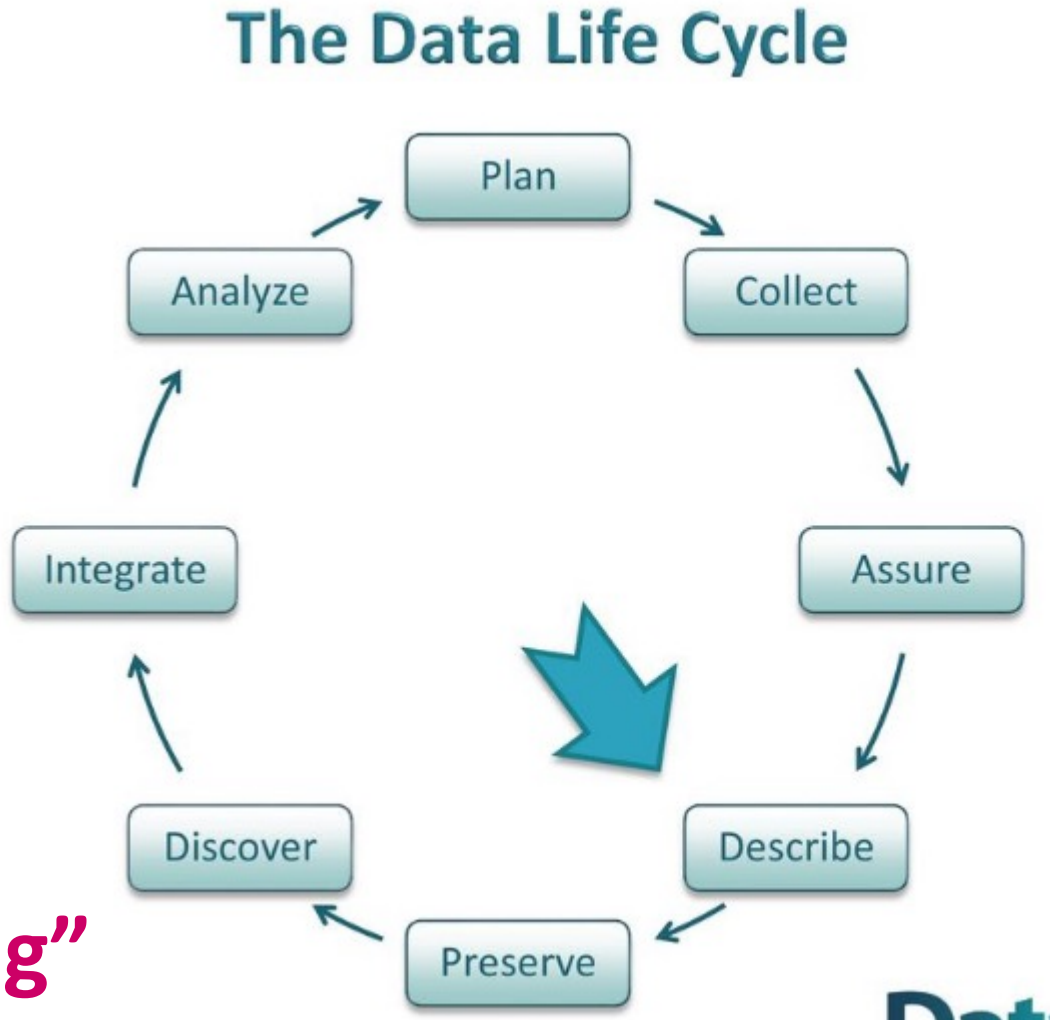
<ddi>

FAIRSFAIR
Fostering Fair Data Practices in Europe

The Value of Metadata



- **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”

Your DATA with(out) METADATA

Bad

Better…

Best
(Rich, Structured)

Nutrition Facts*

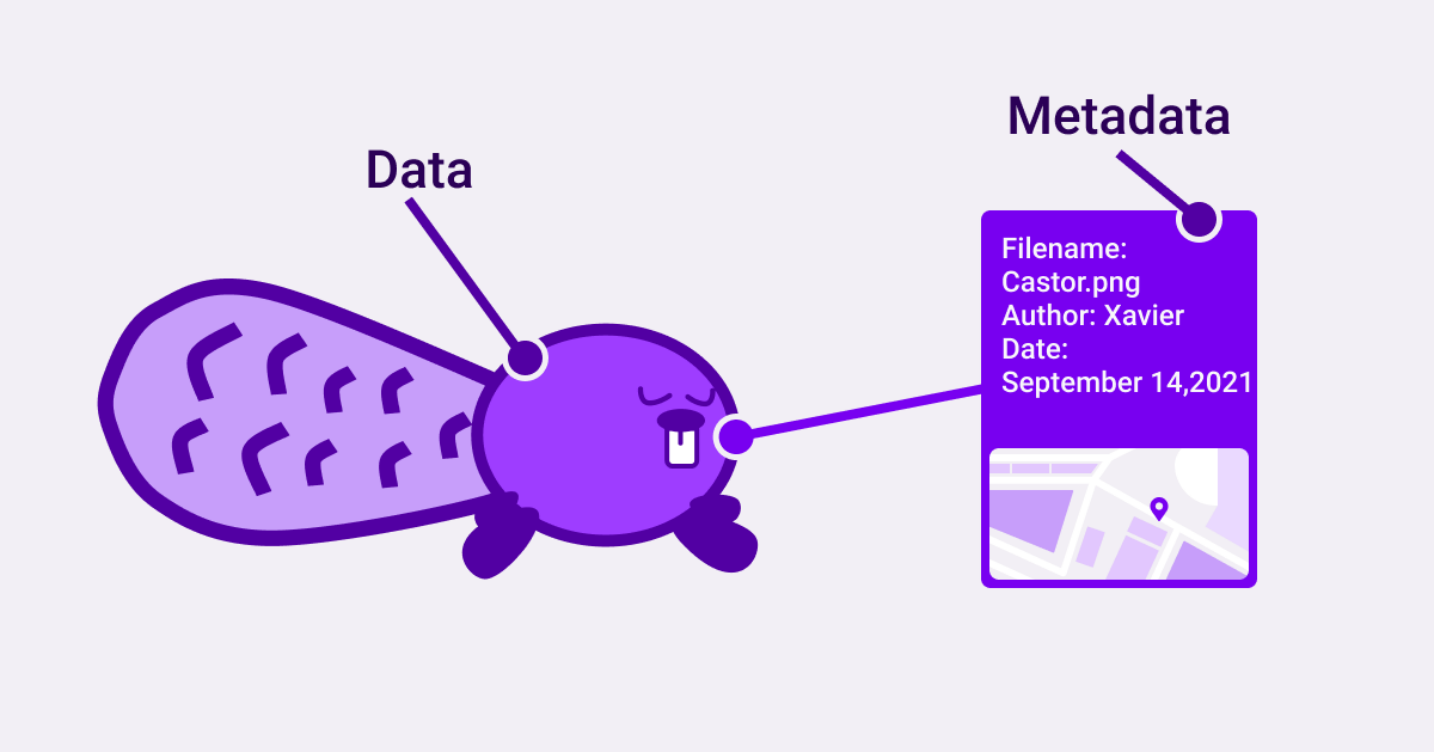
Amount Per Serving (serving size) = 1/2 cup (120mL) condensed soup

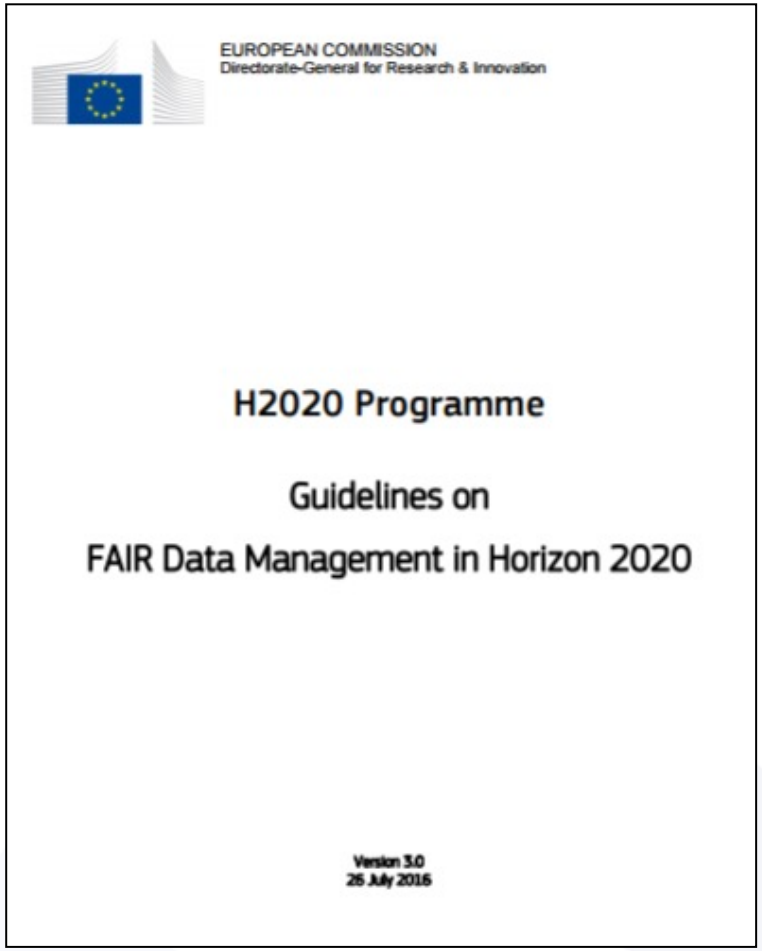
Calories 100	Dietary Fiber 2g
Fat Calories 50	Sugars 1g
Total Fat 6g	Protein 1g
Sat. Fat 1g	Potassium 60mg
Trans Fat 0g	
Polyunsat. Fat 4g	**% Daily Values****
Monounsat. Fat 1g	Vitamin A 0%
Cholesterol 5mg	Vitamin C 0%
Sodium 870mg	Calcium 0%
Total Carb. 9g	Iron 0%

* The nutrition information contained in this list of Nutrition Facts is based on our current data. However, because the data may change from time to time, this information may not always be identical to the nutritional label information of products on shelf.

** % Daily Values (DV) are based on a 2,000 calorie diet.

https://dmp.data.jhu.edu/metadata-for-effective-research-data-management




EUROPEAN COMMISSION
Directorate-General for Research & Innovation

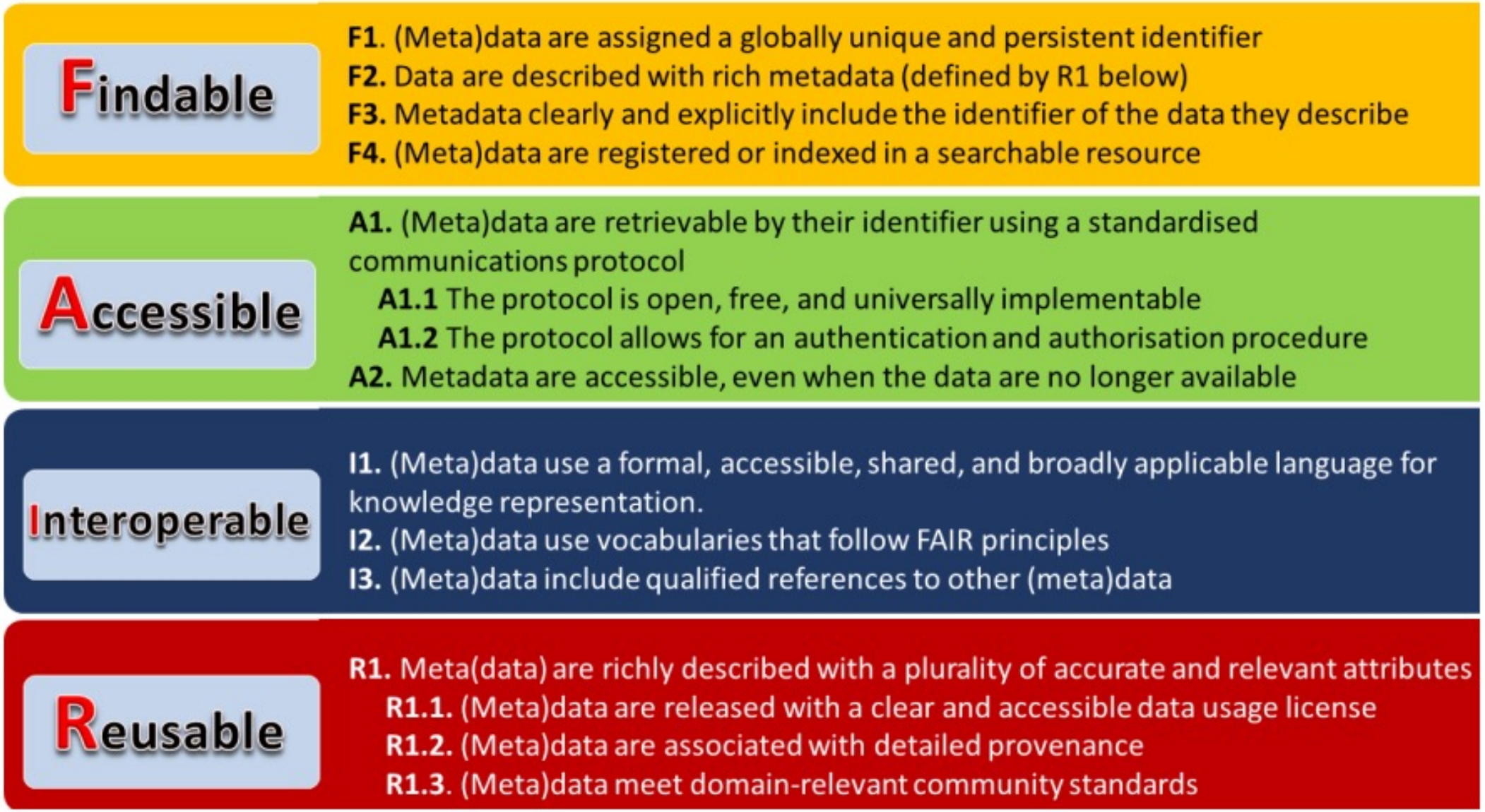
H2020 Programme

Guidelines on

FAIR Data Management in Horizon 2020

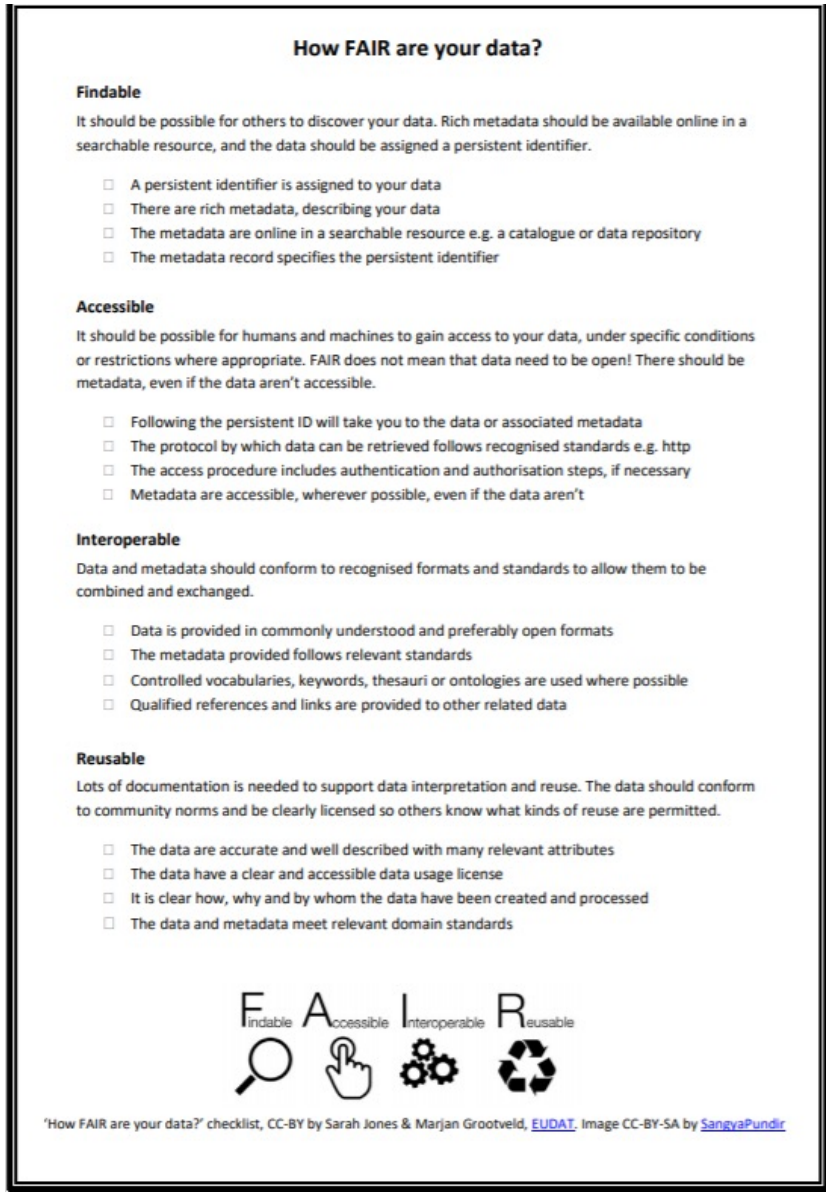
Version 3.0
26 July 2016

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf



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

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

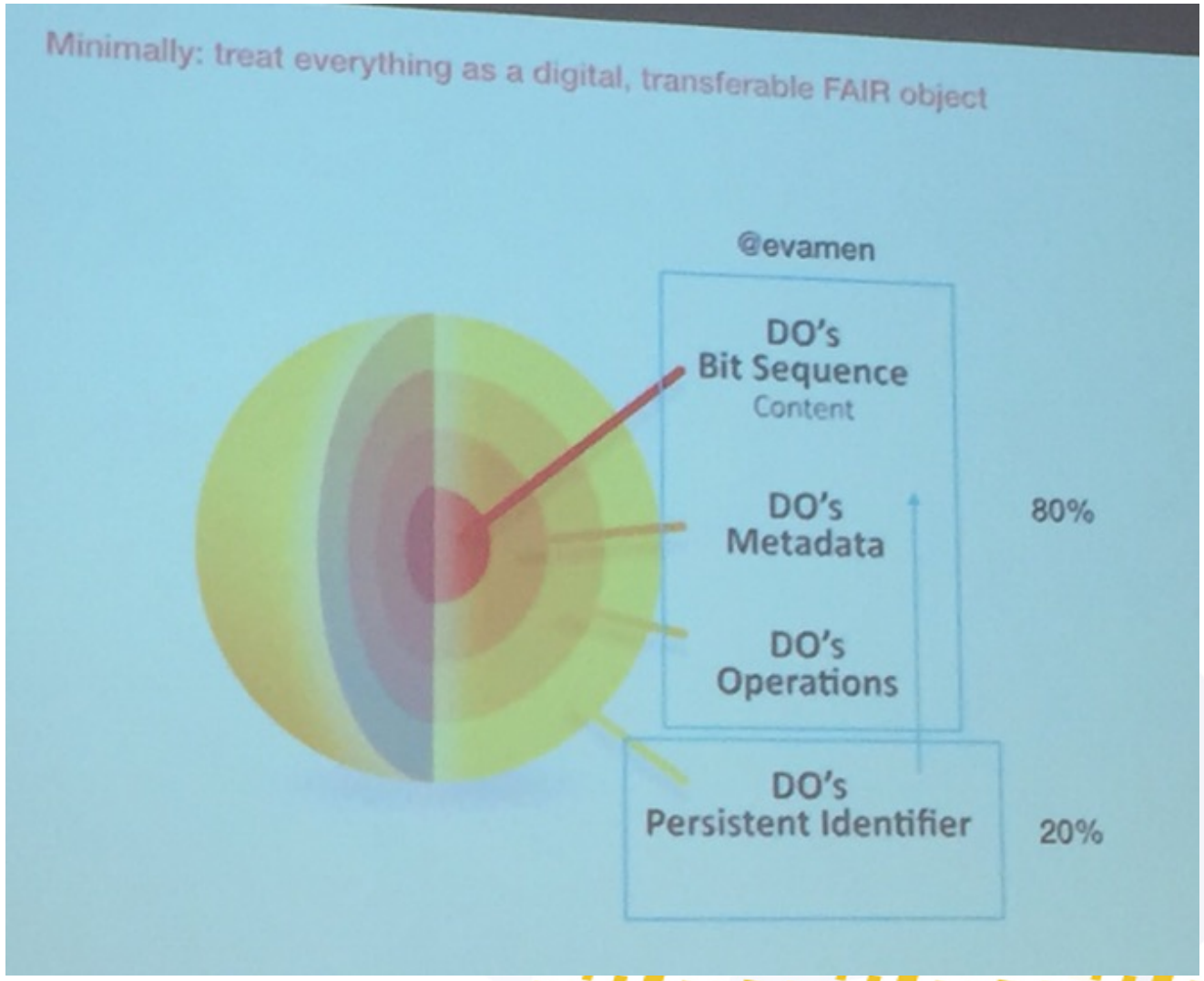
'How FAIR are your data?' checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT. Image CC-BY-SA by SangyaPundir

- *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.

FAIR Guiding Principles: all about metadata

80% metadata
20% PIDs

@



Metadata principles (CoolM4OS)

- The only difference between metadata & data is mode of use
- Metadata is not just for publications and data, it is also for users, software services, computing resources, etc.
- Metadata is not just for description and discovery; it is also for contextualization, (relevance, quality, provenance, restrictions (rights, costs) and for coupling users, software and computing resources to research outcomes
- Metadata must be machine-understandable as well as human understandable for autonomicity (formalism)
- Management (meta) data are also relevant (research proposal, funding, project information, research outputs, outcomes, impact, etc.)
- We must create: **“COOL” metadata**

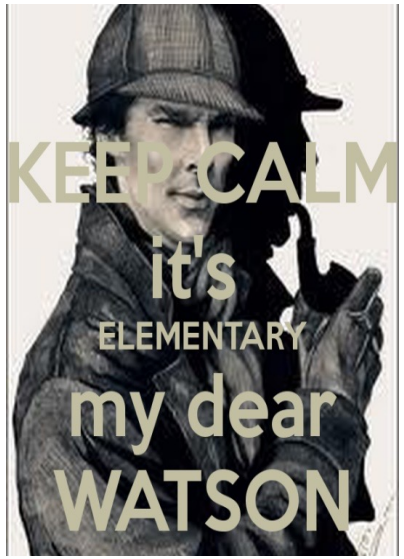
https://www.rd-alliance.org/system/files/documents/RDA%20Metadata%20Principles%20and%20their%20Use%2020141114.docx

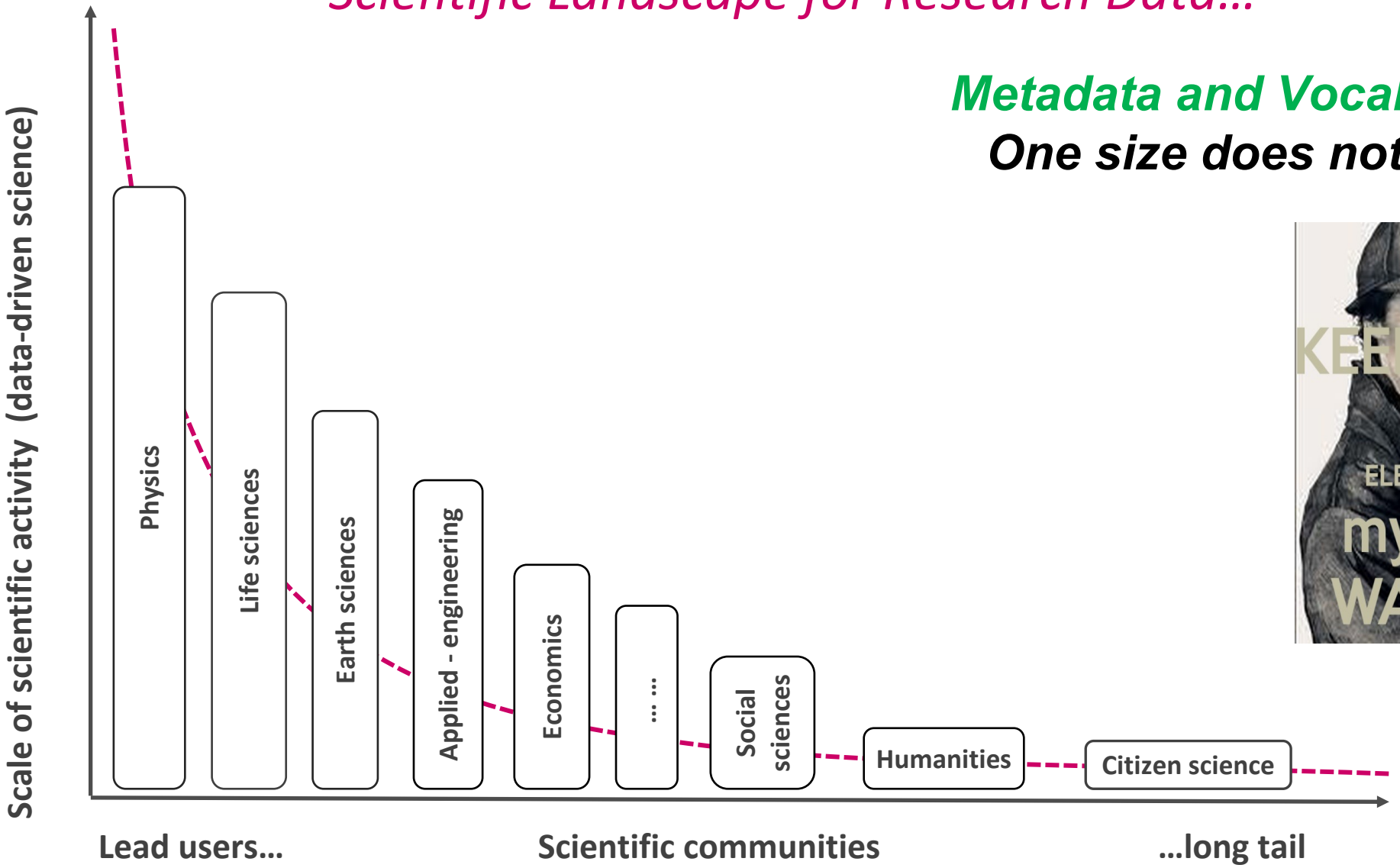
EOSC & FAIR DATA

Scientific Landscape for Research Data…

Metadata and Vocabularies

One size does not fit all!!!



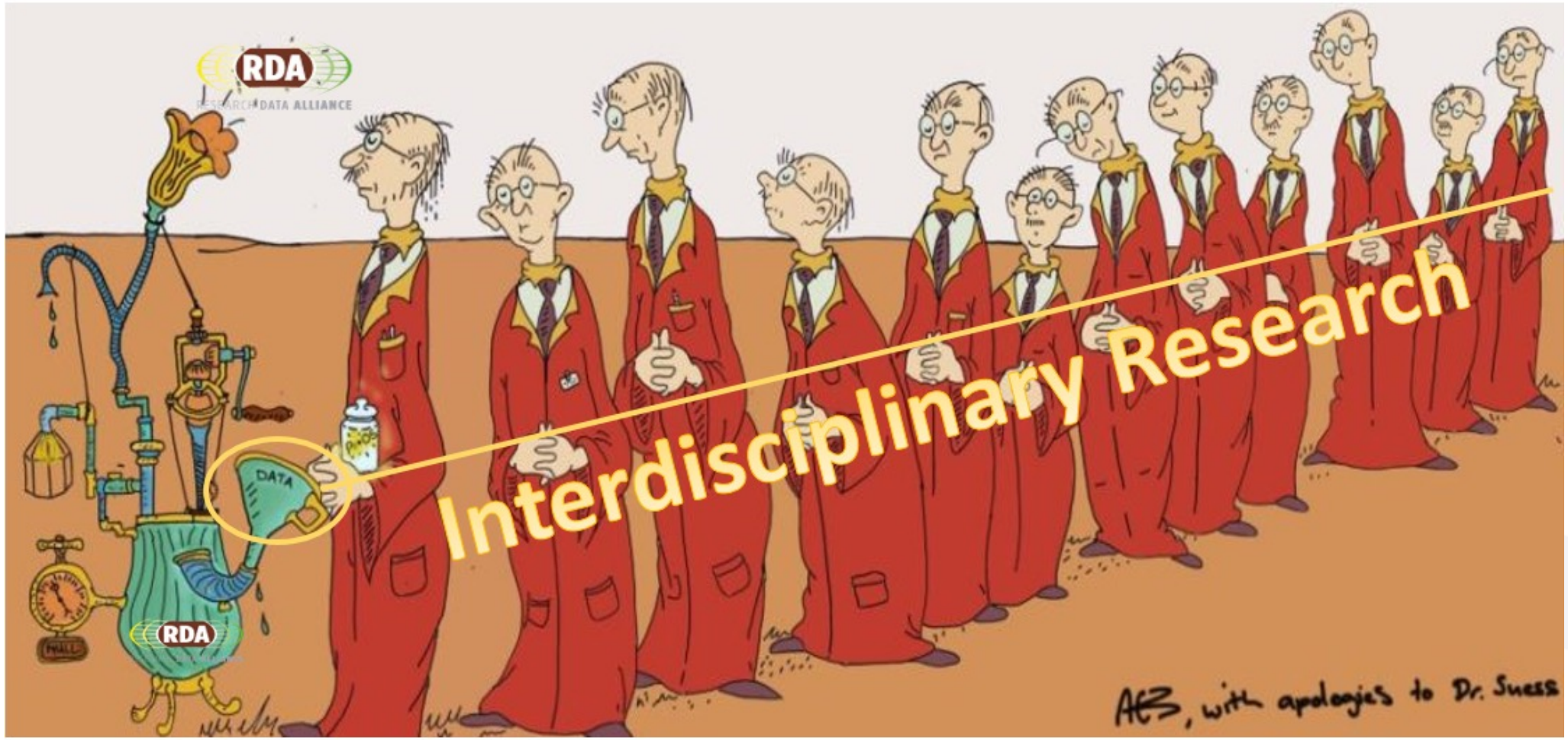


Back to the problem

Metadata are like toothbrushes...

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

https://www.rd-alliance.org/rda-disciplines/rda-interdisciplinary-research

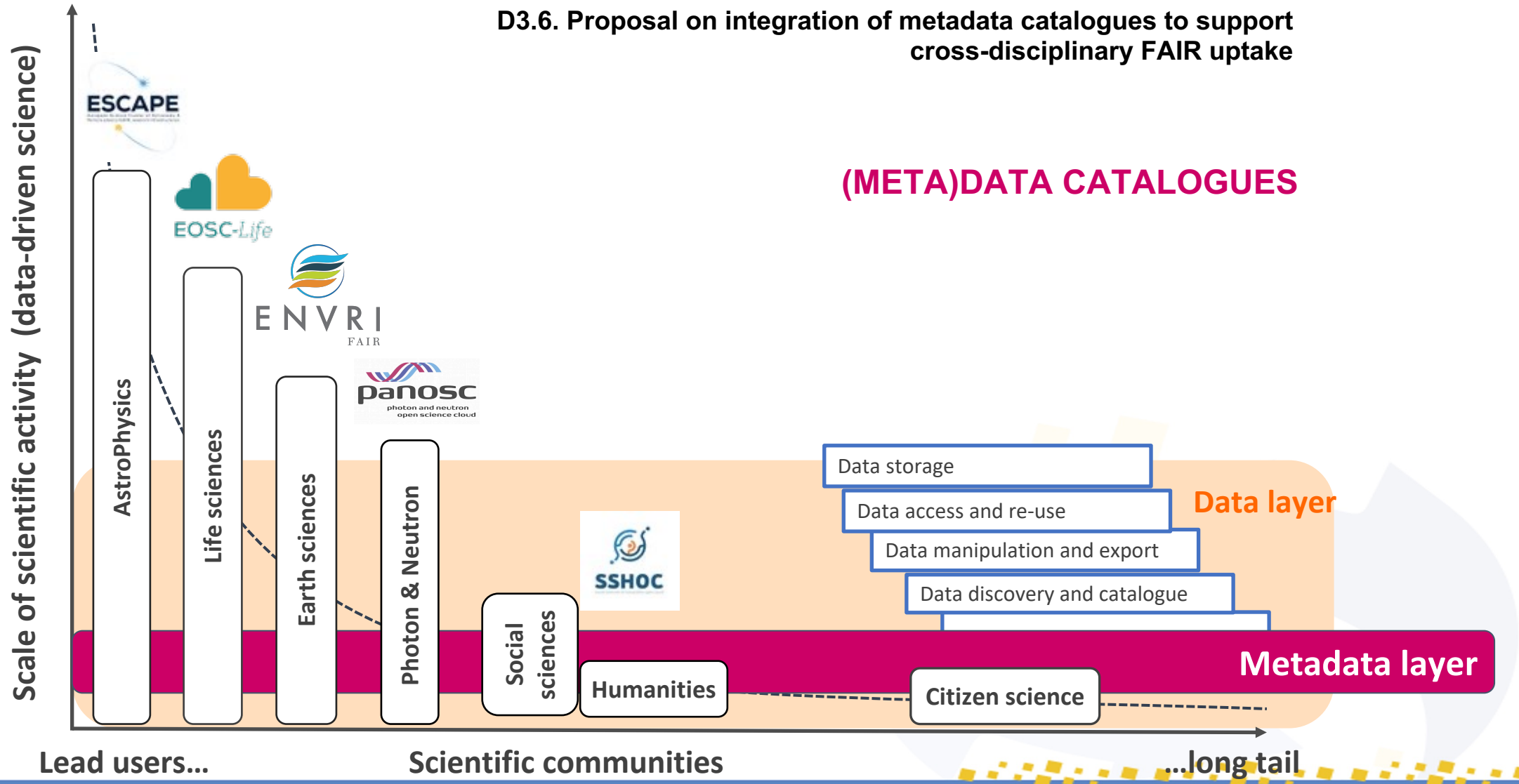


Source: https://imerg.info By Dr Seuss (with obvious modifications by Ashley Barnett contextualization by Eva Méndez

EOSC & FAIR DATA

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

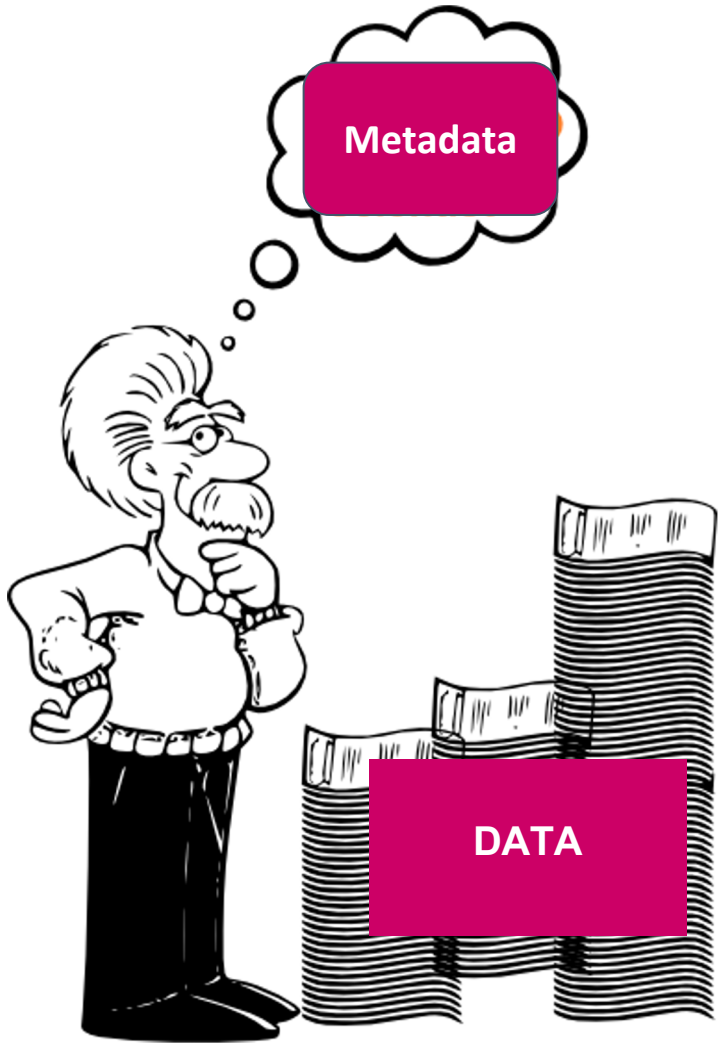
(META)DATA CATALOGUES



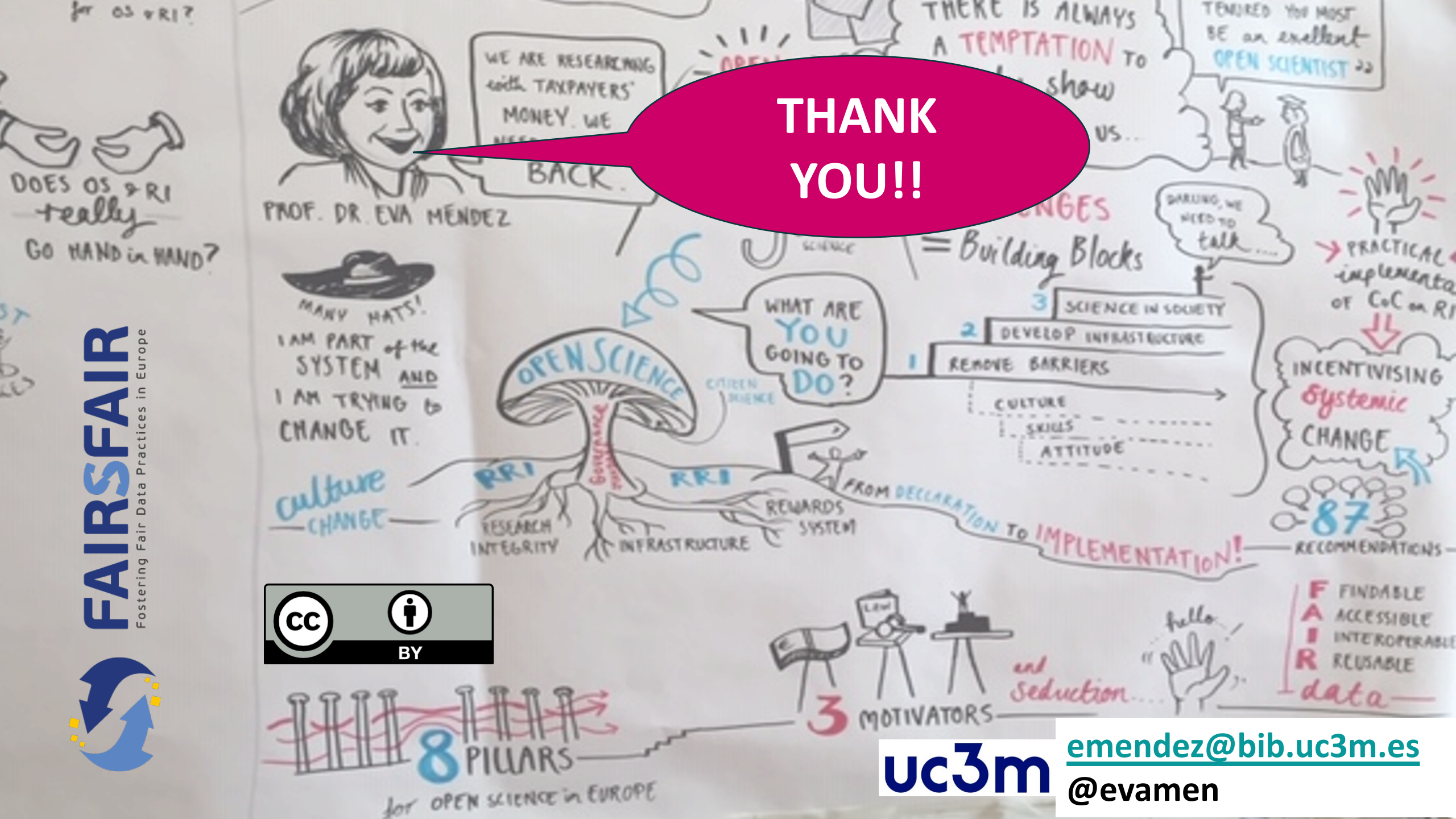
Based on Burgelman

Challenges

Metadata



- **Inter-cross disciplinary** research
- Metadata **standards**: avoid the *“toothbrush effect”*
- (Meta)data **catalogue integration**
- **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.



emendez@bib.uc3m.es
@evamen
uc3m