HELMHOLTZ METADATA COLLABORATION

Needs and activities around semantic interoperability in the physical sciences

NFDI-ErUM data workshop

Oonagh Mannix
HMC Hub Matter

HZB Helmholtz Zentrum Berlin

HMC: Helmholtz Metadata Collaboration

- Make Helmholtz Data FAIR - findable, accessible, interoperable and reusable
- Work across scientific boundaries to support efficient metadata handling
- Close collaboration with NFDI and initiatives outside Helmholtz

Turning FAIR into reality* to enable reuse of data on all levels

** From: Turning FAIR into Reality, Final Report and Action Plan from the European Commission Expert Group on FAIR Data , doi: 10.2777/1524*

Semantic interoperability in matter

Variable I: intensity
Variable I: current

Variable A: Angstrom
Variable A: Ampere

Variable au: arbitrary units
Variable au: astronomical units

And many more…….

Need to make sure we* are talking about the same thing

*humans and machines

Interoperability

The potential for two independent agents to work on the same data in a coordinated manner.

https://purls.helmholtz-metadaten.de/hob/HDO_00007010

FAIR: Findable, Accessible, Interoperable, Reusable

Findable

F1 metadata are assigned a globally unique and eternally persistent identifier.

F2 data are described with rich metadata.

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

F4 metadata are registered or indexed in a searchable resource.

Accessible

A1 metadata are retrievable by their identifier using a standardized communications protocol.

A2 metadata are accessible, even when the data are no longer available

Interoperable

I1 metadata use a formal, accessible, shared and broadly applicable language for knowledge representation.

I2 metadata use vocabularies that follow FAIR principles

I3 metadata include qualified references to other metadata

Reusable

R1 metadata are richly described with a plurality of accurate and relevant attributes.

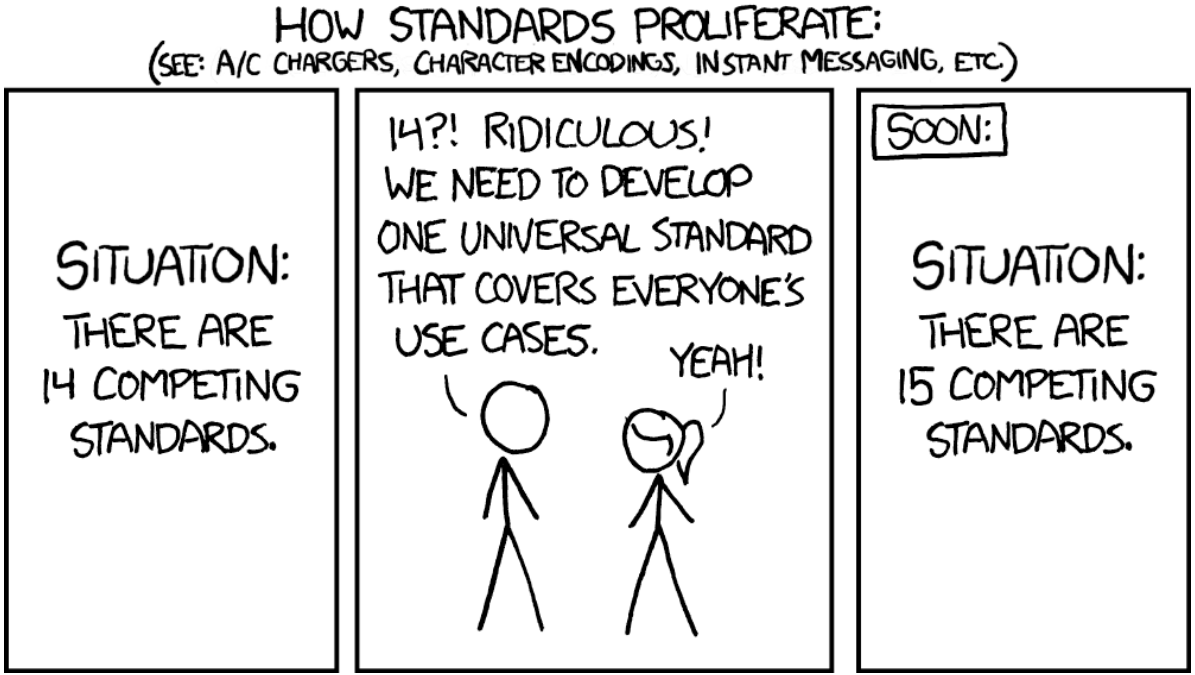
R2 metadata are released with a clear and accessible data usage license.

R3 metadata are associated with detailed provenance

R4 metadata meet domain-relevant community standards.

Wilkenson et al. *Scientific Data* 2016 https://doi.org/10.1038/sdata.2016.18

Standards



https://xkcd.com/927/

Q: Which international standards do you use?

NeXus

Core Scientific Metadata Model (CSMD)
icatproject-contrib.github.io/CSMD/

Crystallographic Information File

Filename extension	.cif
Internet media type	chemical/x-cif
Type of format	chemical file format
Extended from	Self-defining Text Archive and Retrieval
Extended to	mmCIF
Website	www.iucr.org/resources/cif

PDB

Filename extension	.pdb, .ent, .brk
Internet media type	chemical/x-pdb
Type of format	chemical file format

QCDml (guage field configurations)
hpc.desy.de/ildg/specifications/

ISO

openPMD

www.openpmd.org

7 named standards from 125 replies

*** HMC Community Survey 2021 - A survey on research data management practices among researchers in the Helmholtz Association.**
https://doi.org/10.3289/HMC_publ_05

A non-exhaustive list of activities around……

- FAIRmat (NeXus)
- DAPHNE
- PUNCH4NFDI
- EM Glossary
- SECoP
- HMC glossary
- ELN consortium
- CERN (standards at individual experimnets e.g. CMS, ATLAS)
- HMC projects (PATOF, HELPMI….)
- PaNnet ontology
- EuroLabs project
- ……

Findable

Findable

F1 metadata are assigned a globally unique and eternally persistent identifier.

F2 data are described with rich metadata.

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

F4 metadata are registered or indexed in a searchable resource.

- Need to consider mid/high-level standards to improve **findability**
  - e.g. PaN net ontology
  - Use case: SearchAPI from ExPANDS/PaNOSC (https://data.panosc.eu/)
  - Can consider filling additional gaps here

Wilkenson et al. *Scientific Data* 2016 https://doi.org/10.1038/sdata.2016.18

Reusability

- Domain specific extensions with high granularity to improve reusability also required
  - e.g. FAIRmat work with NeXus
  - Consider interoperability
  - Workflows are critical to achieve impact

Reusable

R1 metadata are richly described with a plurality of accurate and relevant attributes.

R2 metadata are released with a clear and accessible data usage license.

R3 metadata are associated with detailed provenance

R4 metadata meet domain-relevant community standards.

Wilkenson et al. *Scientific Data* 2016 https://doi.org/10.1038/sdata.2016.18

Things to consider

- Need to consider the purpose of each standard
  - And how connects to other standards
- Interoperability is key!
  - How do we achieve this?
  - How do we maintain this?
  - Need to consider transverse (between disciplines/techniques) and vertical (high/mid/low) interoperability
- Sustainability and governance
  - Survival after project funding ends
  - Clearly defined processes/workflows so survival beyond individuals

Thank you

Get in touch …

Oonagh Mannix - oonagh.mannix@helmholtz-berlin.de

Group page …

https://helmholtz-metadaten.de/en/matter/contact-us

Twitter …

@helmholtz_hmc / #HMCMatter

Mattermost

https://mattermost.hzdr.de/hmc-public/channels/town-square