ICRI 2018
Vienna
13 September 2018

Research Infrastructures ensuring the trust and quality of data
Session 5A

Simon Hodson, Executive Director, CODATA
www.codata.org

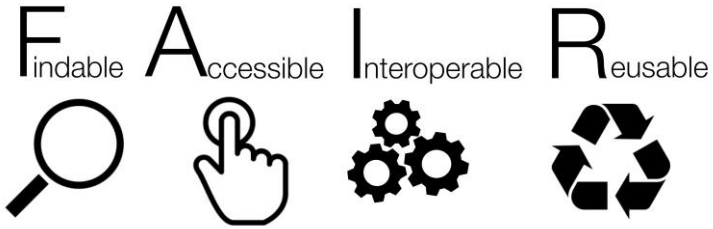
International Science Council

What contributes to quality and trust?

What principles and practices are relevant to RIs with regard to trustworthiness and quality of data?

Quality and trust relates to…

- The data as such
- Metadata and provenance
- Controlled vocabularies and ontologies
- Organisational standards, processes and sustainability



- FAIR principles help us a great deal in thinking about these issues and preparing policy
- FAIR needs to be supported by an ecosystem of policies, standards, services and robust organisations, including Research Infrastructures

The data as such…

- Scientific and (sometimes) operational questions
- Were the instruments calibrated properly, is the methodology robust, were the survey questions well designed…?
- The need to be able to assess data quality, to test reproducibility of a result is perhaps *the* most fundamental and imperative argument for Open Data and Open Science
- Need openness in hardware, in algorithms/code, in any software packages and tools, in steps throughout the chain of data capture, processing, reduction, preparation etc.

THE DEFINING CONSTANTS OF THE INTERNATIONAL SYSTEM OF UNITS

Defining constant	Symbol	Numerical value	Unit
hyperfine transition frequency of Cs	$\Delta\nu_{Cs}$	9 192 631 770	Hz
speed of light in vacuum	c	299 792 458	m s^{-1}
Planck constant*	h	$6.626\,070\,15 \times 10^{-34}$	J Hz^{-1}
elementary charge*	e	$1.602\,176\,634 \times 10^{-19}$	C
Boltzmann constant*	k	$1.380\,649 \times 10^{-23}$	J K^{-1}
Avogadro constant*	N_A	$6.022\,140\,76 \times 10^{23}$	mol^{-1}
luminous efficacy	K_{cd}	683	lm W^{-1}

*These numbers are from the CODATA 2017 special adjustment. They were calculated from data available before the 1st of July 2017.

Metadata and provenance

- Thorough, quality and trustworthy metadata is essential for data discovery and reuse
- Completeness and quality of metadata can be more easily assessed and encouraged
- Various organisations have suggested minimal metadata standards across domains and in particular domain reporting
- Provenance metadata has a considerable bearing on the assessibility and reusability of the data
- Without robust provenance metadata we cannot be fully confident in the quality and trustworthiness of the data
- **Need quality standards for metadata both in terms of standards established by particular disciplines and in terms of completeness**

FAIRsharing.org
standards, databases, policies

DDI®
DATA DOCUMENTATION INITIATIVE

Controlled Vocabularies and Ontologies

- Very important for interoperability and reuse; essential for computational use at scale
- Operate at the datum level, i.e. of a particular measurement. Or at the level of defining features in images, terms in text etc.
- Do we or do we not have a shared, consistent and robust definition of what has been measured, observed, recorded and how?
- Without the use of shared and well-defined vocabularies or ontologies, we cannot be fully confident in the quality and trustworthiness of the data
- The data may not be fully usable without these things so they are integral to the 'quality' of the data in that sense

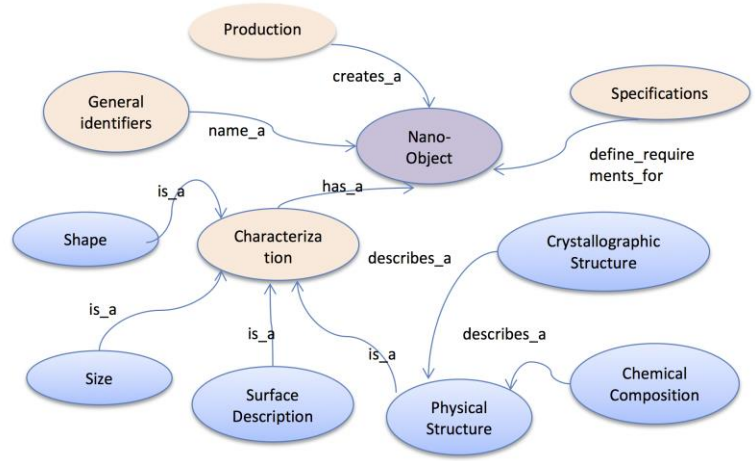


Figure 4. Information categories for describing an individual nano-object

Uniform Description System for Materials at the Nanoscale: http://dx.doi.org/10.5281/zenodo.56720

http://www.obofoundry.org/

Organisational Standards, Processes and Sustainability

- FAIR data depends on an ecosystem of organisations that provide identifiers, metadata standards, vocabularies and ontologies, that steward the data and provide useful tools, computational access to the data
- Important not to focus on ‘making data FAIR’ at the expense of maintaining about this ecosystem
- Data creation, processing, annotation, ingest, stewardship, quality metadata and semantics, providing human and computatational access: core roles for Ris and essential for quality and trustworthiness
- To do this, it is important to ensure the robustness of business processes for these things, significance of mission, value proposition and business model for sustainability etc.
- Quality and maturity standards exist for these things and should be explored, refined and adopted

Business models for sustainable data repositories
https://doi.org/10.1787/302b12bb-en

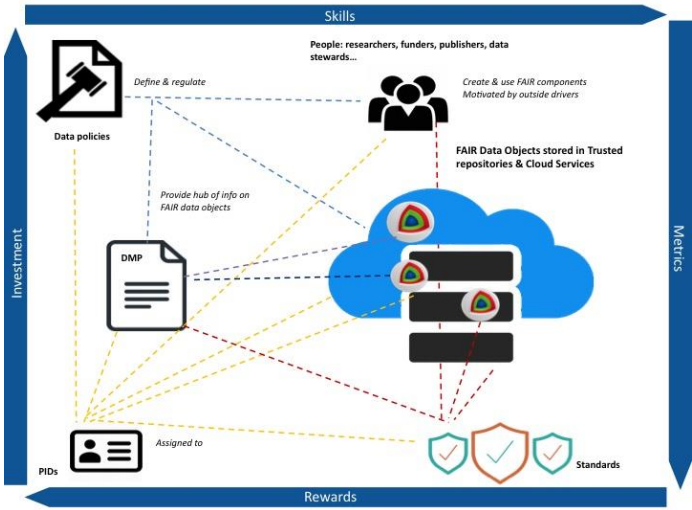
RIs and policy implications

CoreTrustSeal
https://www.coretrustseal.org/

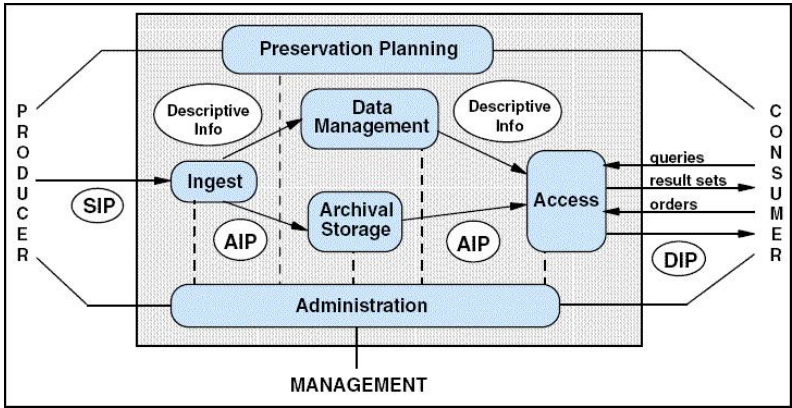
- **RIs have a role in relation to all of these dimensions**

Policy Implications: policies should

- Support FAIR data as an important contributor to some of these components of quality
- Take steps to ensure the development, support and sustainability of the **FAIR data ecosystem**, particularly identifiers, metadata standards and vocabularies / ontologies, trusted services
- RIs and other institutions that steward data and provide computation access should look seriously at standards for organisational processes such as CoreTrustSeal, ISO 14721 and 16363; contribute to their further development and adoption



Interim FAIR Data Report
https://doi.org/10.5281/zenodo.1285272



OAIS Reference Model https://public.ccsds.org/pubs/650x0m2.pdf

Thank You for your attention

- European Commission Expert Group Interim Report on **Making FAIR Data a Reality** https://doi.org/10.5281/zenodo.1285272
- ISC and CODATA Initiative on Data Interoperability and Integration for Interdisciplinary Research Programmes http://dataintegration.codata.org/
- African Open Science Platform Strategy Document with some insights on coordinating Research Infrastructures for Open Science http://www.codata.org/strategic-initiatives/african-open-science and https://doi.org/10.5281/zenodo.1407488
- **International Data Week, comprising SciDataCon and the RDA 12th Plenary Meeting, 5-8 November, Gaborone, Botswana: http://internationaldataweek.org/**

- simon@codata.org | www.codata.org | @simonhodson99

International Science Council