Applying the FAIR principles to the worlds of research and measurement

Simon Hodson
Executive Director
CODATA
www.codata.org

BIPM CGPM 2022
Versailles Convention Centre
17 November 2022

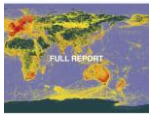
CODATA's mission and operation

- **The mission of CODATA is to "Connect data and people to advance science and improve our world".**
- As the 'Committee on Data of the International Science Council (ISC)', CODATA supports the ISC's mission of 'advancing science as a global public good' by promoting Open Science and FAIR data.
- CODATA convenes a global expert community and provides a forum for international consensus building and agreements around a range of data science and data policy issues, from the fundamental physical constants to cross-domain data specifications.
- **CODATA's membership includes national data committees, scientific academies, International Scientific Unions and other organisations.**

Making Data Work…

- **Decadal Programme: Making Data Work for Cross Domain Grand Challenges**
- Recommendations on core interoperability and FAIR
- FAIR Vocabularies with ISUs
- Cross-Domain Case Studies
- Global Open Science Cloud initiative
- Regional Open Science Platforms

Data Policies



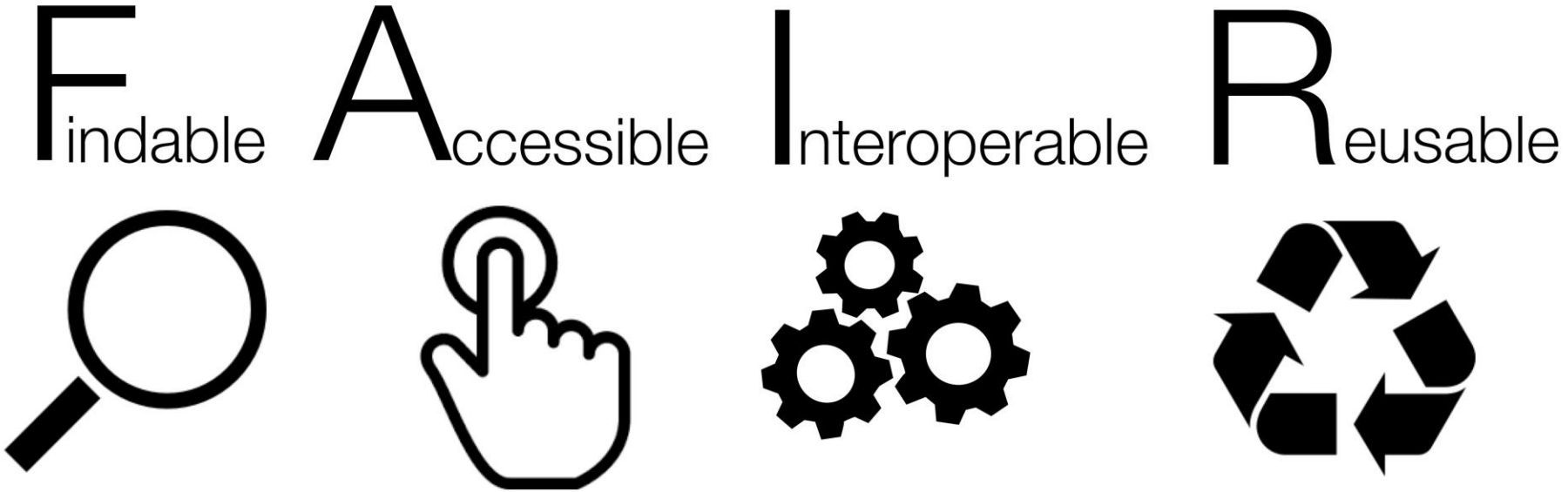
- International Data Policy Committee http://bit.ly/data-policy-committee;
- One major policy report per year.
- 20-Year Review of GBIF published in May 2020
- Preparing Independent Review of CAS Earth data policy and practices

Data Science

- Data Science Journal: https://datascience.codata.org/
- International Data Week and CODATA Conference series.
- Task Groups and Working Groups.

Data Skills

- CODATA-RDA School of Research Data Science.
- Beijing and other training workshops.
- #terms4FAIRskills and FAIRsFAIR



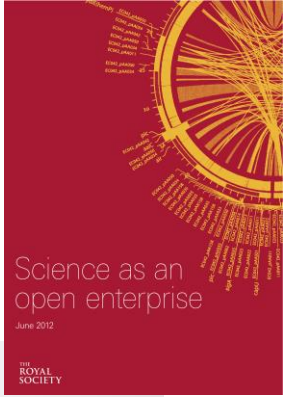
(Wilkinson, M., et al., The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, http://dx.doi.org/10.1038/sdata.2016.18)

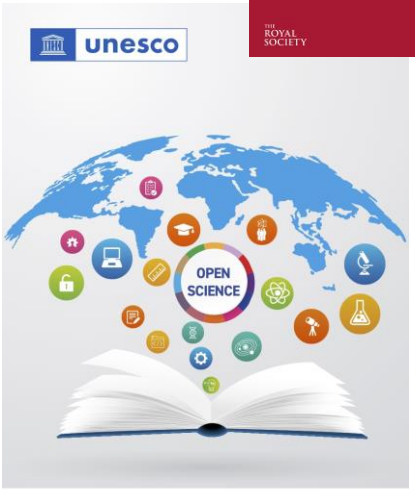
FAIR and Open

- **FAIR is an important component of Open Science, but… FAIR does not necessarily mean Open**
- **Drivers for Open:** reproducibility and transparency, public benefit.
    - Research data should be as Open as possible and only as closed as necessary; Open by default.
- **Drivers for FAIR:** maximize the utility, usability of data.
    - Principles of good data stewardship: not enough to put make data Open, dump it raw onto the Web
- **FAIR applies as much to data that MUST be restricted as to data that can be Open**

The case for Open Science

- **Good scientific practice depends on communicating the evidence.**
  - Open research data are essential for transparency, scrutiny, reproducibility, self-correction.
  - **Boulton, Science as an Open Enterprise:** 'to fail to communicate the data that supports scientific assertions is tantamount to malpractice'.
  - Editorial, Miyakawa, T. **No raw data, no science:** another possible source of the reproducibility crisis. Molecular Brain, 13, 24 (2020). https://doi.org/10.1186/s13041-020-0552-2
  - Increasingly strong position from funders, journal editors, publishers: open data is essential to the conduct of our science.
- **Societal and scientific benefit: essential for addressing global and societal challenges.**
  - **UNESCO Recommendation:** strong emphasis on global benefit; benefit for society and involvement of societal actors; very strong ethical dimension.
  - **OECD Principles:** research data produced by publicly funded research are a public asset.
  - **European Commission:** essential to accelerate scientific discovery, economic benefits of Open data.
  - Open data practices have transformed certain areas of research: genomics and related biomedical sciences; crystallography; astronomy; areas of earth systems science; various disciplines using remote sensing data…





UNESCO Recommendation on Open Science

The case for FAIR (data, metadata, code)

- **FAIR: encompasses in an easy communicable acronym, some core principles of good data stewardship**
  - Increases the usability and utility of data, metadata, code.
- **Emphasis of the benefits of machine-actionability**
  - FAIR principles designed to support the use of data at scale, by machines, harnessing technological potential.
  - Vision of harnessing the technologies of the web, to improve querying of vast, dispersed and heterogenous data.
- **Increases the value of data for science and the economy**
  - PWC report, 2019: **Opportunity cost to the European science system of NOT having FAIR data: 8.2 Bn Euros.**
  - (at least) **80% of project effort goes into downstream 'data wrangling', rather than upstream 'data stewardship'.**

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

FAIR Guiding Principles

To be Findable:

- 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 it describes
- F4. **(meta)data** are registered or indexed in a searchable resource

To be 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

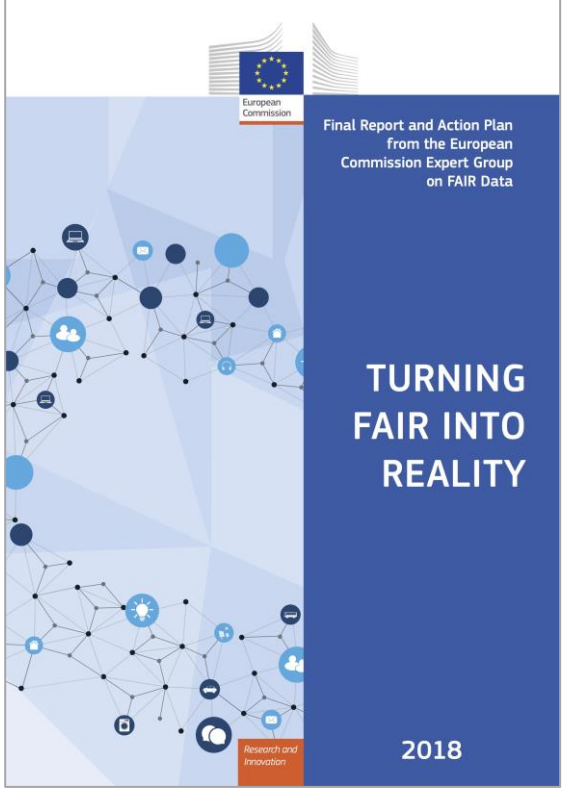
To be Interoperable:

- 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

To be 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

(Wilkinson, M., et al., The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, http://dx.doi.org/10.1038/sdata.2016.18)



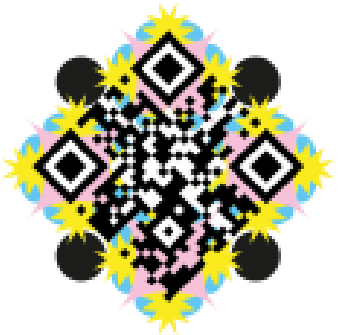
European Commission Expert Group, Chaired by Simon Hodson, Turning FAIR into Reality (2018)
https://doi.org/10.2777/1524

- **Findable:** have sufficiently rich metadata and a unique and persistent identifier, to enable discovery.
- **Accessible:** retrievable by humans and machines through a standard protocol; authentication and authorization where necessary.
  - **Allows programmatic access for analysis.**
- **Interoperable:** metadata use a 'formal, accessible, shared, and broadly applicable language for knowledge representation'.
  - **The descriptions of variables etc follow a shared specification and are commensurable.**
- **Reusable:** metadata provide rich and accurate information; clear usage license; detailed provenance.
  - **Both humans and their analytical tools know what can be done with the data (license) and can assess its provenance.**

FAIR Digital Objects

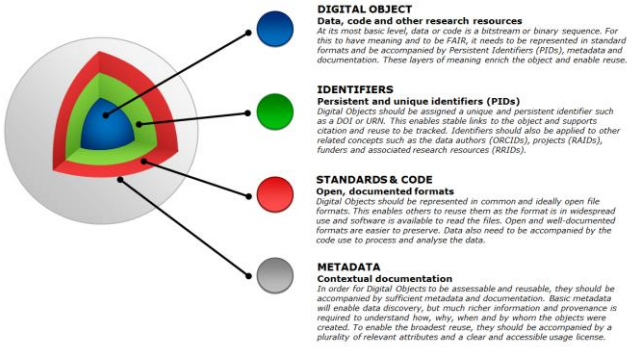
- **FAIR Digital or Data Object = 'FAIR, machine-interpretable and self-explanatory units of information'.**
- What is the precise **identity and location** of this object? (PID / GUPRI)
- What **may** I do with it? (License, protections)
- What **can** I do with it? (Data type)
- …
- **How can I process it?**
  - Concepts, variables and units (Metadata and semantics)
  - Data structure
  - Provenance and processing
  - Quality, accuracy…

FDOs bind all critical data about a digital entity in one object. The data is FAIR: findable, acessible, interoperable and reusable (both by humans and machines).

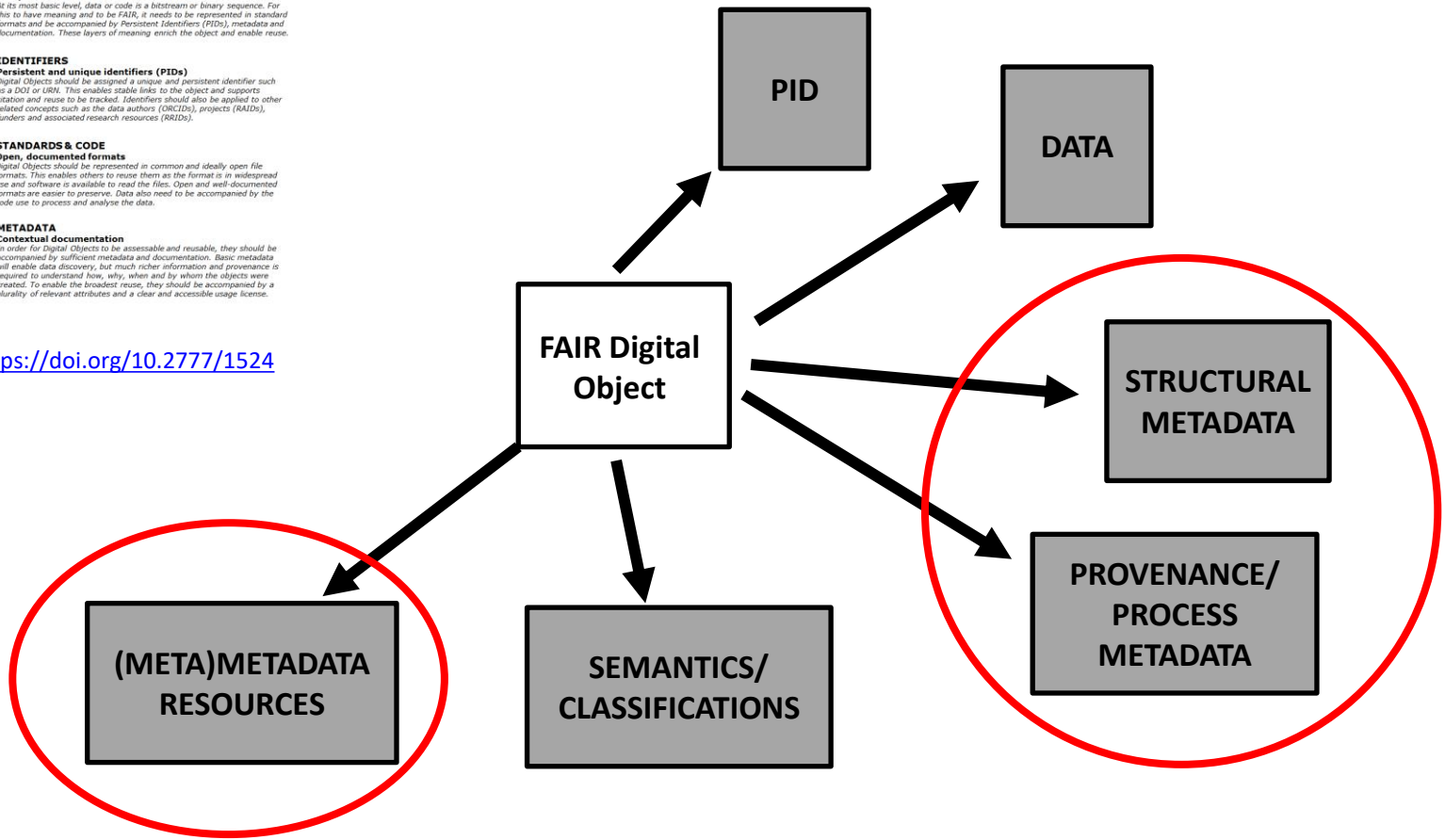


European City of Science Leiden2022

FAIR Digital Objects: a way of thinking about the information needed for reusability / machine actionability



Turning FAIR into Reality https://doi.org/10.2777/1524



What does this mean for measurements and research?

- **Joint Statement of Intent on the digital transformation in the international scientific and quality infrastructure: https://www.bipm.org/en/liaison/digital-transformation**
- CODATA is a party to the Joint Statement of Intent, along with BIPM, ISC, and other standards and metrology organisations.
- States that "maintaining … confidence in the accuracy and global comparability of measurement will require the creation and adoption of **a fully digital representation of the SI**, **including robust, unambiguous, and machine-actionable digital representations of units of measurement and of measurement results and uncertainties**"

Bureau International des Poids et Mesures

Signing of the Joint Statement of Intent. Clockwise from top left: Wynand Louw, CIPM President (on behalf of the BIPM); Mathieu Denis, ISC Science Director and Acting CEO; Barend Mons, CODATA President; Roman Schwartz, CIML President (on behalf of the OIML); Frank Härtig, IMEKO President.

What does this mean for measurements and research?

- **Need for digital (FAIR) representation of units:** Digital SI; cooperation and coordination, interoperability / translations with other systems for representing units.
- **Need for digital (FAIR) representation of quantities:** online referenceable definitions of the measurand / property / variable…
- Research or metrology 'domains' increasingly needing to publish definitions of quantities as 'FAIR semantic artifacts' (ontologies, terminologies, vocabularies.
  - Cox et al., '10 Simple Rules for Making a Vocabulary FAIR' https://doi.org/10.1371/journal.pcbi.1009041
- Some domains have the idea of a 'variable cascade' to assist with machine-processing / data integration:
  - Conceptual variable: the definition of the variable / quantity.
  - Represented variable: definition, plus meaning and code.
  - Instance variable: instanciation in a given dataset.

Digital SI

https://www.bipm.org/en/conference-centre/bipm-workshops/digital-si/

DRUM (Digital Representation of Units of Measurement) TG

- **Mission: Promote cooperation and coordination across initiatives**, and in particular mobilising the input of the various scientific domains, as represented by the ISUs/ISAs: https://codata.org/initiatives/task-groups/drum/
- **Manifesto**, endorsed by the ISC and the ISUs/ISAs, calling for greater action and investment on the issue of units of measure (their definition, digital representation and conversion): https://doi.org/10.5281/zenodo.4081656
- **Call to Action 'Stop squandering data: make units of measurement machine-readable' Nature Comment Article https://doi.org/10.1038/d41586-022-01233-w**
- **Sessions at IDW and a Units Summit: https://bit.ly/DRUM-BIPM-Units-Summit**

Bureau International des Poids et Mesures

COMMITTEE ON DATA CODATA INTERNATIONAL SCIENCE COUNCIL

- Contributing to the CIPM WG work on a **'Universal Metrology Data Model'** for Units of Measurement.
- Preparing publication of the **CODATA Recommended Values of the Fundamental Constants** as machine-readable Linked Open Data.
- **Discussion of expanding the remit of the Task Group to cover definition of quantities**

Making Data Work for Cross-Domain Grand Challenges

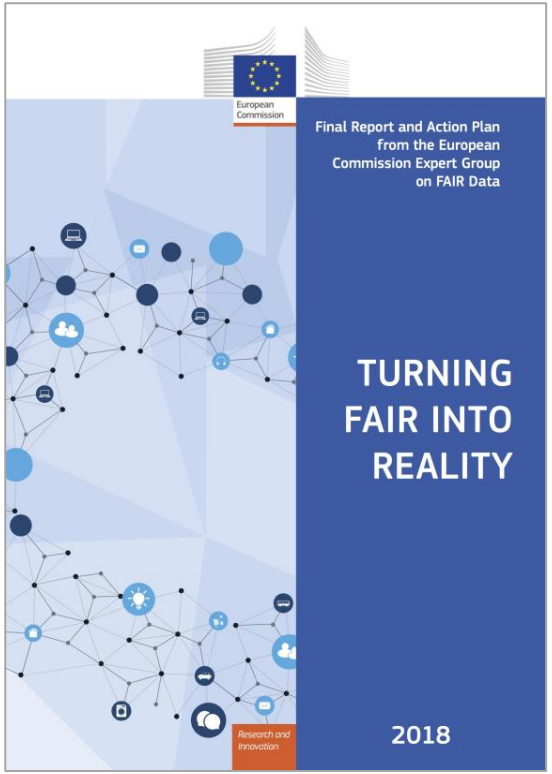
- **ISC Action Plan entrusts CODATA with an initiative 'Making Data Work for Cross-Domain Grand Challenges': establish a global (decadal) programme to address these issues.**
- The major, pressing global scientific and human issues of the 21st century can **ONLY** be addressed through **research that works across disciplines to understand complex systems**, and which uses **interdisciplinary and transdisciplinary** approaches to turn data into knowledge and then into action.
- **Preparatory Phase:** exploring technical issues and case studies through Dagstuhl workshops, TGs and WGs, funded projects etc, to understand the challenges and prepare the programme.
- **Core Interoperability Framework:** units, vocabularies, data structure, data description…
- **Case Studies:** in a range of domain and cross-domain research areas.
- **WorldFAIR Project:** takes this approach forward in the context of an EU-funded project.
- **Approach extended to Global Open Science Cloud (GOSC) Initiative.**

WorldFAIR: Global cooperation on FAIR data policy and practice

- Two year project to advance implementation of the FAIR principles in a range of disciplines, or cross-disciplinary research areas.
- Funded by the European Union, HORIZON-WIDERA-2021-ERA-0 — Project: 101058393
- Two years from 1 June 2022.
- Nineteen partners.
- Institutions in France, Belgium, Cyprus, Denmark, Germany, Ireland, Norway; Australia, Brazil, Kenya, New Zealand, USA; UK.
- Important partnership between CODATA and Research Data Alliance.

WorldFAIR

- Among the most important, but most challenging, recommendations of the **Turning FAIR into Reality** report, is R.4:
- ‘Develop **interoperability frameworks** for FAIR sharing within disciplines and for interdisciplinary research: Research communities need to be supported to develop interoperability frameworks that define their practices for data sharing, data formats, metadata standards, tools and infrastructure. To support interdisciplinary research, these interoperability frameworks should be articulated in common ways and adopt global standards where relevant.’
- UNESCO Recommendation on Open Science and ISC Action Plan make similar recommendations.
- **What is the current practice in metrology? Is there activity to define Open Science and FAIR practices, and an interoperability framework?**



WorldFAIR Case Studies

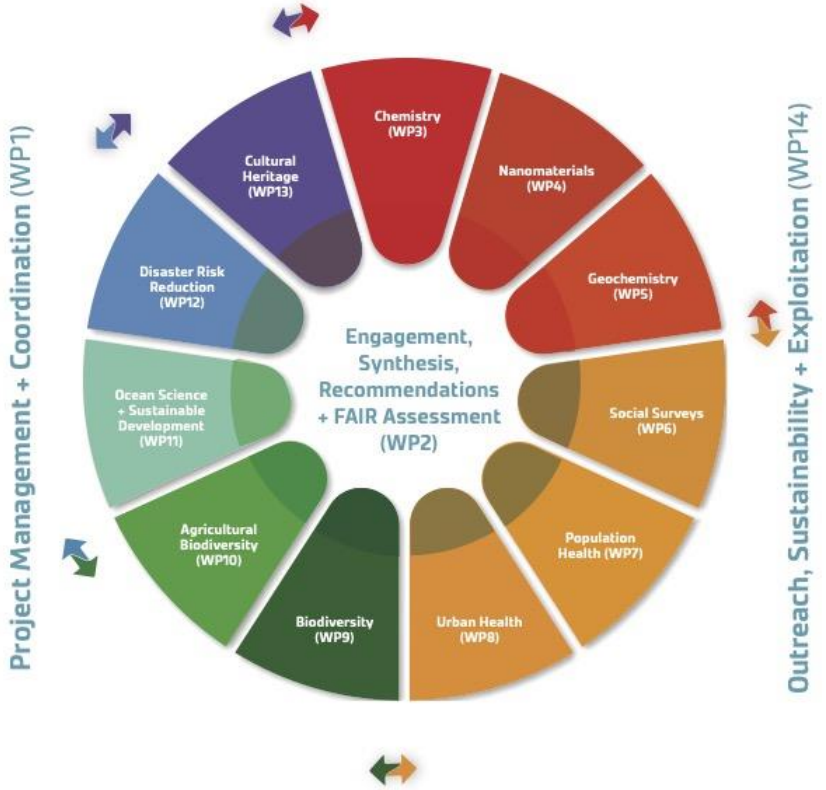
- **Chemistry** - making IUPAC assets FAIR
- **Nanomaterials** - applying NanoInchi and FAIR recommendations in Nanosafety.
- **Geochemistry** - recommendations for FAIR in geochemistry, particularly vocabularies.
- **Social Surveys Data** – data harmonisation between ESS and AussiESS.
- **Population Health** - INSPIRE - Integration of population surveys with clinical and genomics data for COVID-19 research in eastern and southern Africa.
- **Urban Health** - terminologies and making urban health data FAIR
- **Biodiversity** – improving GBIF data model in collaboration with TDWG - GBIF (Global Biodiversity Information Facility)
- **Agricultural Biodiversity** - pollinator data (KALRO, Embrapa, Meise, HiveTracks)
- **Ocean Science** - Implementing FAIR in the ODIS (Ocean Data and Information System) for the UNESCO Oceans' decade.
- **Disaster Risk Reduction** - recommendations on making DRR data and terminologies FAIR, case studies in Africa and Pacific Islands
- **Cultural Heritage** - recommendations on making cultural heritage data FAIR (particularly digital representation of heritage artefacts)

WorldFAIR Coordination and Synthesis

- CODATA leading coordination and synthesis WP.
- **FAIR Implementation Profiles** for each case study (a methodology for understanding FAIR practices).
- Findings will lead into recommendations for **domain sensitive FAIR assessment**.
- Recommendations and documentation for a **Cross-Domain Interoperability Framework**.

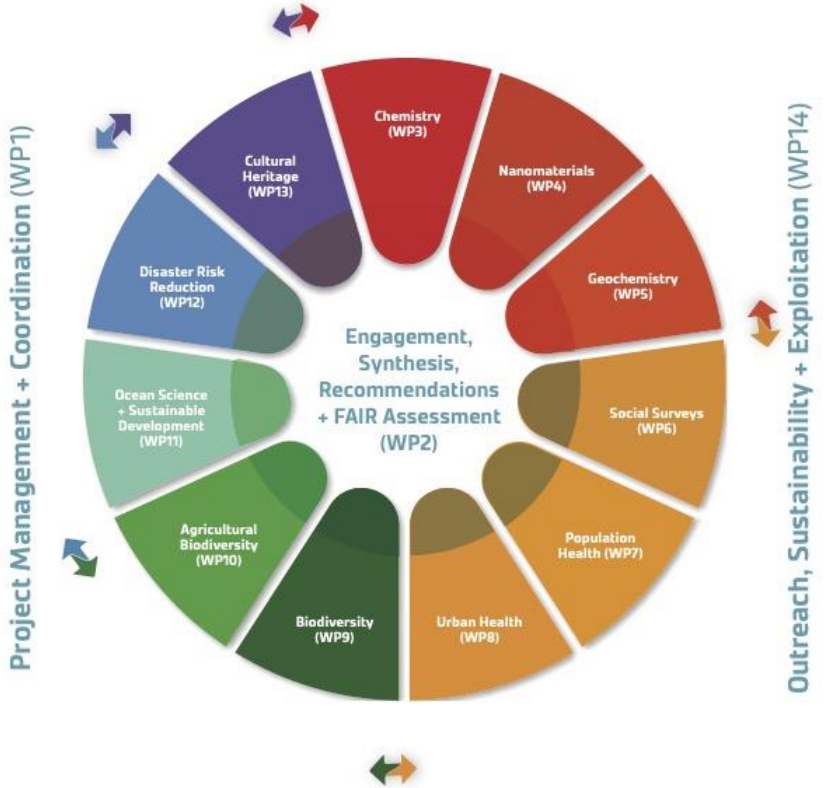
Cross-Domain Interoperability Framework

- FIPs activity prepares WorldFAIR for two 'WP2' outputs: **Cross-Domain Interoperability Framework** and **recommendations for domain sensitive FAIR assessment.**
- Exploring features of a **Cross-Domain Interoperability Framework** with case studies from a range of research areas.
    - Units and quantities/measurands/properties; terminologies/vocabularies/ontologies;
    - Data description/variable cascade, data structure;
    - Provenance and processing; data types; licences; protection and programmatic access…
- A number of webinars / conference sessions: GOSC DataIO Webinar; SciDataCon Session; DCMI Conference.
- Dagstuhl Workshop: discussion paper, report and links at https://worldfair-project.eu/



Where next? CODATA, WorldFAIR and FAIR

- WorldFAIR is a two-year kick start for a set of domain and cross-domain research areas.
- Work of DRUM on digital representation of units of measure, the potential work on FAIR / machine-readable definitions of quantities is very relevant.
- **Hope with the Joint Statement of Intent, DRUM, FAIR Vocabularies, FDOs and WorldFAIR to make a significant contribution to FAIR and to the digital transformation in the international scientific and quality infrastructure**



COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL

International
Science Council

Thank you for your attention

Simon Hodson, CODATA
www.codata.org
simon@codata.org
@simonhodson99 ; @CODATAnews