COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL

International Science Council

FAIR and Open Science to Enable Cross-Domain Research

Simon Hodson
Executive Director, CODATA

https://worldfair-project.eu/
https://codata.org/

Polish Open Science Conference 2024
Data in the Service of Science and Society
Krakow, Poland
10 April 2024

CODATA's vision and strategy

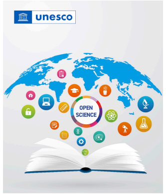
- **CODATA's vision is of a world in which science is empowered to address universal challenges through the transparent, trustworthy and equitable use of data and information.**
- CODATA is a key resource through which ISC, and the ISC membership, can address data-related issues.
- CODATA has a **skilled secretariat** and leverages a **global expert community** to provide 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 serves a **membership** that includes national data committees, scientific academies, International Scientific Unions and other organisations.
    - **We would love to have Poland as a member!**
- CODATA works with many **partners**, including GO FAIR, RDA and WDS; many international and national data organisations; UN and intergovernmental agencies…

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

(Some) CODATA Contributions

- **Mercè Crosas**, CODATA President, **2023+**, WorldFAIR+; data policy contributions with ISC, OECD, UNESCO; rigorous, responsible and relevant; pragmatic and collaborative.
- **Barend Mons**, CODATA President, **2018-2023**, one of the originators of the FAIR principles, GO FAIR, FAIR Implementation Profiles, WorldFAIR.
- Simon Hodson, Executive Director, **2013+**, chair of Expert Advisory Group for Turning FAIR into Reality **2018**; vice chair of Expert Advisory Group for UNESCO Recommendation on Open Science **2021**.
- **Geoffrey Boulton**, CODATA President, **2014-2018**, lead author of 'Science as an Open Enterprise', **2012**; Open Data in a Big Data World, **2016**; African Open Science Platform Strategy,
- **2000s:** CODATA made a major contribution to the OECD Principles on Public Access to Research Data (2007) and the GEO Data Sharing Principles (2005 and 2015).
- **2000:** The proposal for using DOIs for data originated in a CODATA Germany session at the CODATA 2000 Conference.
- **1969+:** Fundamental Physical Constants, recommended values used in physics, cosmology, metrology etc.; since **2019** underpins the core units of the SI System (i.e. all the core units are derived mathematically from the CODATA recommended values of the fundamental physical constants.

WorldFAIR



UNESCO Recommendation on Open Science

Making Data Work…

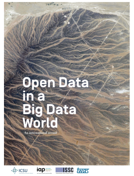
WorldFAIR

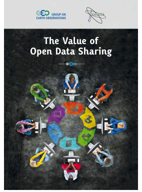
GOSC Global Open Science Cloud

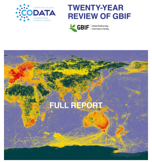
DDI DATA DOCUMENTATION INITIATIVE

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

Improving Data Policies







- International Data Policy Committee [http://bit.ly/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

Advancing Data Science

ubiquity press open scholarship

23-26 OCT 2023 SALZBURG International Data Week A FESTIVAL OF DATA

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

Enhancing Data Skills

CODATA – RDA Data Schools

FAIRSFAIR

ICTP The Abdus Salam International Centre for Theoretical Physics

GO FAIR

- CODATA-RDA School of Research Data Science.
- Beijing and other training workshops.
- CODATA RDM Terminology

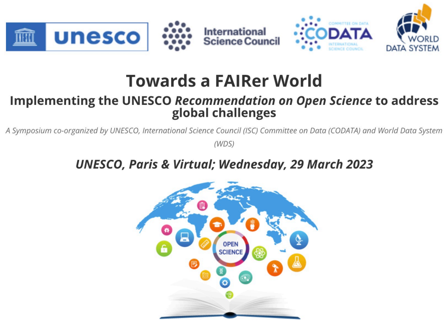
COMMITTEE ON DATA CODATA INTERNATIONAL SCIENCE COUNCIL

International Science Council

Improving Data Policy

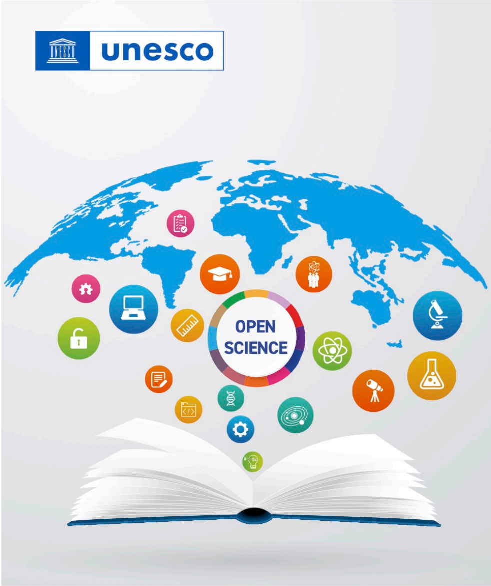
- Impact: The impact of this work will be to **support the effectiveness and ethical dimensions of global science by improving Open and FAIR data policies**, and by ensuring that data policy is an integral part of science policy.
- Flagship initiative: **International Data Policy Committee (IDPC)**
  - New IDPC Action Plan: https://bit.ly/IDPC-Action_Plan
    - **Data policy for…** 1) data quality, reliability, and integrity; 2) science in crisis situations; 3) education (data policy curriculum); 4) AI; 5) Open Science; 6) publication and communication of science; 7) inclusion and diversity.
  - Working on a Handbook for Data Policy and AI
- **CODATA-UNESCO WG on Data Policy for Emergencies: will present Data Policy for Emergencies, contribution to the UNESCO Open Science Toolkit at UN Data Forum in Nov 2024**
  - Towards a FAIRer World Workshop, March 2023: https://codata.org/events/science-and-policy-workshops/towards-a-fairer-world/
  - The ‘PROTECT’ Essential Elements in Managing Crisis Data Policies https://doi.org/10.5334/dsj-2024-012
- **CODATA Data Ethics Task Group**
  - CODATA Data Ethics Working Group Policy Briefs available, feedback requested
  - Data Ethics and Research Integrity https://doi.org/10.5281/zenodo.10933361
  - Data Ethics and Privacy https://doi.org/10.5281/zenodo.10933546
  - Data Ethics and Structural Inequities in Science https://doi.org/10.5281/zenodo.10933602

UNESCO Recommendation on Open Science



Dimensions of Open Science

- **Improve the reproducibility and transparency of science: rigorous and responsible**
  - Open research data (and open code, protocols etc) 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.
- **Enhance the public good dimension of science: relevant**
  - Increase public access to the outputs of science
  - Increase public engagement with the conduct of science (citizen science, transdisciplinary research
  - **UNESCO Recommendation:** strong emphasis on global benefit; benefit for society and involvement of societal actors; very strong ethical dimension
- **Increase the economic, social impact of science: RoI**
  - **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.

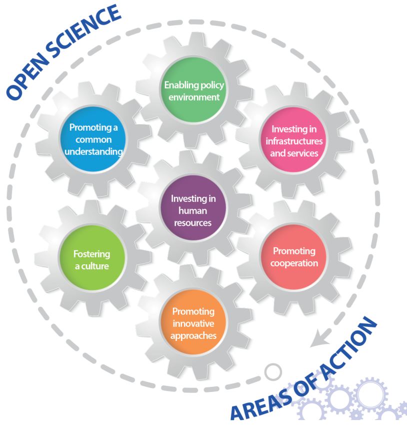


UNESCO Recommendation on Open Science

Open Science Transformations

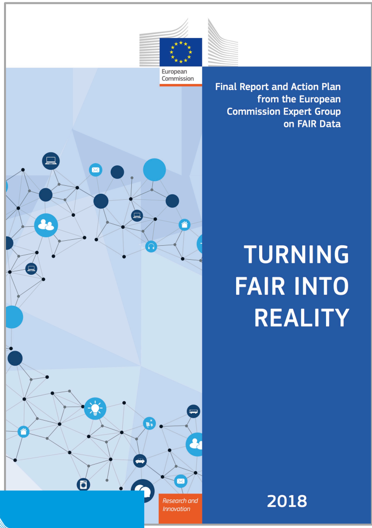
- **In the conduct of science**
  - Openness and transparency in methods, protocols, models, data (provenance).
  - Nature abhors a vacuum, (Open) Science abhors a black box…
  - As open as possible, *only* as closed as necessary.
- **In scientific communication**
  - Science as a global public good requires diamond OA and data / code deposit as a precondition of publication.
  - Requires (in many disciplines) the aggregation of data as an end in itself for which research groups are recognized.
- **In the engagement of science with society**
  - Increasing importance of mission oriented science (SDGs), of transdisciplinary research, of citizen or participatory science.
  - Need for a science that is more globally, linguistically, societally inclusive (in terms of participation, methodology, scrutiny of concepts and categories).
- **The future is here, it's just unevenly distributed**
  - Open data practices have (partially) transformed certain areas of research: genomics and related biomedical sciences; crystallography; biodiversity; astronomy; areas of earth systems science; various disciplines using remote sensing data…

UNESCO Recommendation on Open Science



The role of 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.
- **FAIR is neutral on Open**
  - FAIR does not necessarily mean Open.
  - FAIR can be just as beneficial for data that cannot be shared.
- **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.
  - FAIR makes data 'Fully AI Ready'…
- **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'.**
- **Lorentz Center Workshop: 'The Road to FAIR and Equitable Science'**
  - Three themes: machine-actionability, equity/fairness and AI.



A challenge! We need to transform how we think about data in science!

Need to transform how we think about data in science!

- Too much data is still being extracted, post-facto from academic articles… (biodiversity, pollination, nanomaterials…)
- Prevalence of the bibliographic, publication model for thinking about data.
  - Parsons, M.A. and Fox, P.A., 2013. Is Data Publication the Right Metaphor?. Data Science Journal, 12(0), p.WDS32-WDS46.DOI: https://doi.org/10.2481/dsj.WDS-042
  - Metadata is not just extended DC/DCAT and CERIF… !
  - The use case in which researchers deposit data in support of their publication is not the only, or the most important, use case.
  - Need to pay more attention to the role of data infrastructures that furnish data for researchers and support these practices.
  - This is how we reduce the cost of data wrangling and support cross-domain research
- **Need a shift from a bibliographic to an engineering approach (data orchestration?) to data stewardship, from the file level to the datum level, at least for high value data where interoperability and combination is essential.**
  - Two important use cases and drivers for this are interdisciplinary grand challenge research and fine-grained management of data access.
  - WorldFAIR Policy Brief https://doi.org/10.5281/zenodo.7853170
  - **Release the datums!**
- **Learn from and collaborate with integrative initiatives: GBIF, ODIS, SDG Data Cube, SDMX etc.**

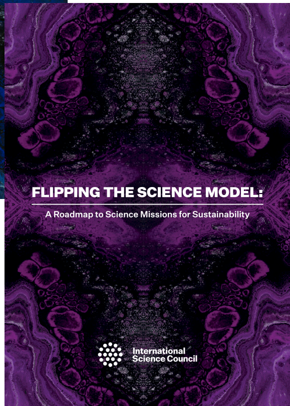
WorldFAIR

GBIF
Global Biodiversity
Information Facility

sdmx
Statistical Data and Metadata eXchange

Why do we need cross-domain integration?

- 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.
- Almost all the SDGs have interdisciplinary / cross-domain dimensions; either directly in the indicators and certainly in the any response.
- Examples: complex, multi-disciplinary and multi-stakeholder issues, ‘wicked problems’.
    - Ocean sciences: key example of human interaction with earth systems; many data dimensions; multi-stakeholder, economic and geo-political interests.
    - Population health: social survey data; clinical data; phylogenetic data; construction of complex queries and models; data protection.
    - Disaster risk reduction: human interaction with a hazard; modelling the hazard, loss, response.
- International Science Council Calls for **mission science for sustainability**.
    - Trans/international; ‘big science’ funding for action-oriented, **transdisciplinary and engaged science.**
    - Network of mission science centres / nodes, supported by a global knowledge sharing platform(s), implements the FAIR and CARE principles, data protection.
    - **Call for international, transdisciplinary pilot projects out now, deadline 31 May!**



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.**
- **Preparatory Phase from 2017:** exploring technical issues and case studies through Dagstuhl workshops, TGs and WGs, funded projects to understand the challenges and prepare the programme. Developed some of the WorldFAIR methodology and approach.
    - **Cross-Domain Interoperability Framework:** units, vocabularies, data structure, data description, provenance…
    - **Case Studies:** in a range of domain and cross-domain research areas: included infectious diseases/ epidemiology; urban health; DRR; Task Groups in Agriculture, Nanomaterials, etc.
- **WorldFAIR Project:** takes this approach forward in the context of an EU-funded project.
- **Global Open Science Cloud initiative:** enable implementation of UNESCO Recommendation, through Open Science platforms / e-Infrastructures.

WorldFAIR

GOSC
Global Open Science Cloud

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**
- Global project with beneficiaries outside the EU.
- Institutions in France, Belgium, Cyprus, Denmark, Germany, Ireland, Norway; Australia, Brazil, Kenya, New Zealand, USA; UK.
- Includes **authoritative international entities** (e.g. IUPAC, OneGeochemistry, GBIF, ODIS); connections with important projects or standards organisations (e.g. NanoCommons, DDI Alliance, OHDSI, TDWG, SalUrbAL).
- **Important partnership between CODATA and Research Data Alliance.**
- WorldFAIR website: https://worldfair-project.eu/
- WorldFAIR outputs: https://zenodo.org/communities/worldfair-project/
- WorldFAIR webinar series: https://bit.ly/WorldFAIR-Webinar-Series
- **Upcoming webinars: https://worldfair-project.eu/events/**
- **Special sessions at the virtual RDA Plenary (on 22 and 23 May)**

WorldFAIR

WorldFAIR Partners

AuScope

Australian National University

NATURAM PRIMUM COGNOSCERE RERUM

AWI

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

dri
Digital Repository of Ireland
Taisclann Dhigiteach na hÉireann

Embrapa

FAPED

GBIF

HiveTracks
know your bees!

IUPAC

KALRO
Kenya Agricultural & Livestock Research Organization

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Meise Botanic Garden

NovaMechanics
Cheminformatics & Nanoinformatics Excellence

RDA
RESEARCH DATA ALLIANCE

SALURBAL
Salud Urbana en América Latina

7P9
Seven past nine

Tonkin+Taylor

UNIVERSITY OF BIRMINGHAM

Interoperability Frameworks

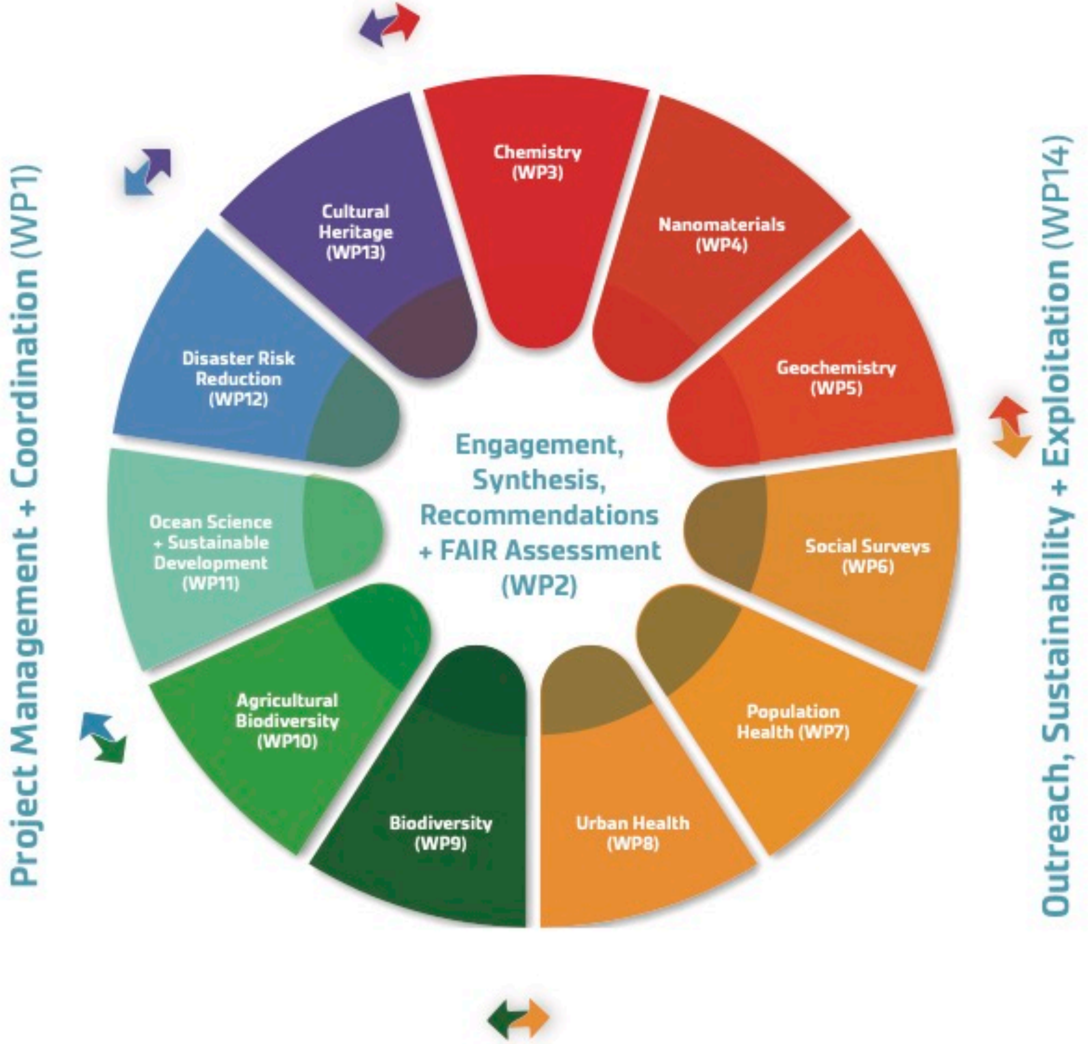
- Among the most important, but most challenging, recommendations of the **Turning FAIR into Reality** report, is R.4:
- ‘Develop **interoperability frameworks** for FAIR sharing within disciplines and for interdisciplinary research: Research communities need to be supported to develop interoperability 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.**’
- Very strong focus on the I and the R of FAIR.
- Core driver of CODATA-ISC Decadal Programme and WorldFAIR project

Community Agreements

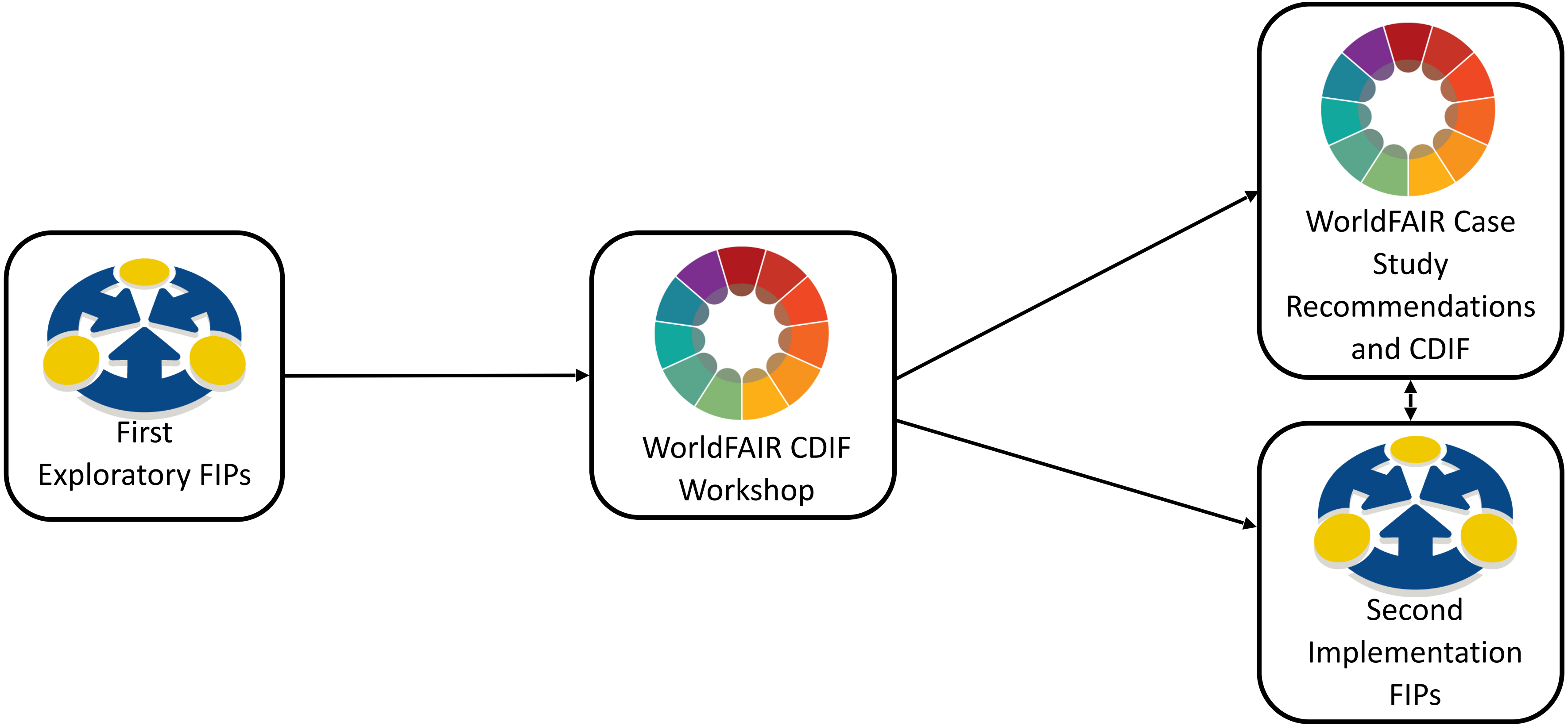
- UNESCO Recommendation calls on Member States, and other stakeholder, to promote:
- 'Community agreements, concluded in the context of regional or global research communities, and which define community practices for data sharing, data formats, metadata standards, ontologies and terminologies, tools and infrastructure. International scientific unions and associations, regional or national research infrastructures and journal editorial boards each have a role to play in helping develop these agreements. **In addition, convergence between the various semantic artefacts (particularly vocabularies, taxonomies, ontologies and metadata schema) is essential for the interoperability and reuse of data for interdisciplinary research.**' (UNESCO Recommendation on Open Science, VI.iii.18.f, p.24)

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 Methodology



WorldFAIR(+) Methodology

WorldFAIR(+) Methodology

1. Identify and scope new Case Studies (or Petals): i.e. a research group, institution, project with which an OSC is working.
2. FAIR Implementation Profile (FIP) Workshop(s)
3. Depending on the Case Study, this could involve one or many FIPs.
    1. Guided and collaboration creation of the FAIR Implementation Profile; can cover the ‘here and now’ and indicate aspirations.
4. (In development) methodology for modelling research processes to identify necessary provenance information and key organising concepts (in DDI terms ‘Unit Types’ and ‘Conceptual Variables’).
5. Interoperability Framework Workshop
    1. Takes the FIP(s) and the ‘organising concepts’ as starting points.
    2. Detailed, structured discussion of data model, metadata, semantics, representation, etc.
    3. Iterative discussion in relation to domain practice and the emerging CDIF.
    4. Recommendations for an Interoperability Framework from the Case Study.
6. Co-design an implementation plan and an aspirational FIP.

WorldFAIR

What are FAIR Implementation Profiles (FIPs)?

- A methodology for understanding the practices of a community in relation to FAIR.
- Developed by the GO FAIR initiative: ‘a collection of FAIR implementation choices made by a community of practice for each of the FAIR Principles’.
- Set of questions, relating to each of the FAIR principles and sub-principles, for data and metadata, that allow a ‘community’ to state the FAIR Enabling Resources it uses to make data and metadata FAIR for each of the principles.
- Supported by an online tool: https://fip-wizard.ds-wizard.org/
- Allows the publication of the FIPs as nanopublications (RDF with provenance), allowing machine referencing and visualisation of practice across a range of research areas.
- **Valuable as a tool to enable a community to reflect on current and potential practice to improve FAIRness.**
- **Report ‘FAIR Implementation Profiles (FIPs) in WorldFAIR: What Have We Learnt?’: https://doi.org/10.5281/zenodo.7378109**

FIP Wizard

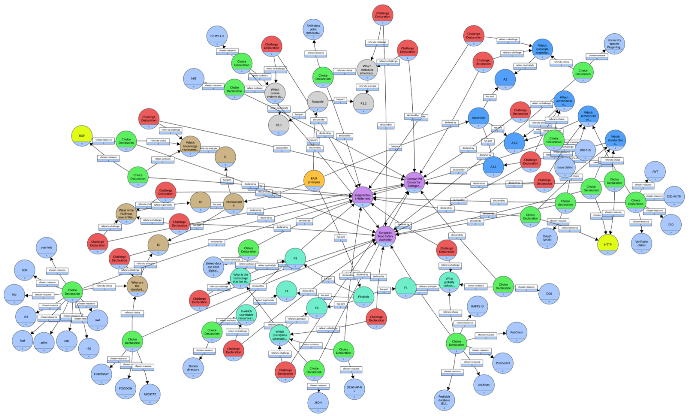


Figure from https://www.go-fair.org/how-to-go-fair/fair-implementation-profile/, Figure courtesy of Barbra Magagna, Umweltbundesamt GmbH and Kristina Hettne, CDS University Library Leiden

How have WorldFAIR Case Studies used FIPs?

- FIPs used to assist comparison of practice across the European Social Survey and the Australian Social Survey. Helped identify shared infrastructural needs (registries) and fed into recommendations.
    - See D6.1 ‘Cross-national Social Sciences survey FAIR implementation case studies’, pp.24-27: https://doi.org/10.5281/zenodo.7599652
- Used in analysis to identify ‘both generic and domain-specific (meta)data exchange conventions and FERs with high potential to bridge other WorldFAIR case studies to and with ODIS (Oceans Data Information System).
    - See D11.1 ‘An assessment of the Ocean Data priority areas for development and implementation’, pp.16-28: https://doi.org/10.5281/zenodo.7682399
- Two FIPs used to help understand 1, the ‘FAIR application of IUPAC standards in supporting chemistry data exchange’, and 2, ’the FAIR status of IUPAC standards for those who need to use them’.
    - See D3.1 ‘Digital recommendations for Chemistry FAIR data policy and practice’, pp.42-44 (discussion of FERs, FIPs and CDIF in chemistry and across domains) and 59-67 (FIPs): https://doi.org/10.5281/zenodo.7887283
- Extended discussion of the FIPs process and reflections on implications for diverse data in nanomaterials / nano-safety domains.
    - See D4.1 ‘Nanomaterials domain-specific FAIRification mapping’, pp.59-74 (As-Is’ Nanomaterials FIP): https://doi.org/10.5281/zenodo.7887341
- FIPs as a basis guide (work in progress) for good practice in urban health and for the SALURBAL project.
    - See D8.1 ‘Urban Health Data - Guidelines and Recommendations’, pp.11-14 (discussion) and pp.28-57 (FAIR Primer for the SALURBAL data platform): https://doi.org/10.5281/zenodo.7887523

FIPs: value to WorldFAIR

- Very useful tool for understanding current and potential FAIR practices in a given community.
- Assists in focusing reflection on practice, how this may be improved.
- Useful as a reference statement of good practice in a given community.
- Essential for the WorldFAIR approach, as it:
  - Helps identify shared practices across domains.
    - See analysis of this in WorldFAIR D11.1 'An assessment of the Ocean Data priority areas for development and implementation' https://doi.org/10.5281/zenodo.7682399
  - Helps towards the functional breakdown and solutions in CDIF.
- **Final synthesis of findings and experience, including final FIPs, matrix and graph of FIPs, and recommendations will be published by 31 May 2024!**

Aussi-ESS FIP published in WorldFAIR D6.1 Cross-national Social Sciences survey FAIR implementation case studies https://doi.org/10.5281/zenodo.7599652

9.2 Appendix Two: FAIR Implementation Profile for AUSSI-ESS (managed by ADA)

This table summarises the FAIR Implementation Profile for the ADA services supporting the AUSSI-ESS through the ADA Dataverse and related services.

Table 6 FAIR Implementation Profile (FER) for AUSSI-ESS – Australian Data Archive

FAIR Principle name	Referring to MetaData/Data	FIP question	FER Enabling Resource used in WP06 Social Surveys
F1	MD	What globally unique, persistent, resolvable identifier service do you use for metadata records?	DataCite DOI resolution service
F1	D	What globally unique, persistent, resolvable identifier service do you use for datasets?	DataCite DOI resolution service
F2	MD	What metadata schemas do you use for findability?	DDI Codebook Version 2.1
F2	MD	What metadata schemas do you use for findability?	DataCite metadata schema version 3.1
F3	D	What is the schema that links the persistent identifiers of your data to the metadata description?	No implementation choice has been made by this community
F4	MD	Which service do you use to publish your metadata records?	ADA Dataverse
F4	D	Which service do you use to publish your datasets?	ADA Dataverse
A1.1	MD	Which standardized communication protocol do you use for metadata records?	HTTPS|Hypertext Transfer Protocol Secure
A1.1	MD	Which standardized communication protocol do you use for metadata records?	REST|Representational state transfer
A1.1	D	Which standardized communication protocol do you use for datasets?	HTTPS|Hypertext Transfer Protocol Secure
A1.1	D	Which standardized communication protocol do you use for datasets?	REST|Representational state transfer
A1.2	MD	Which authentication & authorisation service do you use for metadata records?	None for open records; SAML2|Security Assertion Markup Language 2.0
A1.2	D	Which authentication & authorisation service do you use for datasets?	SAML2|Security Assertion Markup Language 2.0
A2	MD	What metadata preservation policy do you use?	RDA Core Trust Seal Certification
I1	MD	What knowledge representation language (allowing machine interoperation) do you use for metadata records?	JSON|JavaScript Object Notation
I1	MD	What knowledge representation language (allowing machine	XMLS|eXtensible Markup Language Schema

FAIR Implementation and FAIR Assessment

FAIR Implementation and FAIR Assessment

- A lot of effort being put into FAIR Assessment.
- ‘We now have so many FAIR Assessment tools, we will need a tool to assess these tools…’, Michel Schouppe, EOSC Lustrum event, Vienna, October 2023.
- The priority for WorldFAIR is to help enable research communities to develop, articulate and implement FAIR practices.
- How do we assess FAIRness effectively when practices are still emerging? Need for domain sensitive FAIR recommendations.
- **The FAIR Implementation Profiles horse should come before the FAIR assessment cart.**

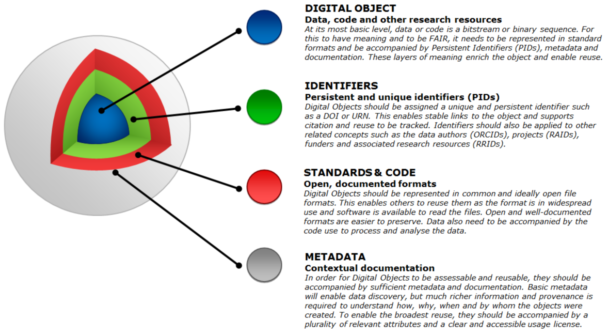
FIPs and CDIF

- Utility in identifying technical (metadata and semantic) points of contact across domains.
- FIPs and FER graph could provide a visualisation of practice.
- Genuine potential in a platform for sharing FIPs as readily visualisable and explorable.

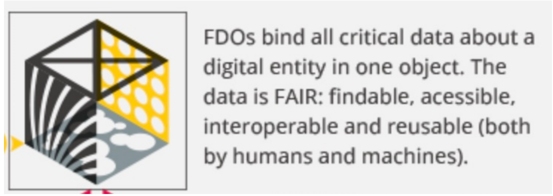
Image from https://www.phrases.org.uk/meanings/put-the-cart-before-the-horse.html

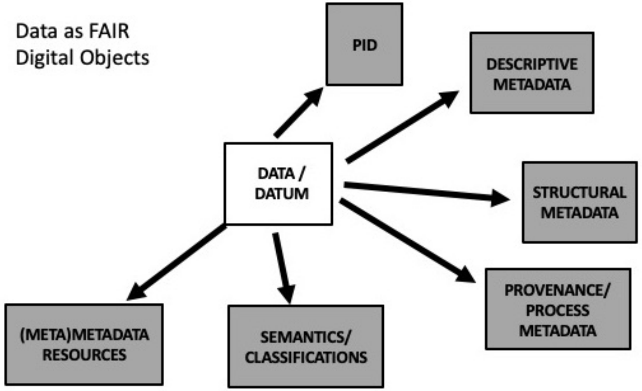
Machine Actionability

- **How can we increase the automation of data integration / combination?**
- **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, API)
- …
- **How can I process, combine it, analyse it?**
    - Concepts, variables and units (metadata and semantics)
    - Data structure
    - Provenance and processing
    - Quality, accuracy…



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





Terminologies and Semantics

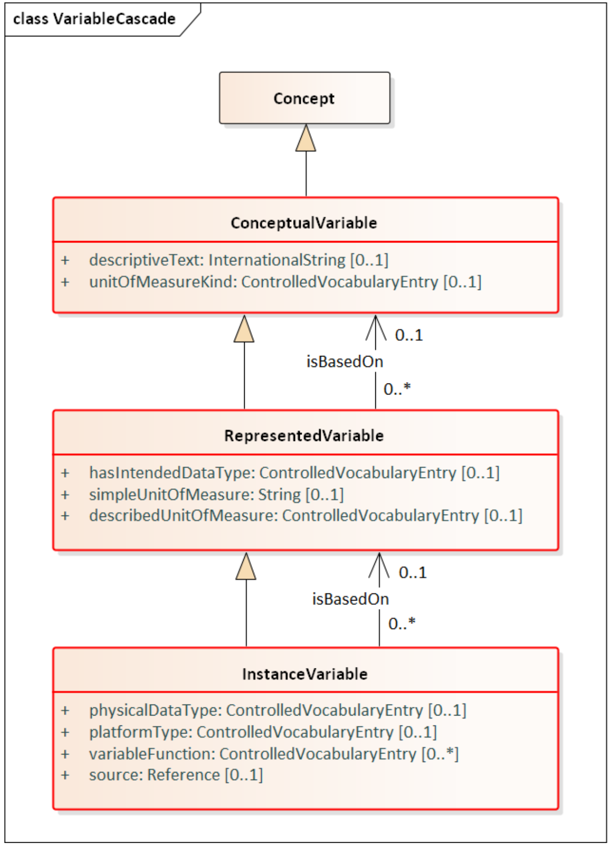
- **Need for digital (FAIR) representation of units:** Digital SI; cooperation and coordination, interoperability / translations with other systems for representing units.
- **Need for digital (FAIR) representation of data elements / conceptual variables:** online referenceable definitions of the quantity (metrology) / measurand / property / variable…
- Research ‘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**
- Vocabularies should represent scientific expertise and consensus, need to have good governance, need to be FAIR and usable by machines.
- There is a role and responsibility for ‘authoritative organisations’ to prepare and maintain FAIR vocabularies.
- Role for international scientific unions and similar bodies:
    - **IUSSP-CODATA Report on FAIR Vocabularies in Population Research: https://doi.org/10.5281/zenodo.7818157**
- Importance of initiatives like Research Vocabularies Australia, the NERC Vocabulary Server, OBO Foundry etc…
- Need for crosswalk registries, expressed in SSSOM.

INTERNATIONAL UNION OF PURE AND APPLIED CHEMISTRY

Bureau International des Poids et Mesures

Data Structure, Provenance, Universals

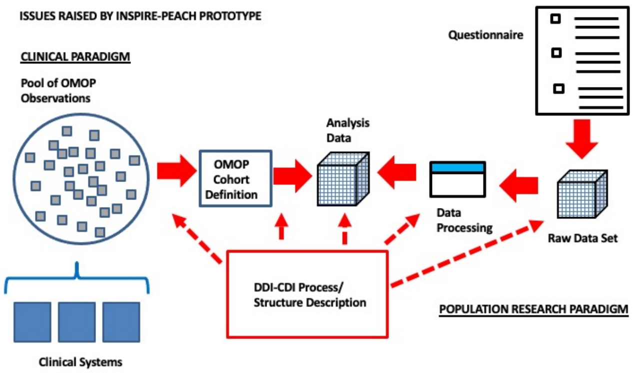
- Important to think about how we combine data for cross-domain research.
- Data Documentation Initiative (DDI) Cross-Domain Integration (CDI) specification contains three modules to assist with this:
  - **Structural Description:** assists processing of data structure transformations across four data structures.
  - **Data Description: Variable Cascade** describes data at an atomic level, describes relationships between concepts, representations and instances (assists with combining data and documenting information loss).
  - **Provenance and Processing:** module uses PROV-O and SDTL to provide and relay provenance and processing information.
  - Currently under final review for public release by DDI Technical Committee https://ddialliance.org/
  - See: The Role of DDI-CDI in EOSC: Possible Uses and Applications, final report https://doi.org/10.5281/zenodo.4707263
  - See: Optimising Your Data Description for Integration and Reuse https://bit.ly/ddi-cdi-workshop-materials



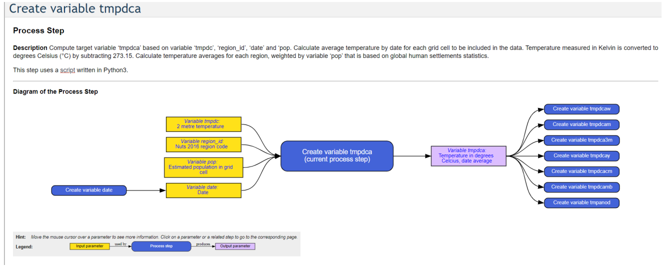
- **Applications: INSPIRE (WF WP07); EOSC Future Science Project 9 combines social survey data with environmental data (meteorological and air pollution); UKDA fine-grained access to sensitive data.**
- **Key topics of 2023 Dagstuhl Workshop https://bit.ly/Dagstuhl-CDIF-2023**

Interdisciplinary data combination

- **INSPIRE PEACH / Data Science Without Borders:** Collaborations with African Population Health Research Centre and others. WorldFAIR Case Study / WP07
    - Combines population health (HDSS) and clinical data, WHO questionnaires, phylogenetic data.
    - Uses the CDIF discovery approach (combining schema.org/JSON-LD with OMOP).
    - Uses DDI-CDI for structural and provenance/process description.
    - See Implementation Guide (D7.1 https://doi.org/10.5281/zenodo.7887385) and Resources (D7.2 https://doi.org/10.5281/zenodo.10010936).

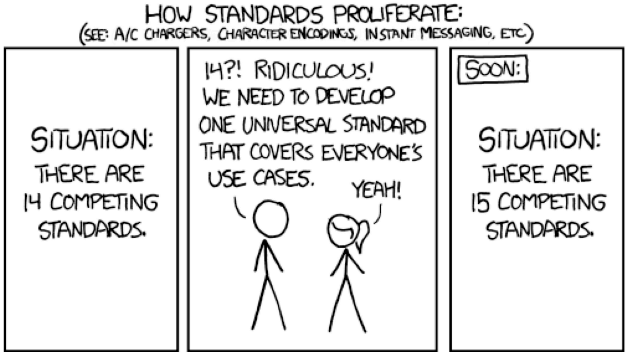


- **EOSC Future Science Project 9 'Climate-Neutral and Smart Cities':** Not WorldFAIR but involves CODATA and Sikt colleagues and uses DDI-CDI.
    - Combines social survey data, climate data and pollution data (European Social Survey (ESS) Social Survey data; European Environmental Agency (EEA) air quality data; European Centre for Medium Range Weather Forecasts reanalysis (ERA5))
    - Uses DDI-CDI to describe the data integration process and variable computation.
        - Allows researchers to review and understand the new compound variables and determine how they can be analysed.
        - Facilitates increased automation of the production of these compound variables as a time series.
    - See https://eoscfuture.eu/data/climate-neutral-and-smart-cities/

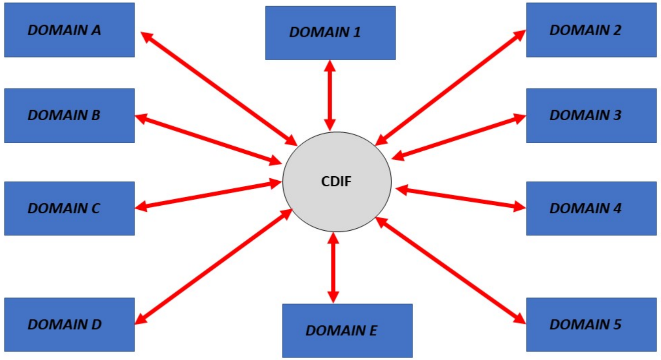


What is the CDIF (Cross-Domain Interoperability Framework)?

- Roots in CODATA-ISC work (nanomaterials, Making Data Work…) and then in the DDI-CODATA collaboration (particularly DDI-CDI).
- Adds (significant) detail to the EOSC Interoperability Framework and the Semantic Interoperability recommendations.
- Pushes beyond the F and A, to the I and R (and harder parts of A).
- **Identifies a set of functional requirements for interoperability, particular for steps in data combination, and identifies good practices for each of these requirements.**
- Use cases: domain or cross-domain projects or data services that need to combine data for analysis, modelling etc.
- **Categorically not a new standard. Rather it is a framework of existing and emerging standards.**
- A framework of standards/specifications to provide a *lingua franca*.
- Significant proportion of CDIF rests on good web practice, domain neutral standards and good practice: disciplines can adopt or map.



Source: xkcd.com



CDIF Functional Requirements and Standards/Specifications

- Aims to discuss and recommend good practice and standards for a set of functions related to the FAIR principles
  - Discover **(DCAT, schema.org as JSON-LD): https://doi.org/10.5281/zenodo.10252564**
  - Access, negotiate access to non-public data **(ODRL, DPV, DUO)**
  - Assess and Integrate/Combine:
    - Understand and transform data structure **(DDI-CDI)**
    - Understand and align semantics **(SKOS/XKOS, OWL, SSSOM)**
    - Determine origin/context **(PROV-O, I-ADOPT/O&M)**
    - Understand processing **(DDI-CDI, SDTL)**
  - 'Universals' **(time, place, units of measurement, authoritative domain recommendations)**;
  - Organising concepts and connectors **(e.g. 'unit types' and 'conceptual variables')** *NB: different terms for these in different domains…*
  - Metadata for ML; Packaging…
- CDIF WG subgroups on Design Principles, Discoverability, Data Integration, API/Architecture, Semantics: https://worldfair-project.eu/cross-domain-interoperability-framework/
- **V01 CDIF to be published on 31 May; consultation on as many components as possible prior to this…**

WorldFAIR

WorldFAIR Recommendations

- **It is essential for work on FAIR / metadata to retain an international dimension.**
  - WorldFAIR shows how important it is to engage with international data initiatives and standards organisations, and demonstrates how this can be done.
- **Cross-domain research is essential to scientific, social and European policy/strategic priorities.** Supporting the needs of cross-domain research, with FAIR data and metadata should be an EOSC priority.
- **The Cross-Domain Interoperability Framework (CDIF) can be useful for Helmholtz centres, NFDI and other initiatives for interoperability of their data and metadata.**
  - Support the further development and refinement of the CDIF.
- **Put the FAIR Implementation Profiles horse ahead of the FAIR Assessment cart!**
  - Support and help refine the FAIR Implementation Profiles methodology and tooling (developed by GO FAIR).
- **Pay attention to the outputs from the 11 WorldFAIR Case Studies.**
  - Support further development of the WorldFAIR approach to ‘Modelling Research Processes’, for provenance information and organising concepts.
- For some of these recommendations, see the WorldFAIR Policy Brief https://doi.org/10.5281/zenodo.7853170

WorldFAIR

WorldFAIR Recommendations

- **Need a shift from a bibliographic to an engineering approach to data stewardship, from the file level to the datum level, at least for high value data where interoperability and combination is essential.**
    - Need to enable research data infrastructures to be able to do this. This is how we reduce the cost of data wrangling (cf. PWC report).
    - Two important use cases and drivers for this are interdisciplinary grand challenge research and fine-grained management of data access.
- Arguments and recommendation presented in the WorldFAIR Policy Brief https://doi.org/10.5281/zenodo.7853170

WorldFAIR+ Activities and Next Steps

WorldFAIR+

- Aim to create a federation of additional Case Studies, with parallel funding and supported by a coordinating mechanism with technical expertise.
- Seeking to identify mechanisms for additional resources and contribution of expertise and effort for the coordinating function.

Concrete Progress

- **New ISC supported Case Studies in emergencies data (earthquake and flooding/cholera).**
- **Project proposals: for an IUGS Deeptime Digital Earth (DDE) Case Study.**
- Preparing OSCARS proposal(s) to explore implementation of CDIF in domain, cross-domain contexts.
- Discussions with various partners and CODATA members (including CESSDA; ARDC, Australia; Helmholtz Metadata Collaboration, Germany; KISTI, Korea; Malaysian Open Science Platform; African Open Science Platform)
- MoUs concluded with CESSDA, DDE, and planned with ARDC.
- **Keen to discuss with any additional Case Studies or partners for coordination! Can we identify Polish case studies, in important domain or cross-domain areas?**

WorldFAIR

Thank you for your attention

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

CODATA Officers

CODATA President

- **Mercè Crosas**, head of Computational Social Sciences, Barcelona Super Computing Centre.
- See https://codata.org/about-codata/message-from-president-merce-crosas/

CODATA Officers

- **Richard Hartshort, Vice President**, Professor of Chemistry, University of Canterbury, New Zealand.
- **Daisy Selematsela, Vice President**, Director of Wits Libraries, South Africa.
- **Christine Kirkpatrick, Secretary General**, **Greenfield**, Data, Metadata and Systems Expert, INSPIRE projects and DSWG.
- **Steve McEachern, Treasurer**, Director of the Australian Data Archive.
- **Barend Mons, Past President**, Professor of Biosemantics, Leiden University Medical Centre.
- **See** https://codata.org/about-codata/executive-committee/

CODATA Executive Committee

Elected Members

- **Toshihiro Ashino**, Professor, Toyo University, Japan.
- **Tyng-Ruey Chuang**, associate research fellow at the Institute of Information Science, Academia Sinica, Taiwan.
- **Leo Lahti**, professor in Data Science at the Department of Computing, University of Turku, Finland.
- **Pam Maras**, Past President IUPsyS, Emerita Professor of Psychology at the University of Greenwich, UK
- **Audrey Masizana**, Senior Lecturer in Computer Science University of Botswana.
- **Virginia Murray**, Head of Global Disaster Risk Reduction, UK Health Security Agency; ISC Fellow.
- **Marc Nyssen**, IUPESM, professor emeritus, medical informatics and biomedical engineering.
- **Elena Rovenskaya**, Advancing Systems Analysis (ASA) Program Director at the International Institute for Applied Systems Analysis (IIASA), Austria.
- **Andrew Young**, Chief Research Scientist at the Commonwealth Scientific and Industrial Research Organisation, Australia; director of National Research Collections.
- **Cyrus Walther**, IUPAP, President of the International Association of Physics Students; ISC Fellow.

Ex Officio Members

- **Francis Crawley**, Good Clinical Practice Alliance, Belgium; Chair of the CODATA International Data Policy Committee.
- **Shaily Gandhi**, co-lead of the CODATA Connect Early Career and Alumni initiative; ISC Fellow.
- **Mark Parsons**, Research Scientist for NASA at the University of Alabama in Huntsville; co-editor-in-chief of the CODATA Science Journal.
- **Lili Zhang**, research scientist in the scientific data center at the Computer Network Information Center of Chinese Academy of Sciences; Executive Director of the CODATA International Programme Office for the Global Open Science Cloud.
- See https://codata.org/about-codata/executive-committee/

CODATA Secretariat: Increased Capacity

CODATA Secretariat

- **Arofan Gregory**, Metadata and Technical Expert, 'Making Data Work', WorldFAIR and INSPIRE projects
- **Simon Hodson**, Executive Director
- **Asha Law**, Programme Assistant: communications, technical platforms and website updates
- **Laura Molloy**, Senior Research Lead: WorldFAIR and RDA Tiger projects, vocabularies and terminologies, project and research development
- **Hana Pergl**, Operations Manager: strategic and management activities, membership relations, GOSC
- info@codata.org

Simon Hodson

Laura Molloy

Asha Law

Arofan Gregory

Hana Pergl

CODATA Team on INSPIRE Projects

- **Jay Greenfield**, Data, Metadata and Systems Expert, INSPIRE projects and DSWG.
- **Dorothy Mailosi**, Research Assistant, INSPIRE Network and MH Projects.
- **Doug Fils**, Data and Software Craftsman with a focus in semantics, linked data and web architecture, INSPIRE Network project.

CODATA Task Groups and Working Groups

Task Groups

1. Advancing Data Science for Sustainability
2. CODATA-WDS TG on Citizen-generated data for the SDGs
3. Data driven social change towards society promoting cognitively healthy aging
4. Data Ethics Task Group (DETG)
5. Data Systems, Tools, and Services for Crisis Situations Task Group (DSTS_CS-TG)
6. Digital Representation of Units of Measurement (DRUM)
7. FAIR Data for Disaster Risk Research (FAIR-DRR)
8. Geographical Indications Environment & Sustainability (GIES)

See https://codata.org/initiatives/task-groups/

Standing Task Group

1. Task Group on Fundamental Constants

Working Groups

1. integrity in data and AI science, ethics, and policy (Integrity-WG)
2. One Geochemistry Working Group
3. GOSC Working Groups

See https://codata.org/initiatives/working-groups/

Interested people can contact co-chairs to join TGs and WGs.

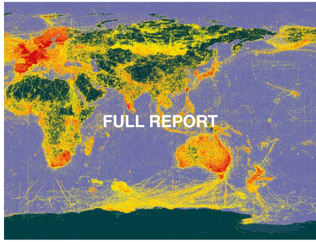
Making Data Work for Global Grand Challenges

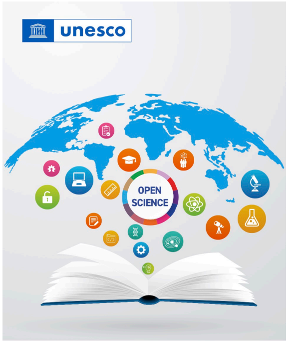
- **Making Data Work… and WorldFAIR+**
- Impact: The impact of this work will be to **better enable interdisciplinary science** for grand challenge issues, by **improving capacity to combine data and metadata across domains and ensuring that science missions for sustainability are supported by good data practices**.
- Flagship initiatives:
    - WorldFAIR Project > WorldFAIR+ initiative.
        - Part of the ISC Action Plan. Looking at FAIR (particularly I and R) for interdisciplinary science.
        - Main product is the Cross-Domain Interoperability Framework (CDIF).
        - Case Studies and collaboration led to the WorldFAIR Project: https://worldfair-project.eu/
        - Expanding and sustaining the WorldFAIR initiative as WorldFAIR+
        - Importance of FAIR vocabularies: see IUSSP-CODATA Report on FAIR Vocabularies in Population Science https://doi.org/10.5281/zenodo.7818157
        - Working with GOSC to align with WorldFAIR methodology.

Improving Data Policy

- Impact: The impact of this work will be to **support the effectiveness and ethical dimensions of global science by improving Open and FAIR data policies**, and by ensuring that data policy is an integral part of science policy.
- Towards a FAIRer World Workshop, March 2023: https://codata.org/events/science-and-policy-workshops/towards-a-fairer-world/
- Flagship initiative: **International Data Policy Committee (IDPC)**
  - New IDPC Action Plan: https://bit.ly/IDPC-Action_Plan
    - Data policy for data quality, reliability, and integrity
    - Data policy for science in crisis situations
    - Data policy for education
    - Data policy for AI
    - Data policy for Open Science
    - Data policy for the publication and communication of science
  - CODATA-UNESCO WG on Data Policy for Emergencies: **will present Data Policy for Emergencies, contribution to the UNESCO Open Science Toolkit at UN Data Forum in Nov 2024**

TWENTY-YEAR REVIEW OF GBIF





UNESCO Recommendation on Open Science

Towards a FAIRer World

Implementing the UNESCO *Recommendation on Open Science* to address global challenges

A Symposium co-organized by UNESCO, International Science Council (ISC) Committee on Data (CODATA) and World Data System (WDS)

UNESCO, Paris & Virtual; Wednesday, 29 March 2023

Advancing the Science of Data and Data Stewardship

- Impact: The impact of this work will be to help **advance scientific practices around data** by raising the profile of **the science of data** and by **stimulating dialogue between practitioners of data science and data stewardship.**
- **Interface of data science, the science of data, data stewardship and AI.**
- **Transparency and reproducibility in science.**
- Flagship initiatives:
  - Data Science Journal
  - International Data Week (IDW) and SciDataCon
  - Working Groups and Task Groups
  - Task Groups on Fundamental Constants (TGFC) and Digital Representation of Units of Measure (DRUM)
  - RDM Terminology https://bit.ly/CODATA-RDM-Terminology

23-26 OCT 2023 SALZBURG

International Data Week

A FESTIVAL OF DATA

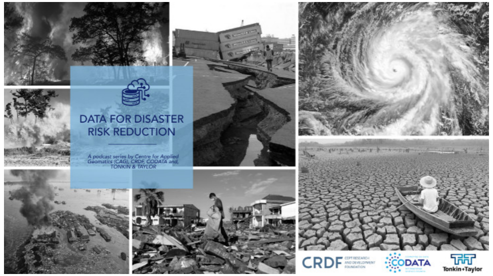
ARDC

Research Vocabularies Australia

Enhancing Data Skills

CODATA – RDA Data Schools

- Impact: The impact of this work will be to **increase capacity for data stewardship**, through targeted and effective **training and train-the-trainer initiatives**, and by facilitating the engagement of early career researchers in data issues.
- Flagship initiatives:
  - CODATA Connect (Early Career and Alumni Group)
    - Current podcast series on Open Geo AI: https://crdf.org.in/podcast/open-geo-ai-unveiling-satellite-insights-through-open-data
  - CODATA-RDA Schools of Research Data Science.
  - CODATA-CNIC Training Workshops



What does WorldFAIR Contribute?

- **Essential international dimension:**
  - Examples from global infrastructures (GBIF, ODIS), strong links with international standards organisations (IUPAC, TDWG, DDI, W3C, OGC, etc), and international stakeholders (UN Stats, UNDRR, PAHO, ISC, CODATA, RDA).
- **11 rich and detailed domain and cross-domain case studies with recommendations for FAIR Implementation.**
- **11 completed, current and prospective, FAIR Implementation Profiles**, encoded as nanopublications and potentially visualisable as a graph, showing good practice, connections and convergence.
- **The Cross-Domain Interoperability Framework:**
  - Adds specific detail to the EOSC Interoperability Framework, for data and metadata interoperability across domains.
  - Identifies a set of functional requirements for cross-domain interoperability.
  - Provides a set of recommendations, in some cases with implementation guidelines and examples from WorldFAIR Case Studies or related work.
- **Recommendations on FAIR Assessment.**
  - Effort should be put into FAIR Implementation Profiles, so we better understand domain requirements and as a precondition to FAIR assessment.
- **Pilot of methodology for modelling research processes to identify necessary provenance information and key organising concepts (in DDI terms ‘Unit Types’ and ‘Conceptual Variables’).**

WorldFAIR