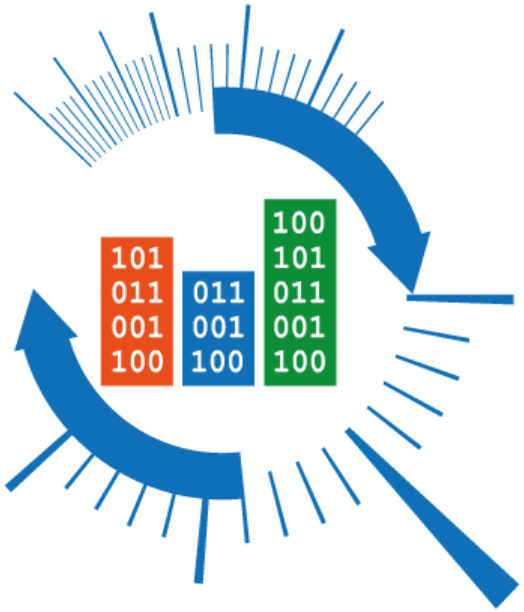



# RDM in EPM projects: Module 3 - Metadata

Jean-Laurent Hippolyte

Principal Scientist

National Physical Laboratory, UK

# Outline


1. What is metadata and how is it useful to RDM?
2. Specifying metadata requirements
3. Making datasets accessible

# Outline


1. What is metadata and how is it useful to RDM?
2. Specifying metadata requirements
3. Making datasets accessible

# What is metadata?

- Definition(s)
  - Data about Data
  - Structured information that facilitate retrieval, use or management of some information resource
- Everyday examples

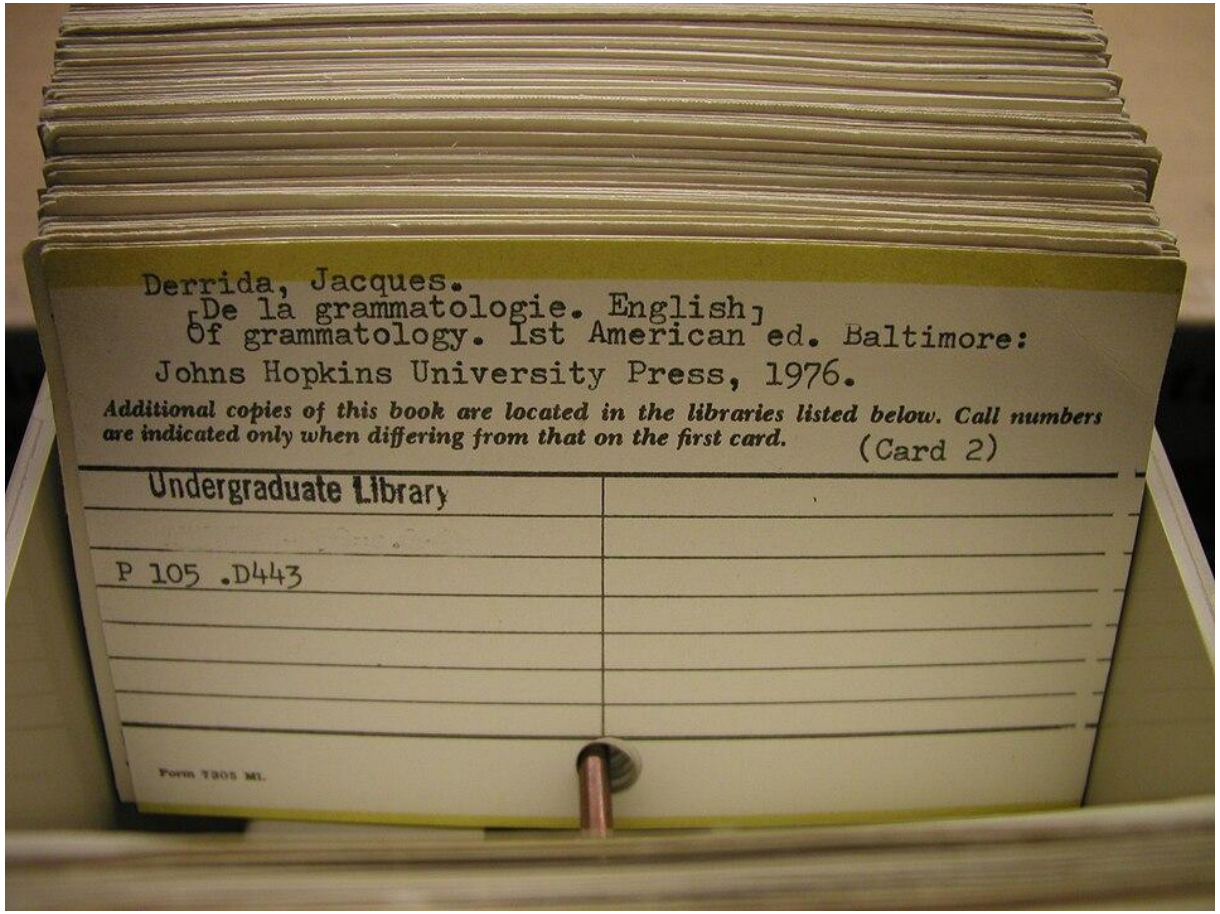
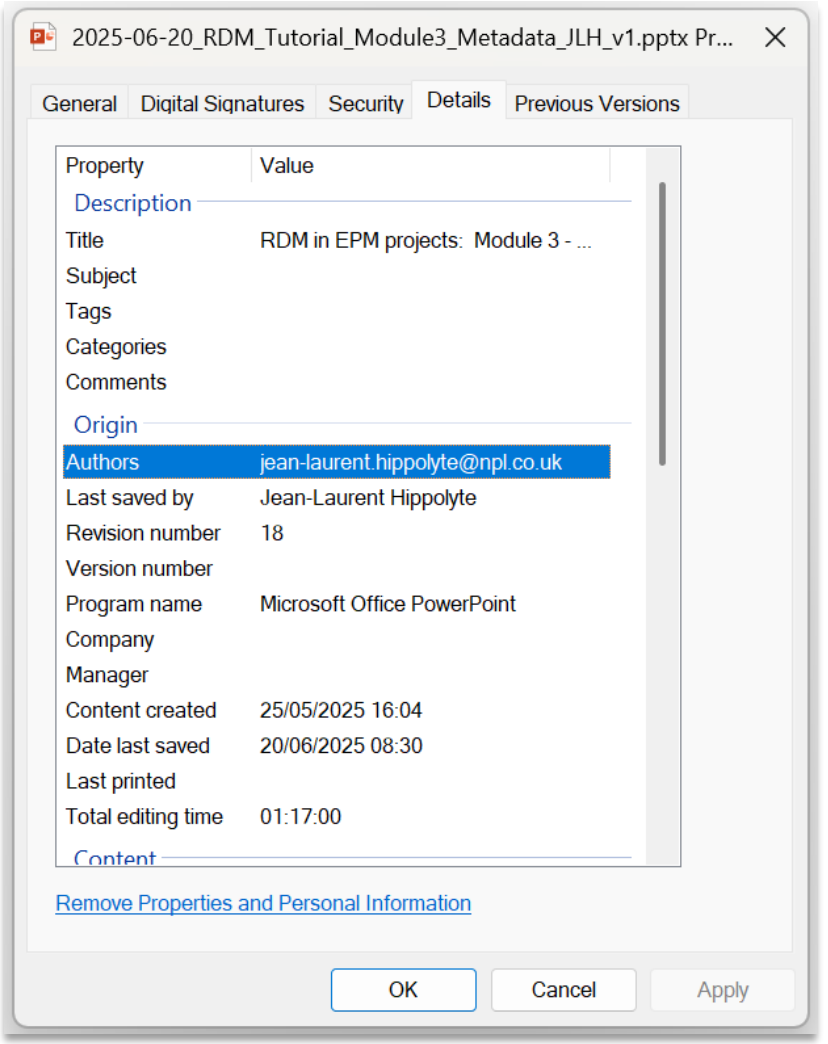



Photo by David Fulmer / CC BY

http://www.niso.org/publications/understanding-metadata-2017

# What is metadata?

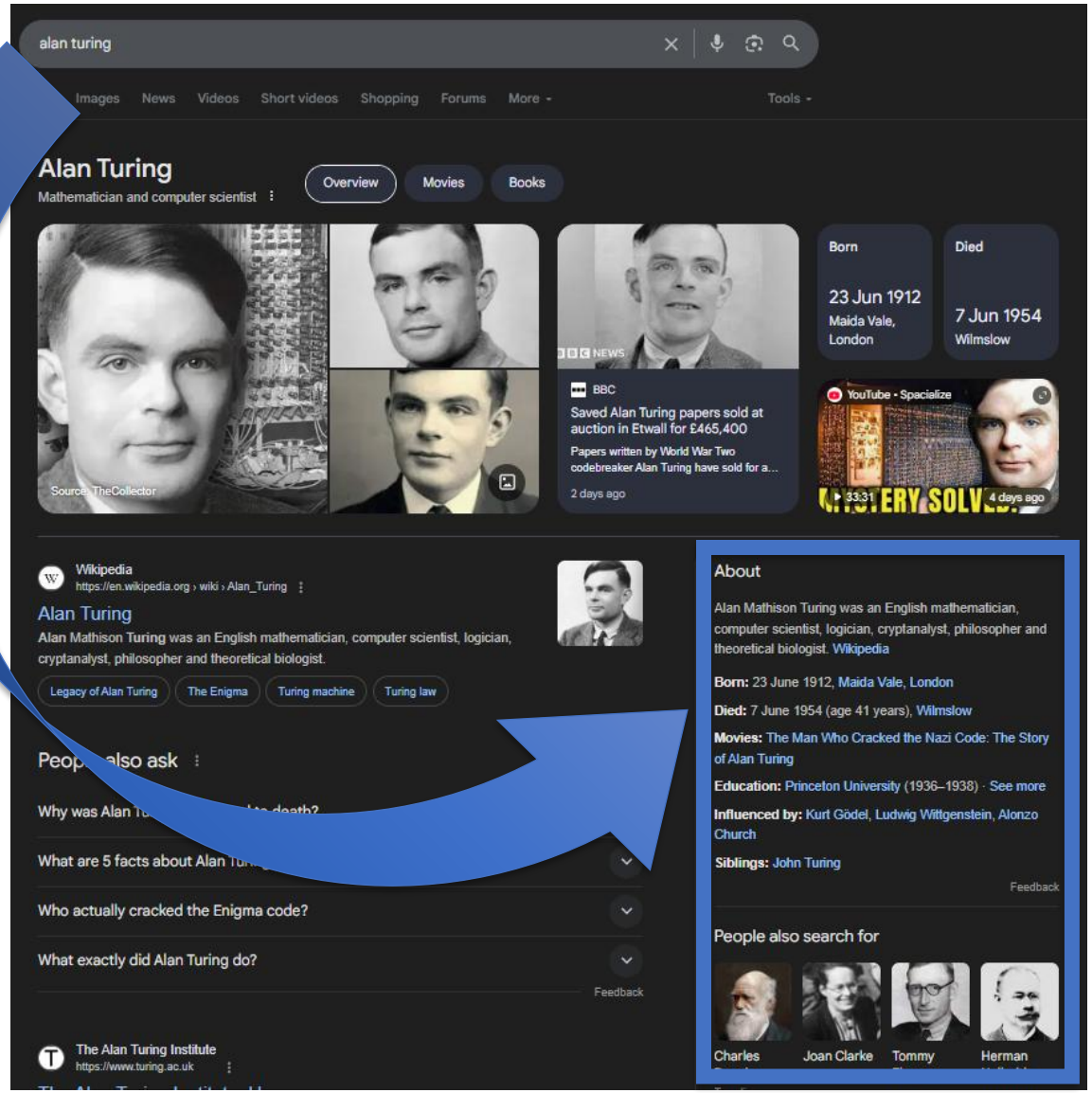
- Definition(s)
  - Data about Data
  - Structured information that facilitate retrieval, use or management of some information resource
- Everyday examples
  - File properties in Operating Systems




http://www.niso.org/publications/understanding-metadata-2017

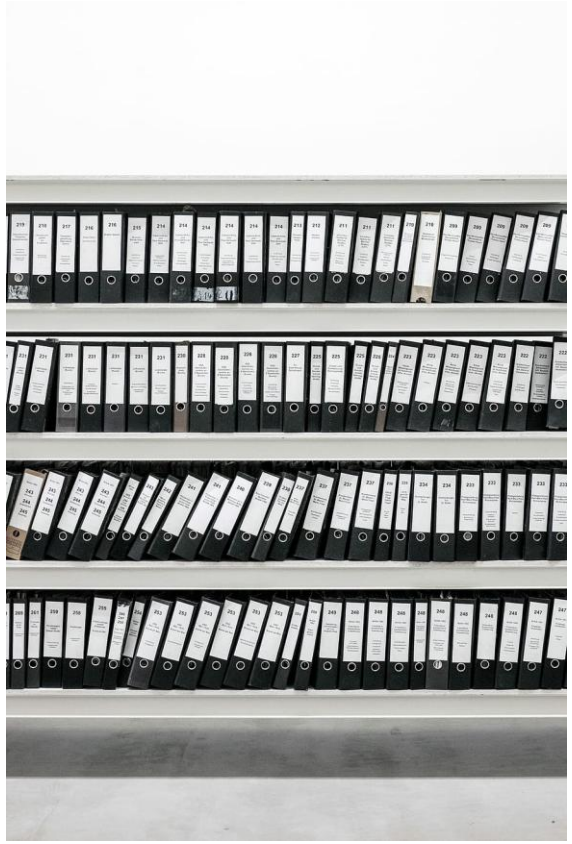
# What is metadata?

- Definition(s)
  - Data about Data
  - Structured information that facilitate retrieval, use or management of some information resource
- Everyday examples
  - File properties in Operating Systems
  - Google Knowledge Graph




http://www.niso.org/publications/understanding-metadata-2017

# What is metadata?

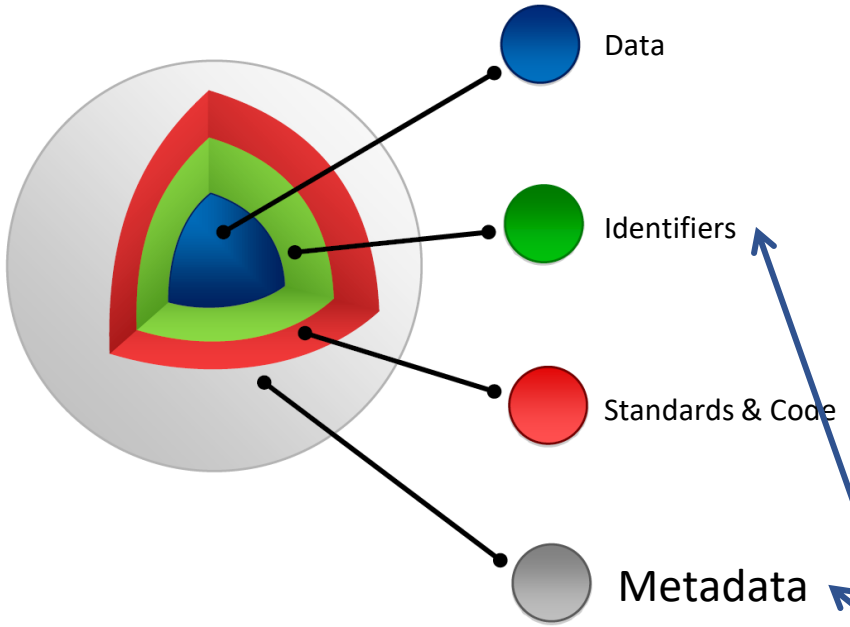


- Gaps in current practices
  - Ad-hoc data organisation
    - file/folder naming conventions
  - Unstandardised description
    - headers in spreadsheets
  - Knowledge embedded in human
    - Understanding of data lost due to employee turnover
    - Data unusable by third parties

https://www.data.cam.ac.uk/data-management-guide/organising-your-data

# How is it useful to RDM?

- Realization of FAIR relies on metadata
  - Findable, Accessible, Interoperable, Reusable




**Findable**
The first step in (re)using data is to find them. Metadata and data should be easy to find for both humans and computers. **Machine-readable metadata are essential** for automatic discovery of datasets and services.

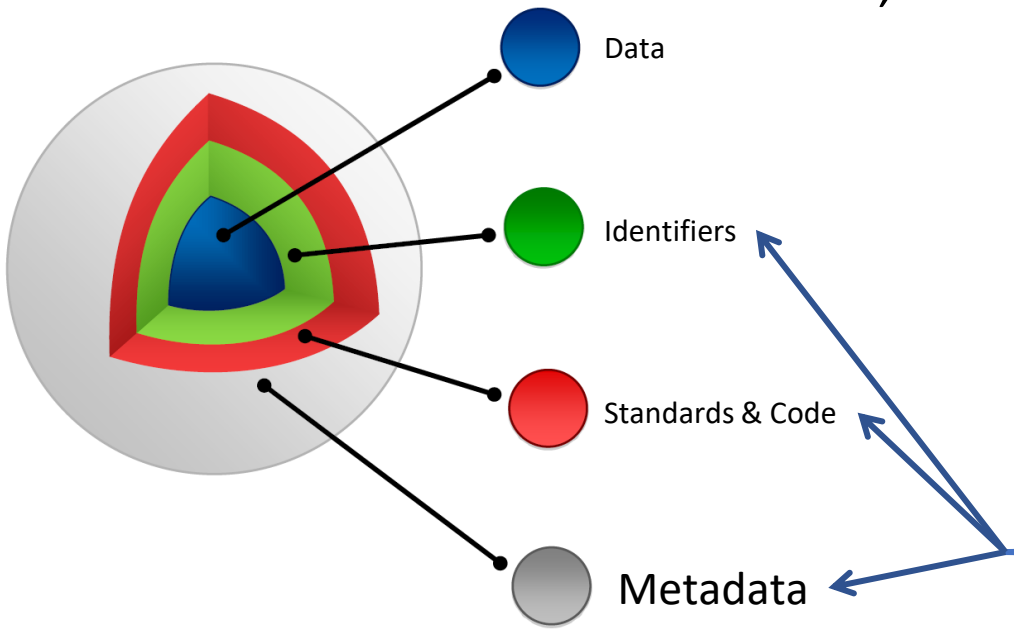
**F1. (Meta)data are assigned a globally unique and persistent identifier**
**F2. Data are described with rich metadata (defined by R1 below)**
**F3. Metadata clearly and explicitly include the identifier of the data they describe**
**F4. (Meta)data are registered or indexed in a searchable resource**

Hodson, Simon et al. 2018. FAIR Data Action Plan: Interim recommendations and actions from the European Commission Expert Group on FAIR data. (Jun. 2018). http://doi.org/10.5281/zenodo.1285290

https://www.go-fair.org/fair-principles/

# How is it useful to RDM?

- Realization of FAIR relies on metadata
  - Findable, Accessible, Interoperable, Reusable




**Accessible**

Once the user finds the required data, she/he/they need to know how they can be accessed, possibly including authentication and authorisation.

**A1. (Meta)data are retrievable by their identifier using a standardised communications protocol**

**A1.1 The protocol is open, free, and universally implementable**

**A1.2 The protocol allows for an authentication and authorisation procedure, where necessary**

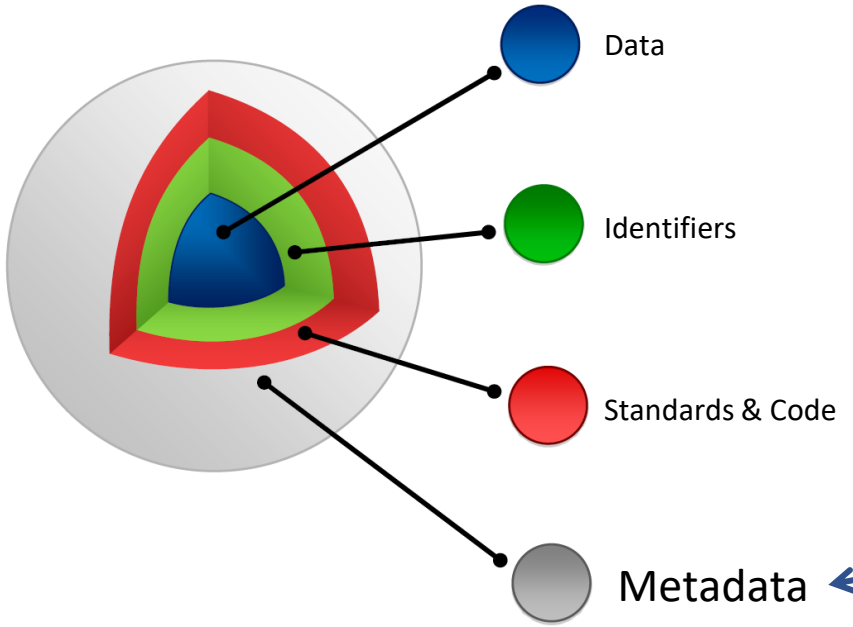
**A2. Metadata are accessible, even when the data are no longer available**

Hodson, Simon et al. 2018. FAIR Data Action Plan: Interim recommendations and actions from the European Commission Expert Group on FAIR data. (Jun. 2018). http://doi.org/10.5281/zenodo.1285290

https://www.go-fair.org/fair-principles/

# How is it useful to RDM?

- Realization of FAIR relies on metadata
  - Findable, Accessible, Interoperable, Reusable




**Interoperable**

The data usually need to be integrated with other data. In addition, the data need to interoperate with applications or workflows for analysis, storage, and processing.

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

**I2. (Meta)data use vocabularies that follow FAIR principles**

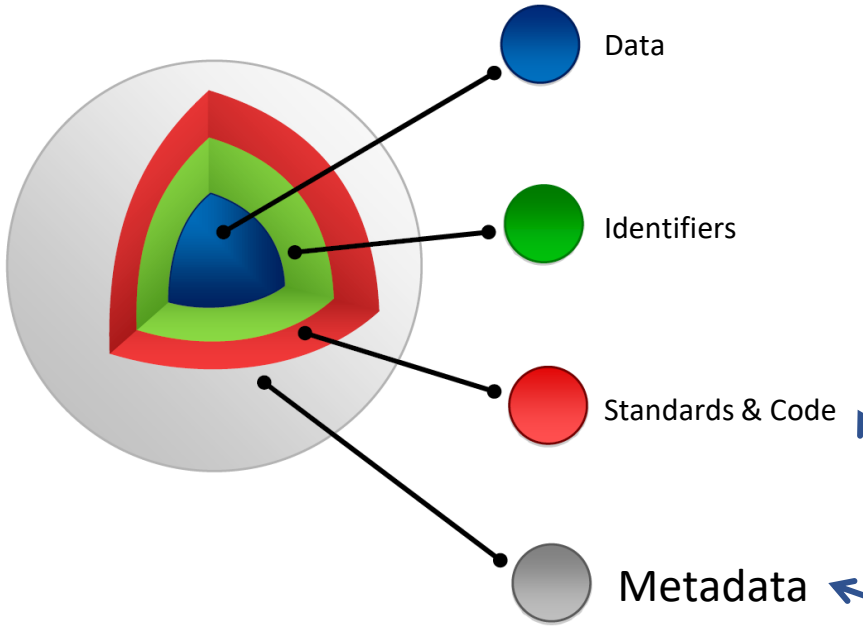
**I3. (Meta)data include qualified references to other (meta)data**

Hodson, Simon et al. 2018. FAIR Data Action Plan: Interim recommendations and actions from the European Commission Expert Group on FAIR data. (Jun. 2018). http://doi.org/10.5281/zenodo.1285290

https://www.go-fair.org/fair-principles/

# How is it useful to RDM?

- Realization of FAIR relies on metadata
  - Findable, Accessible, Interoperable, Reusable




**Reusable**

The ultimate goal of FAIR is to optimise the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

**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 detailed provenance**

**R1.3. (Meta)data meet domain-relevant community standards**

Hodson, Simon et al. 2018. FAIR Data Action Plan: Interim recommendations and actions from the European Commission Expert Group on FAIR data. (Jun. 2018). http://doi.org/10.5281/zenodo.1285290

https://www.go-fair.org/fair-principles/

# How is it useful to RDM?

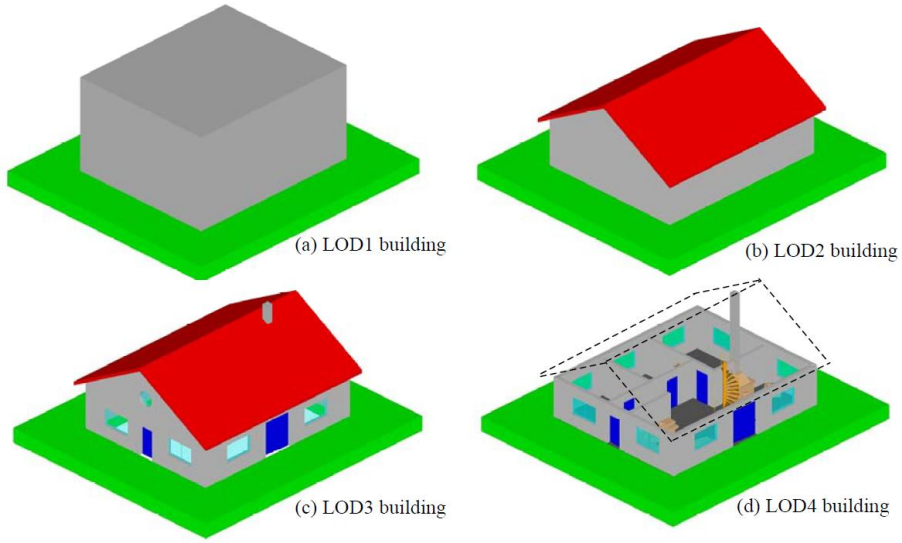
- EURAMET reporting guide
  - Mandates that
    - Data used for research validation is accompanied by metadata
      - "*rich information on the datasets (description, date of deposit, author(s), venue and embargo); the EMPIR funding; the project acronym and number; licensing terms; persistent identifiers for the dataset, the authors involved, and, if possible, their organisations and the project.*"
  - Recommends that
    - Datasets are shared via open access repositories, therefore made **searchable through metadata**
    - Metadata to comply with standard vocabularies or schemas, particularly metadata standards **specific** to the targeted discipline (if they exist…)



https://msu.euramet.org/downloads/documents/EMPIR_Reporting_Guidelines_Part9.pdf

# How is it useful to RDM?

- Leveraging the FAIR principles
  - Data quality
  - Provenance
  - Reproducibility
  - Transparency
  - Trustworthiness
- The more comprehensive the metadata, the more value added to data
- Level of details also depend on intended/potential reuse




http://www.ogc.org

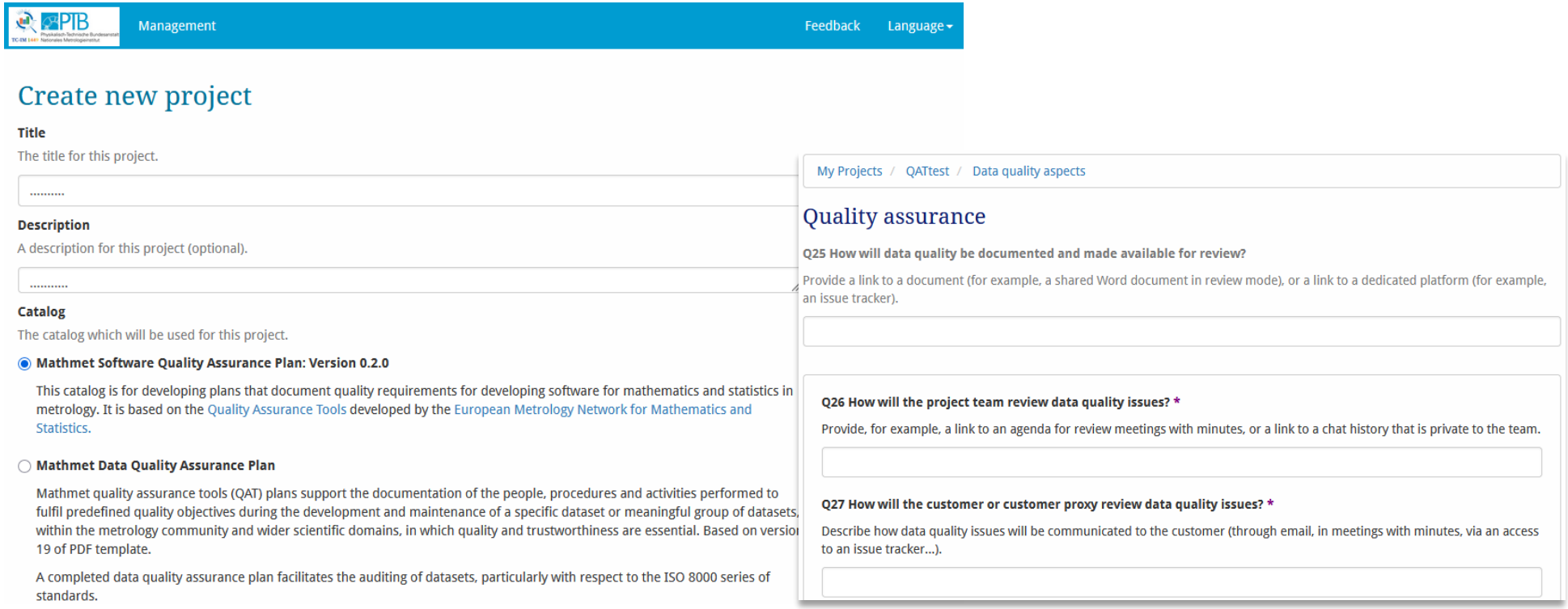
# How is it useful to RDM?

- Many desirable aspects of data quality can't be achieved without metadata:
  - believability, objectivity, reputation, relevancy, interpretability…
  - Mathmet Data Quality Assurance planning tool

https://www.euramet.org/european-metrology-networks/mathmet/activities/quality-assurance-tools

  - Available as an online interactive form on the MetRDMO platform

https://dmp.metrology-rdm.eu/



# Outline


1. What is metadata and how is it useful to RDM?
2. Specifying metadata requirements
3. Making datasets accessible

# Specifying metadata requirements

- Metadata requirements often formally described.
- Example: metadata for scientific papers
  - A BibTeX entry includes mandatory and optional tags which characterize a bibliographic reference (author, title, year, etc.)
  - Multiplicity of tags allows cross-checking of the reference

[1] P. J. Cohen. The independence of the continuum hypothesis. *Proceedings of the National Academy of Sciences*, 50(6):1143–1148, 1963.

```
@article{CitekeyArticle,
  author    = "P. J. Cohen",
  title     = "The independence of the continuum hypothesis",
  journal   = "Proceedings of the National Academy of Sciences",
  year      = 1963,
  volume    = "50",
  number    = "6",
  pages     = "1143--1148",
}
```

# Specifying metadata requirements

Web form


Machine-interpretable metadata

```
<?xml version="1.0" encoding="UTF-8"?>

<dc:title>Metadata for RDM and publications submission for EMPIR projects</dc:title
>
<dc:creator>Jean-Laurent Hippolyte</dc:creator>
<dc:creator>Julia Neumann</dc:creator>
<dc:subject>Metadata</dc:subject>
<dc:subject>Research Data</dc:subject>
<dc:description>Brief overview of metadata for scientific datasets</dc:description>
<dc:publisher>EURAMET TC-IM 1449</dc:publisher>
<dc:date>11/03/2021</dc:date>
<dc:type>Presentation</dc:type>
<dc:format>Microsoft PowerPoint</dc:format>
<dc:source>https://www.euramet.org/</dc:source>
<dc:language>en</dc:language>
<dc:rights>https://creativecommons.org/licenses/by/4.0/</dc:rights>
```

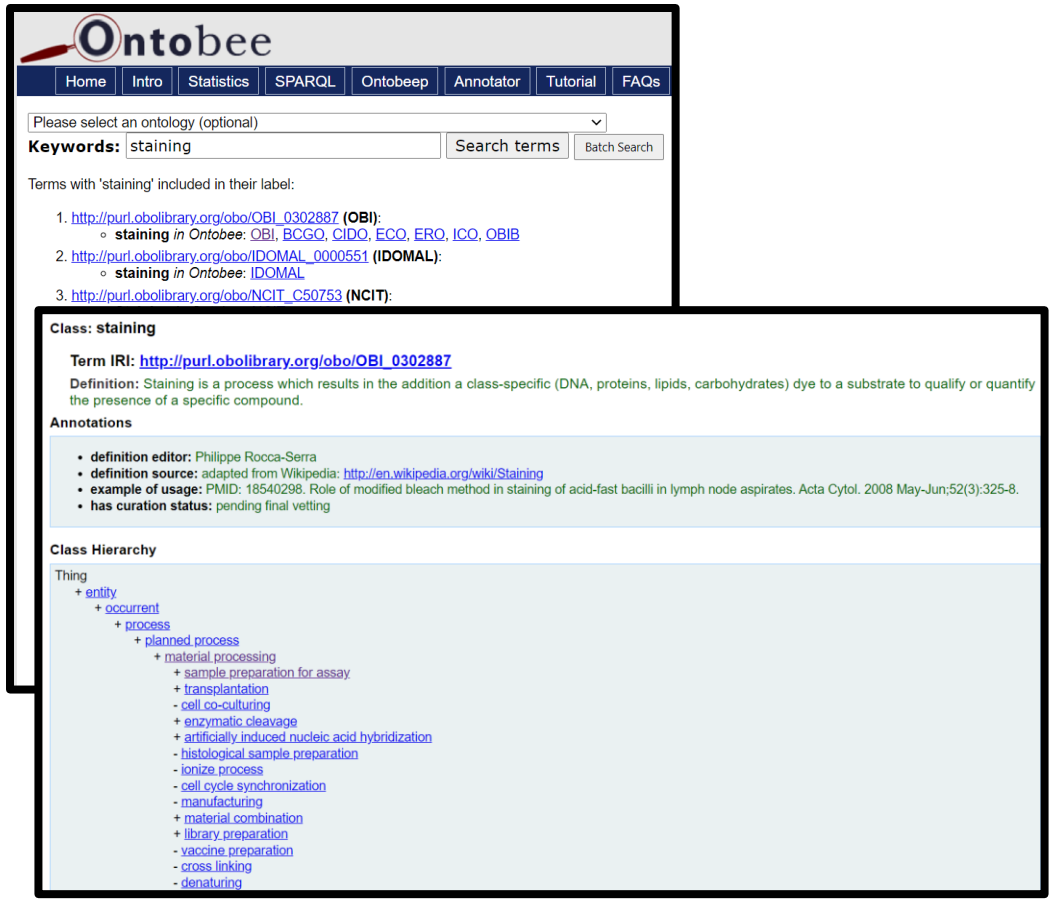
Generated using an online generator:
https://nsteffel.github.io/dublin_core_generator

- Metadata capture as part of the experimental process log

Following talk

# Specifying metadata requirements

- Example of domain-specific metadata schema(s):
  - Open Biological and Biomedical Ontology (OBO) Foundry
  - Metadata concept search engine (OntoBee)
- SI and CIPM MRA
  - SI Digital Framework




Following talk

http://www.obofoundry.org/

# Outline

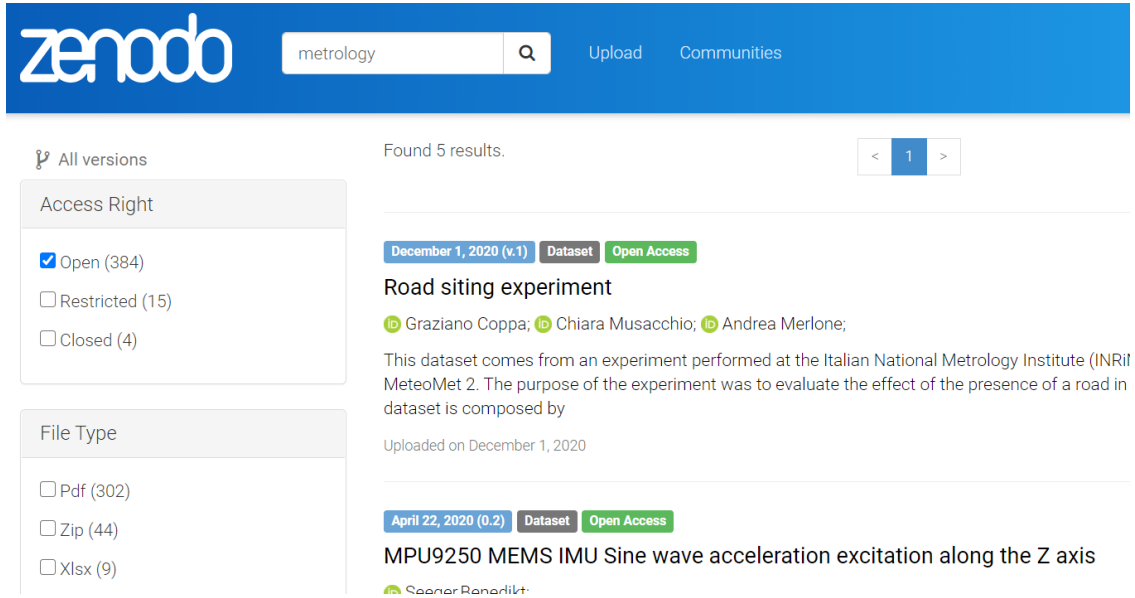

1. What is metadata and how is it useful to RDM?
2. Specifying metadata requirements
3. Making datasets accessible

# Making datasets accessible

- **Zenodo** an open-access repository hosted by CERN
- Attempts to comply with FAIR principles as best as possible
- But some compliance aspects are up to the users!

**Zenodo** provides online tools to:

- assign and resolve dataset persistent identifiers (DOIs)
- generate basic metadata
- search datasets through cross-domain metadata




https://about.zenodo.org/principles/

# Making datasets accessible

- **DataCite** a not-for-profit organization
- Aims to improve data citation for :
  - accessible research data
  - transparent and reproducible research
- Datacite provides online tools to:
  - assign and resolve dataset persistent identifiers (DOIs)
  - generate metadata
  - search datasets through cross-domain metadata

Find what you're looking for by searching millions of records with extensive, reliable metadata.

Share your data and reuse the data of others to create the highest impact in the research community.

Cite your research sources with confidence, and receive proper credit when your work is reused.

Connect your research – publications, datasets, software, authors, institutions, and funding data all in one place.

https://datacite.org/

# Making datasets accessible

- Datacite also provides an online tool to identify what **online repository** is right for your dataset according to:
  - Topic
  - Content type (text, database, source code…)
  - Country


https://www.re3data.org/browse/

# Thank you!


https://www.w3.org/DesignIssues/LinkedData

## Thank you for your attention!

Thanks to Julia Neumann, Federico Toro Grasso and TC-IM 1449 members who helped prepare and review this presentation.

# APPENDIX

- More resources about metadata standards and scientific data sharing:
  - Research Data Alliance
    - https://www.rd-alliance.org/
  - FAIRSharing search metadata standards
    - https://fairsharing.org/
  - CODATA
    - https://codata.org/
  - CASRAI RDM glossary
    - https://casrai.org/