WorldFAIR+
Making data work for cross-domain grand challenges

Simon Hodson, Executive Director, CODATA
(and slides from IUPAC WorldFAIR Case Study)

ISC project updates:
Converging Science & Technology in a Digital Era
Online
18 September 2024

Making Data Work...   Promoting Data Policy   Data science and AI for science

- **Making Data Work for Cross Domain Grand Challenges**
- **WorldFAIR Project and WorldFAIR+**
- Recommendations for the Cross-Domain Interoperability Framework (CDIF)
- FAIR Vocabularies with ISUs
- Cross-Domain Case Studies
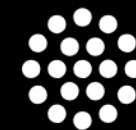- Global Open Science Cloud initiative
- Regional Open Science Platforms

- International Data Policy Committee http://bit.ly/data-policy-committee
- Data Policy in Times of Crisis (UNESCO Open Science Toolkit) https://bit.ly/UNESCO-CODATA-DPTC
- Major policy reports: https://bit.ly/CODATA-Policy-Reports

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

- CODATA Connect ECR Group
- 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

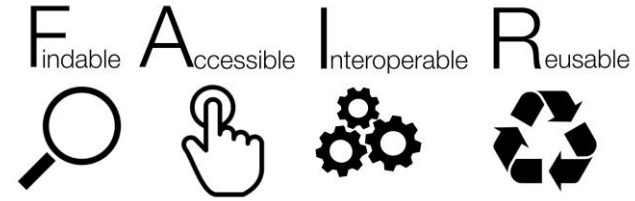# Making Data Work – WorldFAIR – WorldFAIR+

Making Data Work for Cross-Domain Grand Challenges

- The digital and data revolution presents us with huge opportunities and significant challenges.

- **Premise:** 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 a transdisciplinary approach to turn data into knowledge and then into action.

- ISC Action Plans entrusted CODATA with an initiative 'Making Data Work for Cross-Domain Grand Challenges': establish a global (decadal) programme to address these issues.

- **ISC provided funding support for a Preparatory Phase:**

  - Exploratory workshops with Unions and standards organisations.

  - Developed a case study driven methodology.

  - Established a very strong collaboration with the DDI Alliance.

  - Jointly explored cross-domain interoperability issues at a series of Dagstuhl workshops: https://codata.org/initiatives/decadal-programme2/dagstuhl-workshops/

FAIR Principles

- High level principles and guidelines to how to make scientific data as usable as possible.

- Extremely influential (5453 citations Nature; 14244 citations Google Scholar).

  - **Findable:** Does the data have an unambiguous, persistent identifier? Is the data described with metadata to allow it to be discovered?

  - **Accessible:** Is there a clear technical protocol for accessing the data? Are the conditions for access clear and explicit? (Access ≠ Open)

  - **Interoperable:** Are there agreed, authoritative and machine-encoded definitions of key concepts, methods, observations, measurands, variables, units… so I can combine the data?

  - **Reusable:** Is there clear licensing and fair use information, so I know what I may do with it? Is there clear provenance, quality, error, certainty information, so I know what I can do with it?

- **FAIR Implementation Profiles:** a method, developed by GO FAIR, with input from CODATA, to enable disciplines, communities to articulate their response to the FAIR principles.

- **Essential to work on cross-domain interoperability while recognizing the essential role of each research discipline.**

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

Barend Mons and Mercè Crosas, past and current CODATA Presidents, both authors of the FAIR Principles.

Turning FAIR Into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data, 2018, Hodson (chair of working group/lead author), et al., https://doi.org/10.2777/1524

WorldFAIR: Global cooperation on FAIR data policy and practice

- Advances in FAIR implementation in **cross-domain** scenarios, in 11 specific **disciplines** and **globally**.

- **Global** in approach, because research domains, data and metadata standards and specifications need to be global. Leveraged **CODATA and RDA networks** to achieve this.

- 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).

- Considerable emphasis on **case studies** and the recommendations from these organisations.

- Leveraged links to international standards and scientific organisations, as well as reliable articulations of good (web) practice to make cross-domain recommendations. **Were able to have funded partners outside the European Union.**

- Helps reinforce **bidirectional links between EOSC and global developments**.

- **Funded by the European Union, HORIZON-WIDERA-2021-ERA-0 — Project: 101058393**

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)
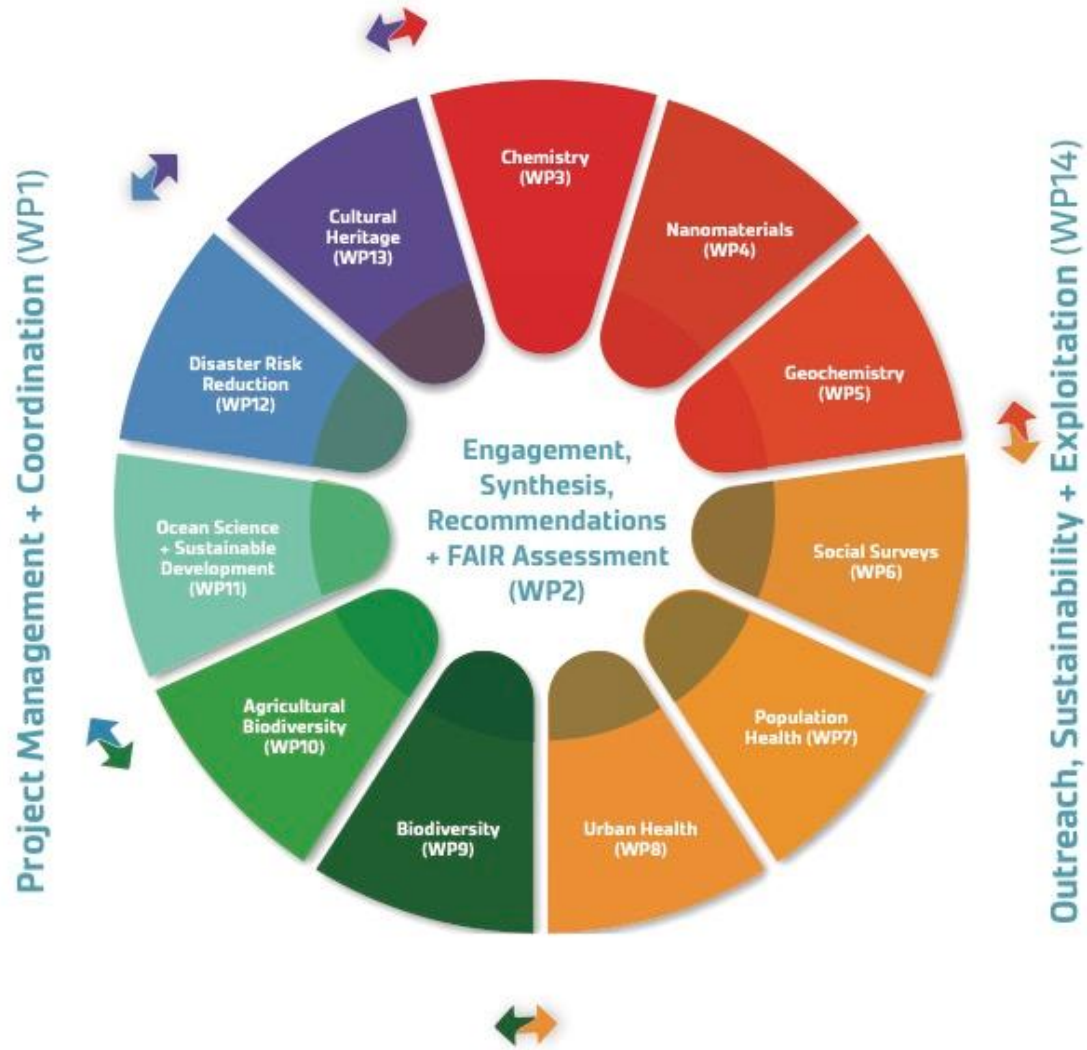
WorldFAIR Methodology
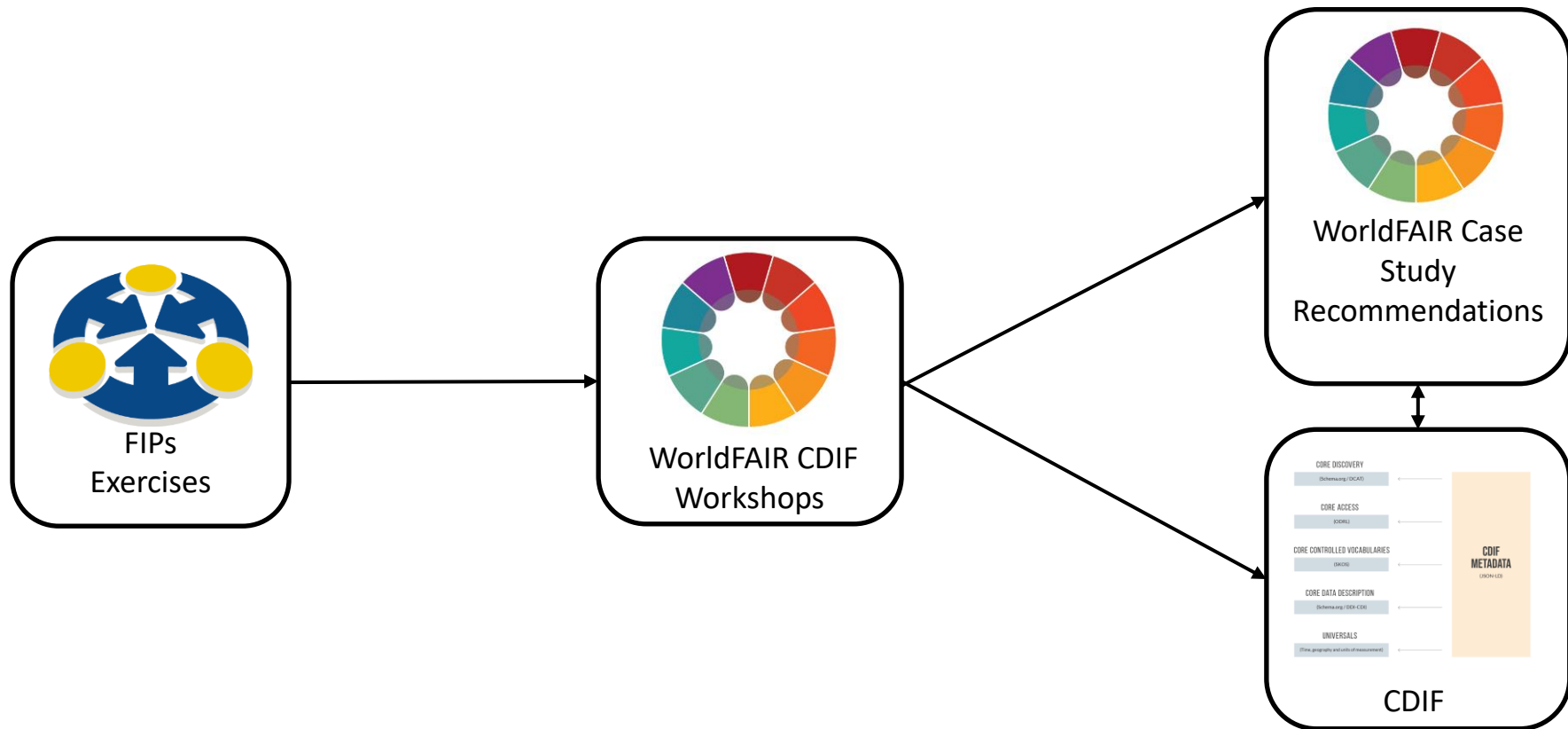
FAIR-enabling chemistry data standards/tools

FAIR attributes	Functionality	Chemical notations (examples)
Findable metadata schema	Indexing, matching	InChI, nomenclature
	Searching	Chemical notations (e.g., SMILES), terms (e.g., properties, methods)
Accessible retrieval protocols	Searching, retrieving (APIs) *(consistent across systems)*	Chemical structure resolver *(API spec prototype in WFC)*
Interoperable knowledge representations, vocabularies, metadata references	File formats for chemical entities and experimental measurements	SDF, CIF, ThermoML, JCAMP-DX, mzML
	Referrable terms and definitions	Gold Book, VIM, MeSH
	Classification, modeling	CHMO, RXNO, ChEBI, *FAIRSpec*
Reusable validation services	Completeness, consistency	checkCIF

Adapted from **GO FAIR**

Envisioning IUPAC at the World's FAIR hub

WorldFAIR

Practical Resources *(in progress)*	Descriptions	WorldFAIR inputs
Cheminformatics Color Book	• Best practices for digital chemical data notation	• D3.1, Enabling Guidance • FIP analysis, CDIF
Global Chemical Representation Resolver	• Notate & validate • Cross-link & federate	• D3.3, API Protocols • InChI in other WPs
IUPAC FAIR Chemistry Cookbook	• Interactive training toolbox • Atomized demos & workflows	• D3.2, Digital Recipes • Collaborations
IUPAC Gold Book Compendium	• Source terminology for data models, metadata	• FIP analysis, CDIF • Collaborations
Chemistry Data Standards Map	• Knowledge graph of data standards parameters	• FIP analysis • Workshops
Digital Units & Quantity Converter	• Digital representations of property measurements	• CDIF *(Events & Samples)* • CODATA DRUM

WorldFAIR Chemistry

➔ *develop **guidelines, training materials and tools** that facilitate use of standards*

WorldFAIR

D3.1 FAIR Chemistry Data Guidance

IUPAC Roadmap & Sustainability Blueprint

D3.2 FAIR Chemistry Training Cookbook

D3.3 FAIR Chemistry Protocol Services

InChITRUST

CCDC

NFDI4Chem

PSDI PHYSICAL SCIENCES DATA INFRASTRUCTURE

PubChem

GO FAIR

CODATA COMMITTEE ON DATA INTERNATIONAL SCIENCE COUNCIL

https://zenodo.org/communities/fairchemistry

RDA RESEARCH DATA ALLIANCE

WorldFAIR Outputs

- See WorldFAIR Final Message from Coordinator https://bit.ly/WorldFAIR-Final-Coordinator-Message and https://worldfair-project.eu/worldfair-plus/
  - **Summary of project, links to key outputs and next steps.**
- Reports, recommendations, guidelines, implementation examples and training materials from 11 Case Studies: https://zenodo.org/communities/worldfair-project/records and https://worldfair-project.eu/project-outcomes-and-documents/
  - **So many useful materials for all the subjects covered by WorldFAIR!**
- Policy Recommendations: https://doi.org/10.5281/zenodo.11242702 :
  - **We urgently need to shift from a bibliographic to an engineering approach to data stewardship.**
  - **We need metadata uplift to support interdisciplinary research and the responsible use of AI.**
- Cross-Domain Interoperability Framework (CDIF): https://doi.org/10.5281/zenodo.11236871 **(approaching 2000 downloads!)**
  - **A practical guide to FAIR implementation! Adopt widely used web standards and use them in line with good practice.**

WorldFAIR

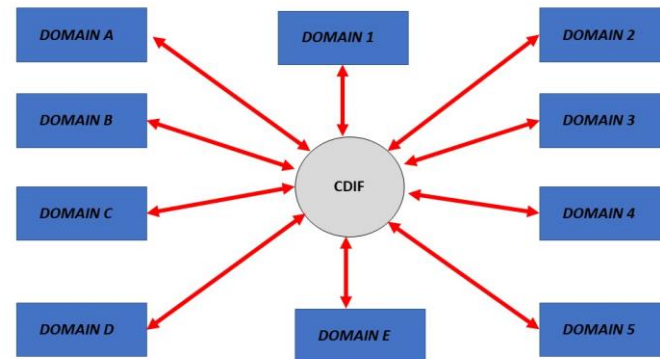| CORE DISCOVERY		
(Schema.org / DCAT)		
CORE ACCESS		CDIF METADATA
(ODRL)		(JSON-LD)
CORE CONTROLLED VOCABULARIES		
(SKOS)		
CORE DATA DESCRIPTION		
(Schema.org / DDI-CDI)		
UNIVERSALS		
(Time, geography and units of measurement)		
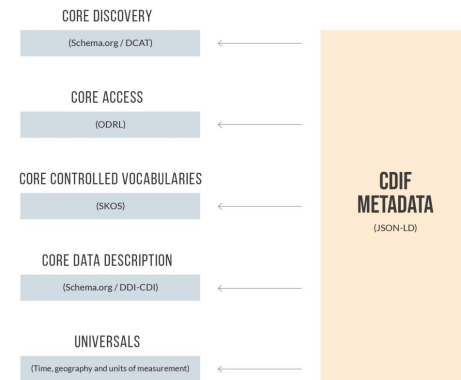
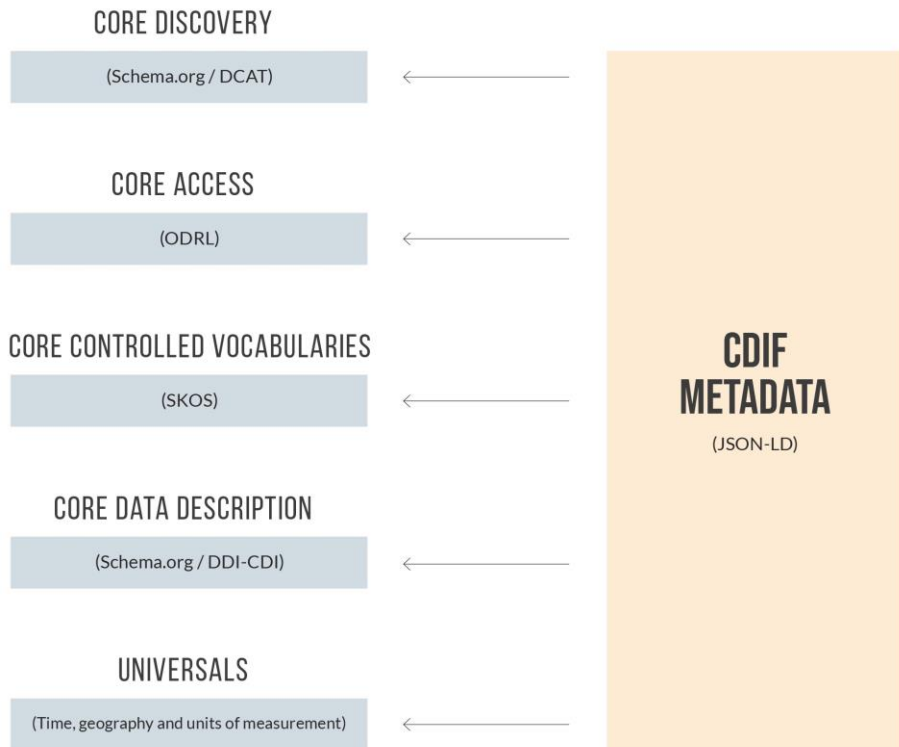# What is the CDIF (Cross-Domain Interoperability Framework)?

- 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.
- Directed at implementers: use cases, standards, how to implement them.

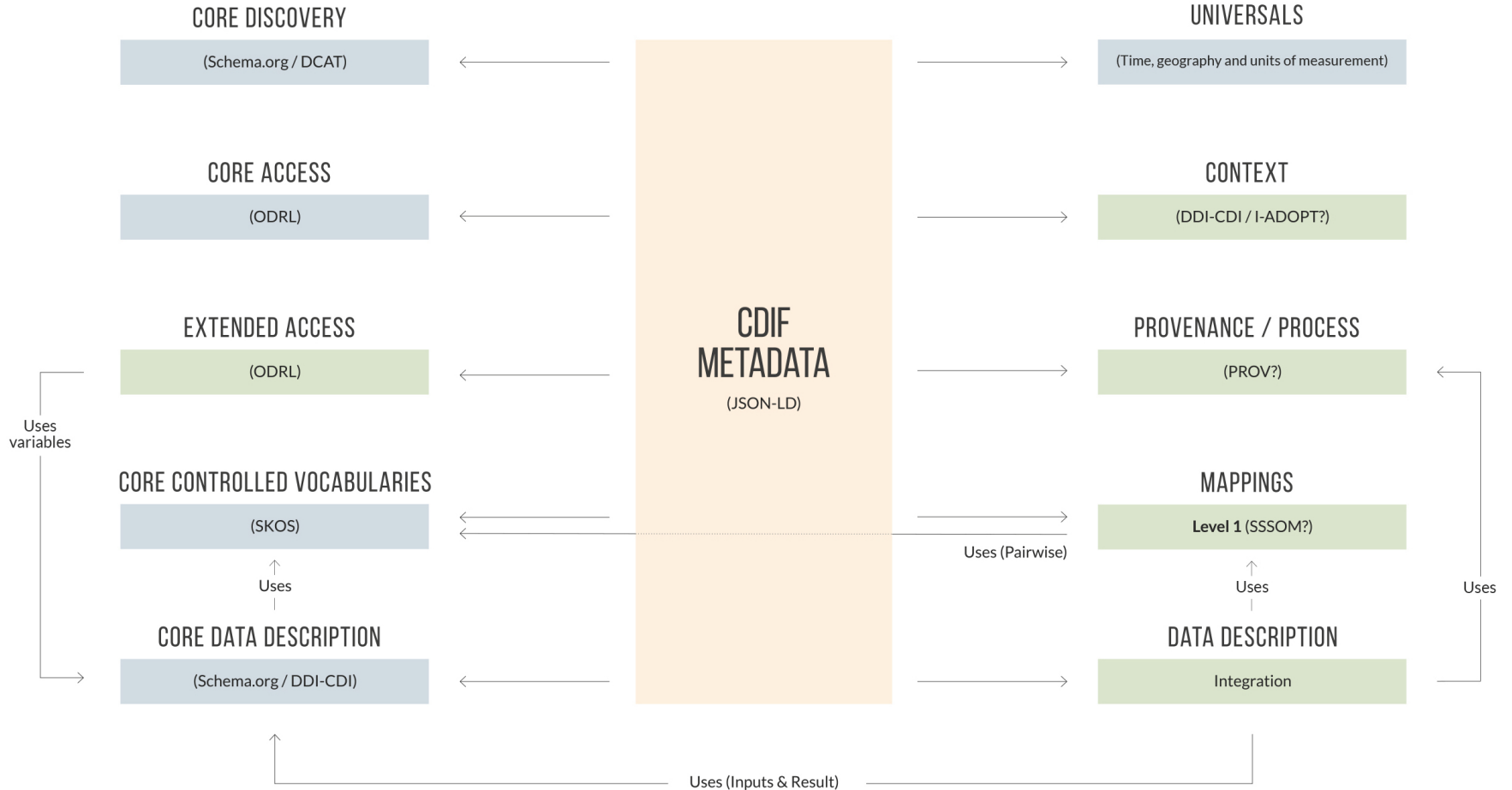# CDIF, the Cross-Domain Interoperability Framework

CORE DISCOVERY

(Schema.org / DCAT)

CORE ACCESS

(ODRL)

CORE CONTROLLED VOCABULARIES

(SKOS)

CORE DATA DESCRIPTION

(Schema.org / DDI-CDI)

UNIVERSALS

(Time, geography and units of measurement)

CDIF METADATA

(JSON-LD)
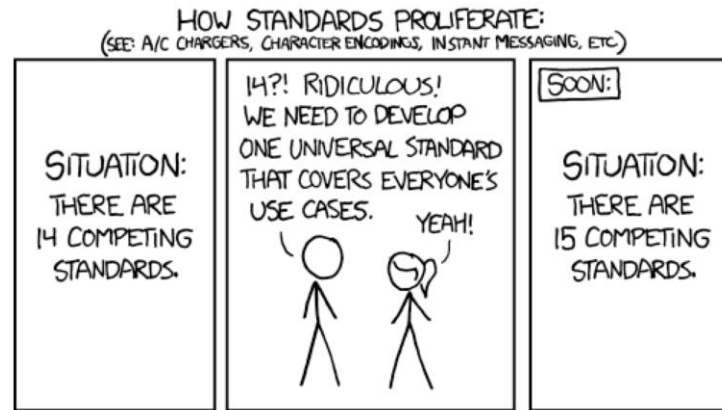
Feedback received from a colleague at a national data infrastructure

"I have a long reading list that I'm working through and initially wasn't too excited to be sitting down to read another technical report, and a massive one at that, but as I started reading, it was like it stopped me in my tracks to ask **"Is it your job to try to work out the design of metadata for a cross-domain repository and would you like us to tell you how you might do that in the best, most FAIR way?" to which I had to reply "Yes, yes that's exactly what I'm trying to do…".** This will make a real difference to guide and frame what we're doing and save me much time by recommending best practices and summarising choices that we would be making along the way. It gives us an achievable first scope for our metadata but will allow us to grow this over time as CDIF develops beyond version 1. We were reassured by how well it aligns well with what we were thinking of doing — we had independently already decided to start at the top-level with a SKOS vocabulary and DCAT catalogue description. As such, we intend to use your report as a starting point for how to implement this and build it up as far as it fits with our project plan."

CDIF, Next Steps

CORE DISCOVERY

(Schema.org / DCAT)

UNIVERSALS

(Time, geography and units of measurement)

CORE ACCESS

(ODRL)

CONTEXT

(DDI-CDI / I-ADOPT?)

EXTENDED ACCESS

(ODRL)

PROVENANCE / PROCESS

(PROV?)

CDIF METADATA

(JSON-LD)

Uses variables

CORE CONTROLLED VOCABULARIES

(SKOS)

MAPPINGS

Level 1 (SSSOM?)

Uses (Pairwise)

Uses

Uses

Uses

CORE DATA DESCRIPTION

(Schema.org / DDI-CDI)

DATA DESCRIPTION

Integration

Uses (Inputs & Result)

Get involved with CDIF!

- **Apply to join the CDIF Working Group or Advisory Group! Deadline 30 September.**

- WG generally meets every two weeks. Focus is on reviewing the existing profiles and developing the new profiles.

- AG to meet 2-4 times a year as required. Receives profile documentation to review.

- Application form: https://forms.gle/DhQ4rnnTWW8DzpGUA

- First review of applications will occur in early October.

- Join the CDIF Community list: cdif-community@lists.codata.org

- Register at: https://bit.ly/cdif-community-list

- CDIF Book: https://cross-domain-interoperability-framework.github.io/cdifbook/introduction.html

- CDIF Github: https://github.com/Cross-Domain-Interoperability-Framework

New 'WorldFAIR+', CDIF Implementation Projects

1. **"Data Science Without Borders":** CODATA is a partner in this Wellcome-funded project which builds on the work of WorldFAIR WP07. The project includes CDIF implementation (particularly data interoperability/integration and privacy management) to enable federated analysis across four African **health research** centres (Kenya, Ethiopia, Senegal, Cameroon). Three years. Underway. Africa.

2. **"FAIR Data and Emergencies:** CODATA will coordinate a project funded by ISC applying the WorldFAIR methodology and implementing CDIF components in **emergencies and cascading hazards** case studies on earthquake data (Turkey) and flooding / cholera (Malawi). 18 months. Started 1 September 2024. Africa and Turkey.

3. **"CDIF-4-XAS":** an OSCARS project cascading grant, to implement, test and refine the CDIF, **to prepare X-ray absorption spectroscopy data for interdisciplinary reuse**. Two years. Starts 1 October 2024, once OSCARS processes are completed. Europe (Germany and UK, but with global relevance).

4. **"CLIMATE-ADAPT4EOSC":** CODATA is a member of the consortium for a major four-year project on FAIR data and innovative services for climate adaptation. A central part of the project will be the implementation and further development of the CDIF and related tooling. The project comprises three case studies: **urban heat** (Greece); **oceans / coastal management data** (Portugal); **clay soils / hydrology / built environment / insurance** (France). Four years. Starts 1 Jan 2025 once EC processes are completed. Europe (Greece, Portugal, France).

5. **"FAIR Principles implementation for DDE":** Implementation of FAIR principles, alignment of IUGS CGI standards with CDIF, for cross-domain research topics and data reuse in **geology**. Three years. From Jan 2025. Global. Funding from IUGS.

WorldFAIR+

Vision:

- federation of case studies (existing and new), with parallel funding and supported by a coordinating mechanism with technical expertise.

Approach:

- Use and refine the Case Studies and FIPs workshop approach.
- Further develop the work on describing 'context' and provenance that was explored at Dagstuhl.
- CDIF implementation and refinement.
- Case study-led interoperability frameworks.
- Continue and expand the CDIF Working Group and Advisory Group (underway).

WorldFAIR

Implementation, mechanisms and governance:

- Approach discussed at WorldFAIR GA, with AC and other stakeholders.
- ISC has approved WorldFAIR+ as part of its portfolio of activities: https://bit.ly/ISC-WorldFAIR-PLUS
- Levels of participation: A, case study; B, case study, plus contribution in kind to coordination; C, case study, plus financial contribution to coordination. **Will shortly reissue a public version of this document.**
- Interest from existing and new case studies, particularly from case studies with new projects to be part of the initiative.
- Challenge remains how to resource the coordinating function.

Concrete Progress

- Discussions with various partners and CODATA members.
- Exploring with ISC, UNEP and UNESCO to explore further case studies and implementation pilots.
- Five new projects for CDIF implementation.
- **Keen to discuss potential case studies with ISC members (national members, unions etc).**
- **Keen to discuss any other partnerships, mechanisms and funding!**

Thank you for your attention

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