The NSF MPS FAIR HACKATHON

Views and Reflections by Maria Cruz
VU Amsterdam

FAIR Hackathon for Mathematics and the Physical Sciences Research Communities 27-28 Feb 2019

The workshop

Organised by University of Notre Dame. Funded by the NSF. Held in Alexandria, VA, USA

- Get expertise from Europe to the US
- Pair FAIR experts with researchers
- Mix of ideas, demonstrations and hands-on sessions
- Gap analysis to accelerate further adoption of FAIR in the US

FAIR Hackathon for Mathematics and the Physical Sciences Research Communities 27-28 Feb 2019

The Participants (52 in total)

- Researchers, mainly from the US and in the areas of
  - Experimental High Energy Physics
  - Chemistry
  - Materials Science
- FAIR experts and service providers
- Librarians and research software engineers
- Funders and program managers

There to share skills, tools and techniques for FAIR research data and software in the physical sciences.

A pair from TU Delft & VU Amsterdam

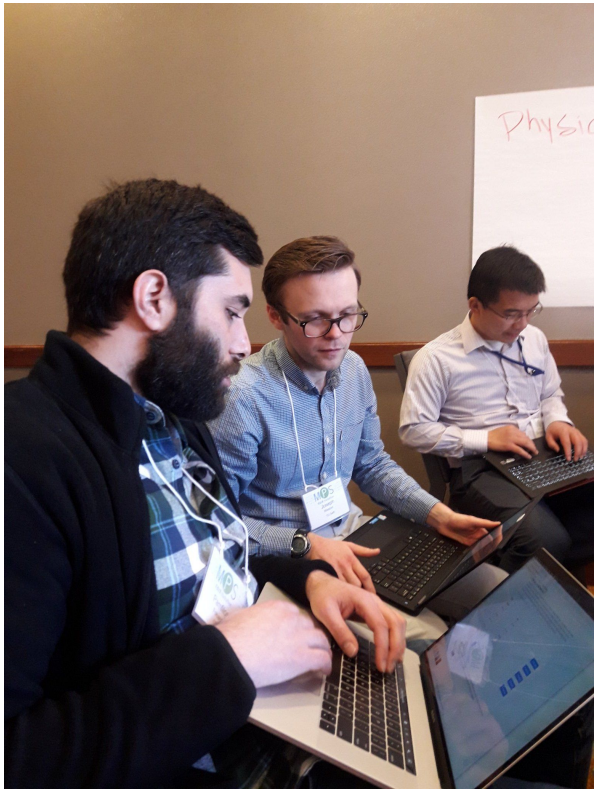
Maria Cruz

Community Manager Research Data, Library of the VU Amsterdam
Previously worked at the TU Delft Library, Research Data Services
Research background in astronomy and space sciences

Joseph Weston

Research Software Engineer and Data Champion at TU Delft
Research background is in theory and simulation of quantum transport in nanoelectronic circuits

Here we are in action during the workshop!



Some initial impressions

Mikala Narlock
@mikalaraethelib

First day of the #MPSFAIRHackathon in the books! It was an incredibly busy day, full of meeting brilliant domain-experts and visionary FAIR specialists (who are sometimes both!). Many thanks to @nkmeyers for helping me get here.

Maria J. Cruz
@gravana

#MPSFAIRHackathon George Strawn: The intellectual excitement around FAIR data is similar to the excitement around the Internet 30 years ago. FAIR Data will be a core infrastructure in the future (2030?).

3:09 PM · Feb 28, 2019 · Twitter Web App

Gordon Watts
@SeattleGordon

Here near DC attending the #MPSFAIRHackathon. Day one really highlighted efforts for various subfields to establish metadata to communicate results digitally. Chemistry folks seem to really have their house in order.

2:29 PM · Feb 28, 2019 · Twitter Web App

Maria J. Cruz
@gravana

#MPSFAIRhackathon Aside from FAIR, the word 'triple', as in a semantic RDF triple, may be the word I'm hearing the most in this Hackathon!

3:50 PM · Feb 28, 2019 · Twitter Web App

And a haiku

FAIR: an ambitious vision to be achieved through metadata

The Grand Vision of FAIR Data

The Internet and FAIR Data

"The internet solved the interoperability of heterogeneous networks problem. FAIR data's aspiration is to solve the interoperability of heterogeneous data problem."

-- George Strawn, https://osf.io/ehk65/

Three eras of computing, according to G. Strawn

- Era one (1951-1995): many computers and many datasets
- Era two (1995-2025?): one computer and many datasets. SUN Computer's line was "the network is the computer."
- Era three (2025?-): one computer and one dataset. The network is the computer and all the data is interoperable

Source: "The Internet and FAIR Data" by G. Strawn https://osf.io/ehk65/

George Strawn and Futures . . .

The Internet changed how science and society function; FAIR Data may bring on a science revolution of the same magnitude as the science revolution of the 17th century (by enabling reuse of all science outputs—not just articles)

George Strawn, *MPS FAIR Hackathon Address*

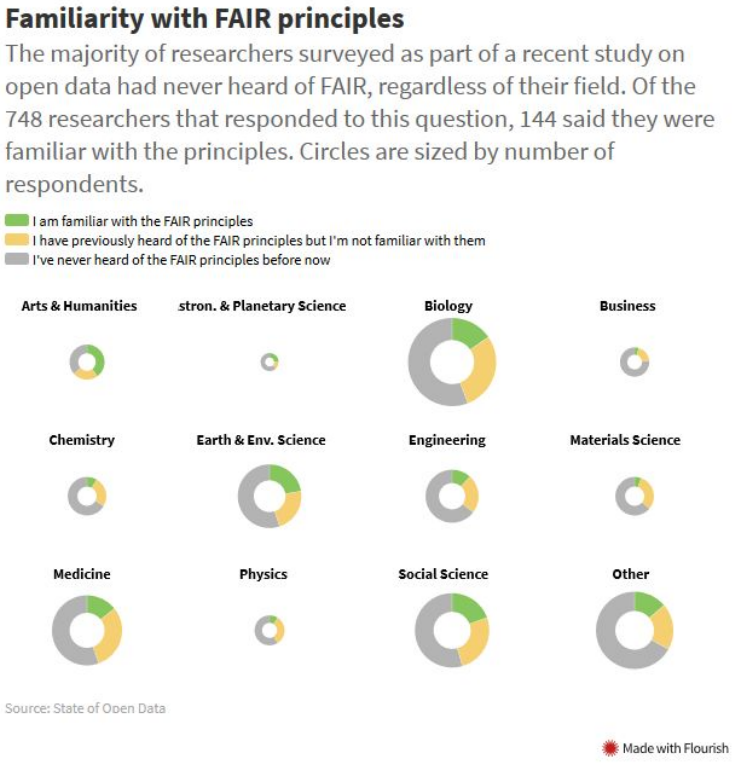
. . . moments of convergence require political, organizational, administrational, economic and technological efforts to turn technology into an infrastructure accessible for all

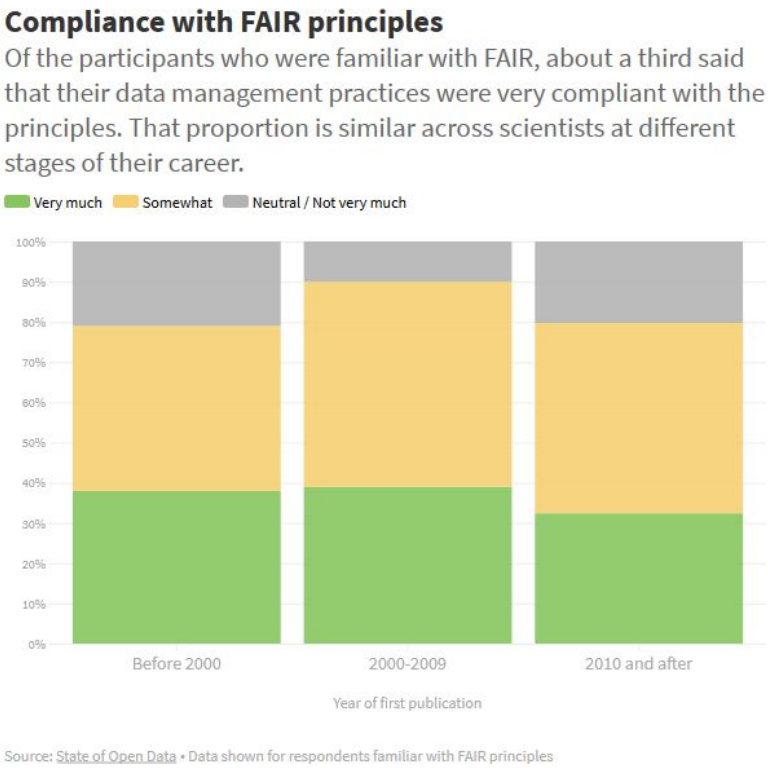
Strawn & Wittenburg in *Common Patterns in Revolution*

Slide credit: Nathalie Meyers
https://osf.io/y8rs5/

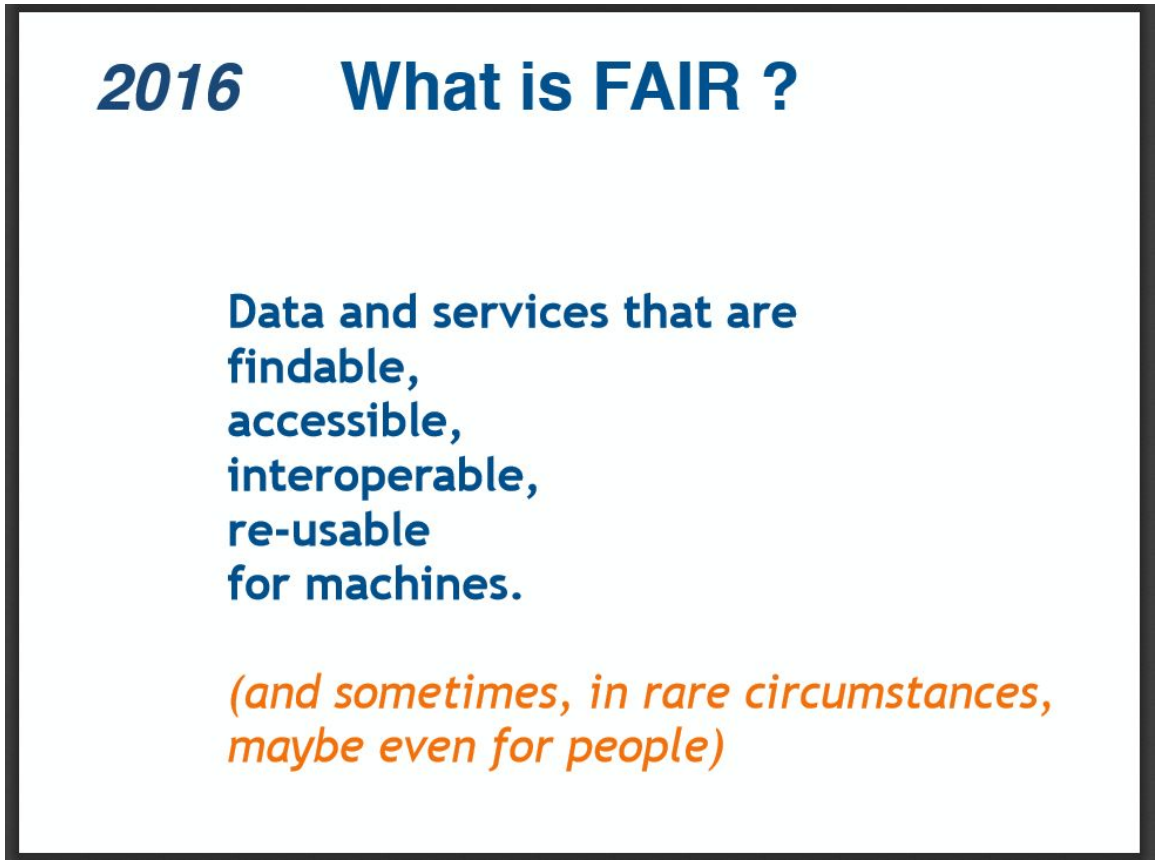
Step back to the present

Majority of researchers do not know about FAIR





https://www.natureindex.com/news-blog/what-scientists-need-to-know-about-fair-data



Erik Schultes (GO FAIR) presentation: https://osf.io/kr7qp/

2016 FAIR Guiding Principles

Sci. Data 3:160018 doi: 10.1038/sdata.2016.18 (2016)

Findable:

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

F2 data are described with rich metadata;

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

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

Accessible:

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

A1.1 the protocol is open, free, 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;

Interoperable:

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;

Reusable:

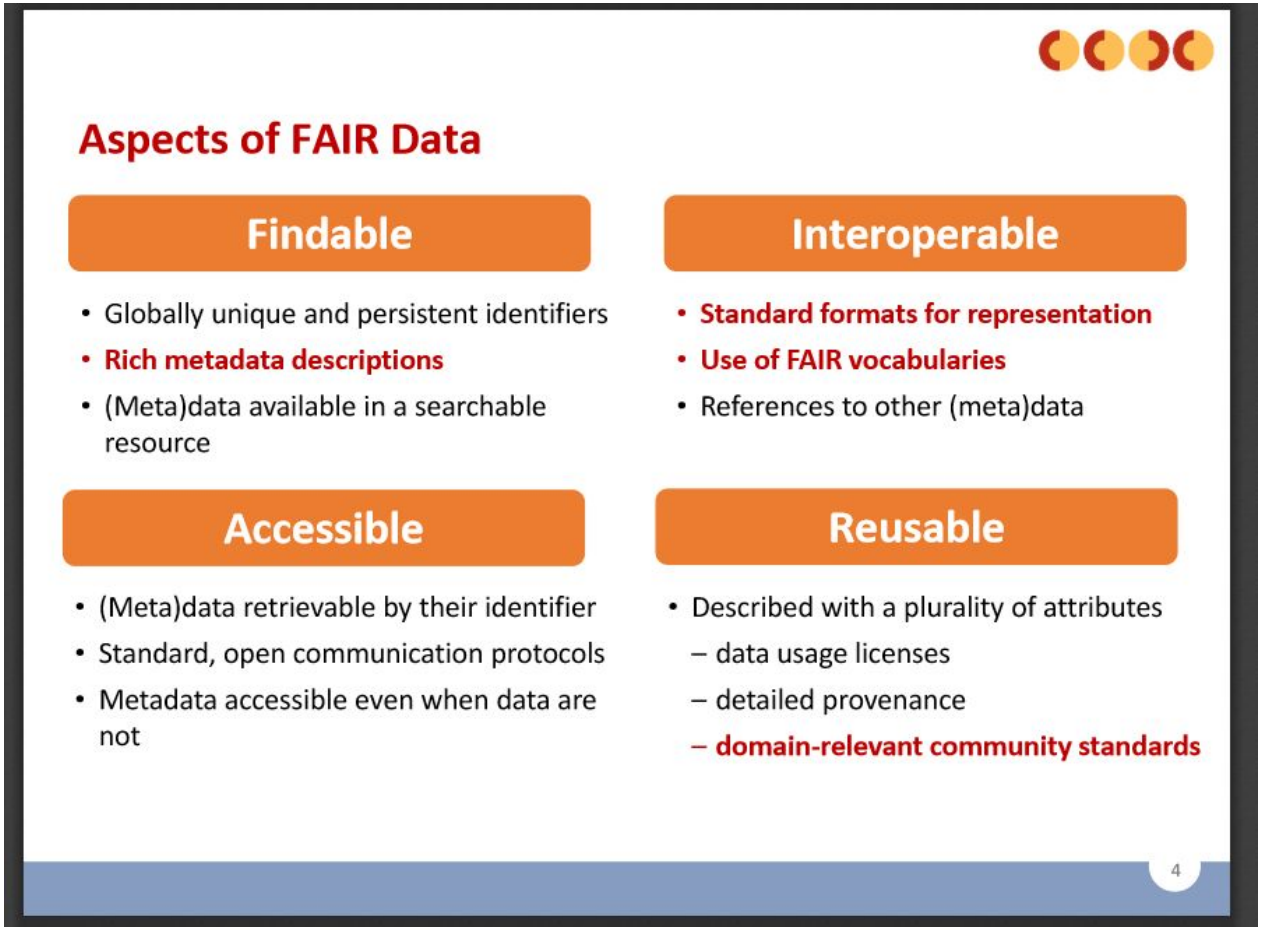
R1 meta(data) are richly described with a plurality of accurate and relevant attributes;

R1.1 (meta)data are released with a clear and accessible data usage license;

R1.2 (meta)data are associated with detailed provenance;

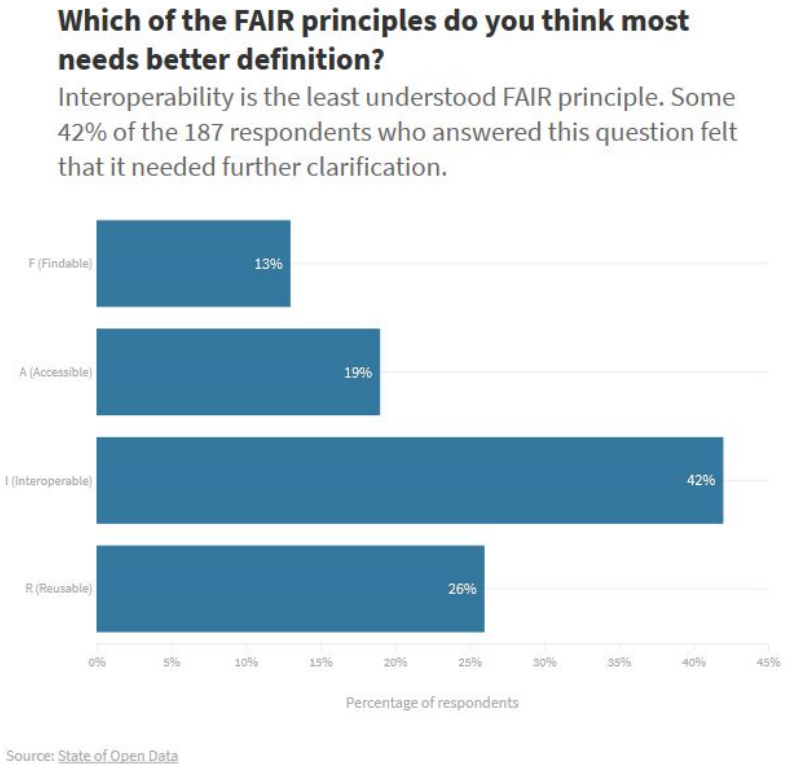
R1.3 (meta)data meet domain-relevant community standards;

Erik Schultes' (GO FAIR) presentation: https://osf.io/kr7qp/; FAIR Principles paper: https://doi.org/10.1038/sdata.2016.18



"Aspects of FAIR Crystallographic Data" by Ian Bruno, **Cambridge Crystallographic Data Centre**: https://osf.io/s7nm9/

Interoperability: the least understood FAIR principle.



https://www.natureindex.com/news-blog/what-scientists-need-to-know-about-fair-data

Interoperability is at the core of (the grand vision of) FAIR

"Data that should be readable for machines without the need for specialised or ad hoc algorithms, translators, or mappings."

"To ensure automatic findability and interoperability of datasets, it is critical to use:

1. Commonly used controlled vocabularies, ontologies, thesauri (having resolvable globally unique and persistent identifiers)

2. A good data model (a well-defined framework to describe and structure (meta)data)." This is process by which a terminology and a framework for the terminology comes together.

https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-broadly-applicable-language-knowledge-representation/

PubChem Case Study

"Making Data Interoperable: PubChem Demo/Use Case" by Evan Bolton: https://osf.io/6mxrk/

PubChem is an archive of chemical substance information and their biological activities.
https://pubchem.ncbi.nlm.nih.gov/

For computers to understand chemistry
they need our help, but our workflows are for humans

Read papers → Do science → Publish papers → Search papers → Read papers

Computers help humans at every step. The better they understand what we do and how we do it, the better they can help us do just about everything.

NIH U.S. National Library of Medicine

Image credits:
http://www.how-to-draw-funny-cartoons.com/cartoon-scientist.html
http://computertutorinc.net/computer-maintenance-safety-tips/

NCBI

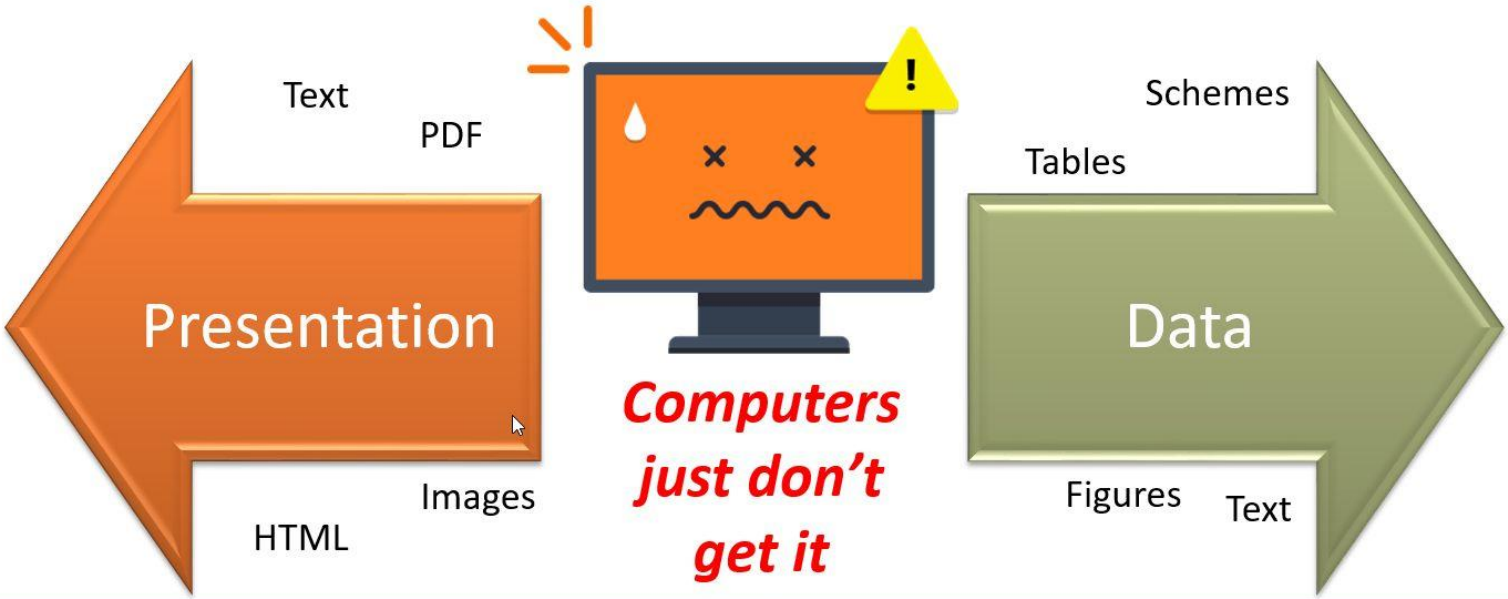
"Making Data Interoperable: PubChem Demo/Use Case" by Evan Bolton: https://osf.io/6mxrk/

Scientific Information

NIH U.S. National Library of Medicine NCBI

"Making Data Interoperable: PubChem Demo/Use Case" by Evan Bolton: https://osf.io/6mxrk/

Scientific Information



NIH U.S. National Library of Medicine

Image credit:
https://static.spiceworks.com/attachments/cms/0000/2161/sad-computer.png

NCBI

"Making Data Interoperable: PubChem Demo/Use Case" by Evan Bolton: https://osf.io/6mxrk/

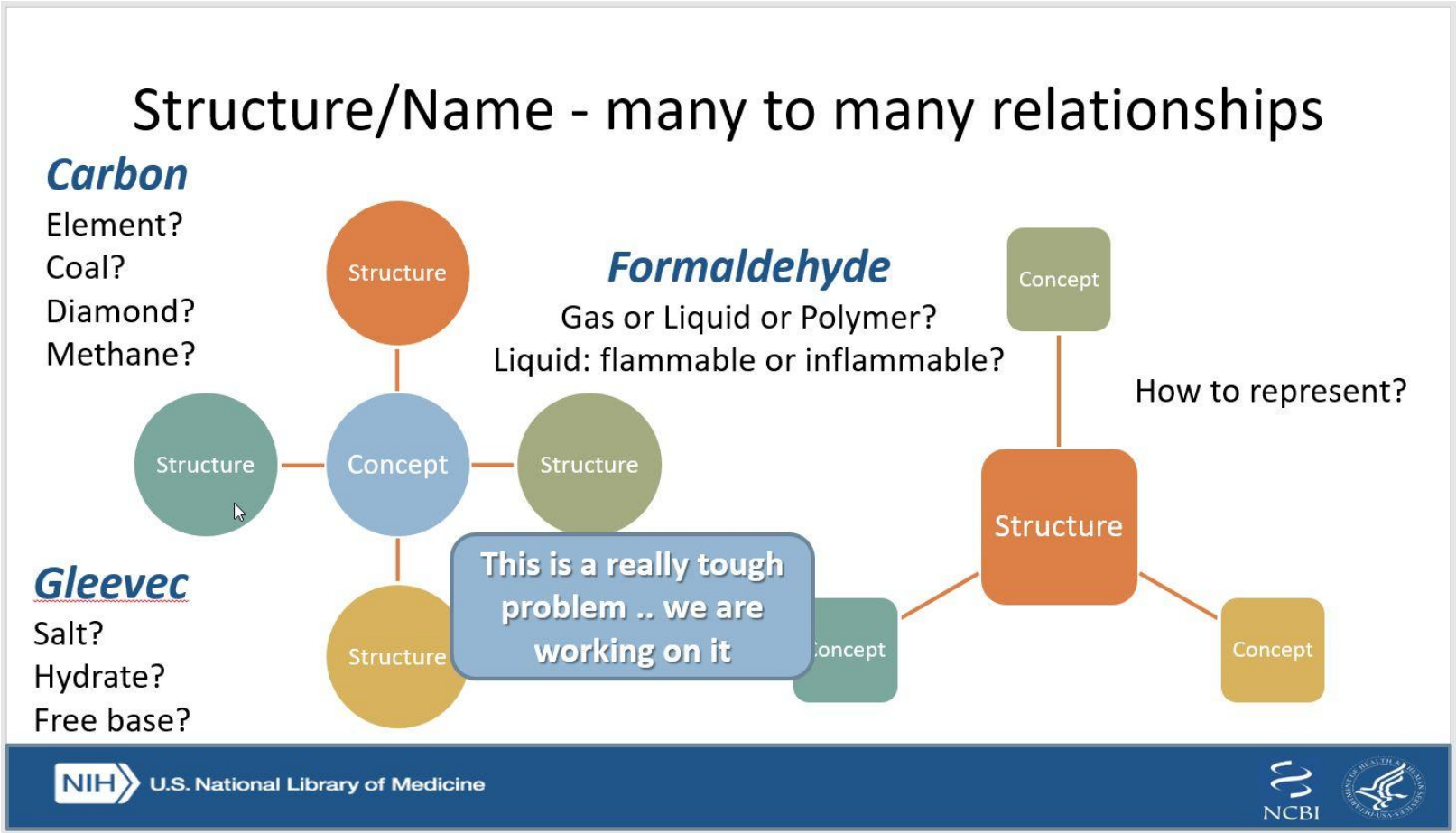
The chemical information understanding divide

NIH U.S. National Library of Medicine

Image credit: https://hcldr.files.wordpress.com/2016/12/1-rr-0116-industry-divided.jpg

NCBI

"Making Data Interoperable: PubChem Demo/Use Case" by Evan Bolton: https://osf.io/6mxrk/

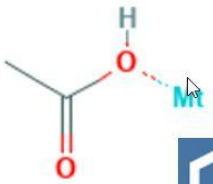


"Making Data Interoperable: PubChem Demo/Use Case" by Evan Bolton: https://osf.io/6mxrk/

Even elements can be troublesome

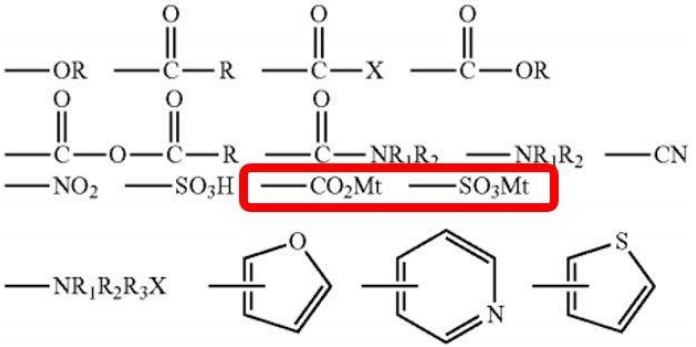
- Chemical diagrams often use abbreviations that can be mistaken for something else

Mt = Meitnerium



CID 60016057

Mt indicates Li, Na or K metal



Patent ID	Patent Title
US7871572	Chemical sensors based on metal nanoparticle encapsulated by mixed ligand and sensor array

NIH U.S. National Library of Medicine NCBI

"Making Data Interoperable: PubChem Demo/Use Case" by Evan Bolton: https://osf.io/6mxrk/

Chasing windmills?

Annotating and FAIR-ifying scientific content can be difficult to navigate:

- Identifiers, licensing/IP, standards, terminologies, normalization, best practices, machine accessibility, scientist education…
- What you can do today may be different from tomorrow
- Everything a work in-progress

Image credit: https://networkingnerd.files.wordpress.com/2012/04/don-quixote-windmill.jpg

"Making Data Interoperable: PubChem Demo/Use Case" by Evan Bolton: https://osf.io/6mxrk/

Going back to George Strawn and Futures...

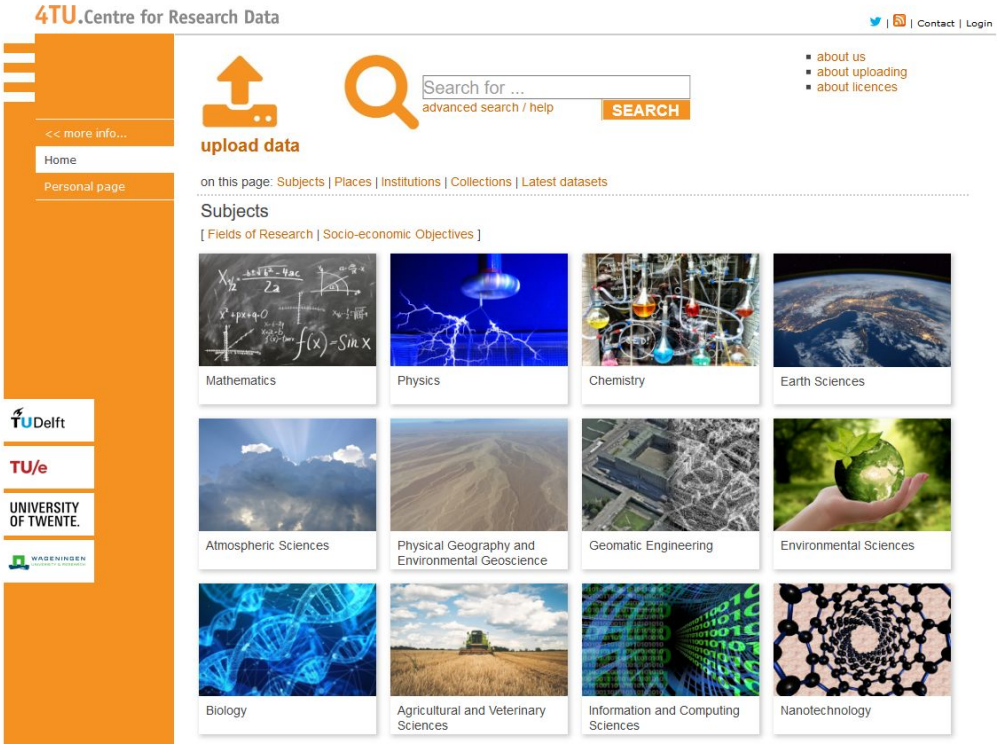
“Scientific revolution will come because of FAIR data, but likely not in a couple of years but in a couple decades.”

“The decade of the 20s will be a decade of development. In the 30s there will be enough consolidation and the technology will be a requirement by the society.”

Source: https://osf.io/ehk65/

In the meantime how can you contribute towards FAIR data?

Upload your data to a certified repository



https://data.4tu.nl/

Get a DOI (persistent identifier) for your data and metadata

Standard metadata descriptions attached to the data; references to other data/metadata

Data and metadata indexed and searchable

License

Access via standard & open protocols

What else could you do?

- Cite and refer to data and software as independent research objects (in the reference list with DOIs).

- If you review papers, request authors to obtain a DOI for their data and software (by depositing in a trusted repository) and ask them to cite data and software according to data and software citation standards.

- Identifiers are also for researchers. Help connecting research and researchers globally by obtaining an ORCID if you don't have one yet.

ORCID
Connecting Research and Researchers

And finally...

- ➢ Befriend your data steward, librarian, research software engineer and your fellow data champions. These are your allies in the road to FAIR.

Important element of culture change is developing a “Community of Practice.”

-- MPS Hackathon Group Notes https://osf.io/csbvj/

Some resources and acknowledgments

- MPS FAIR Hackathon, all materials and presentations available via https://osf.io/km8db/ (DOI 10.17605/OSF.IO/KM8DB)

- TOP 10 FAIR Data & SoftwareThings https://librarycarpentry.org/Top-10-FAIR/
  - TOP 10 FAIR Data & Software Things for Astronomy coming soon. Stay tuned!

With thanks to the University of Notre Dame for covering my travel and accommodation costs. Special thanks to Natalie Meyers, Marta Teperek and the TU Delft Data Stewards for making it possible for me to attend.