

# CODATA

Connecting Data and People to Advance Science and  
Address Global Challenges

Simon Hodson, Executive Director, CODATA

# CODATA in a nutshell

- CODATA serves a **membership** that includes **national data committees, scientific academies, International Scientific Unions (ISC membership) and other organisations.**
- CODATA leverages a **network of expertise** and a **skilled secretariat** to address challenges around data policy, interoperability and reuse of data in all areas of science.
- CODATA works with many **partners**, including
  - **ISC and ISC members** (national members, international scientific unions, international programmes)
  - **UN and intergovernmental agencies** (UNESCO, UN Stats, UNEP, UNDRR; OECD; GBIF, BIPM)
  - **Technical and standards organisations, data infrastructures.**
- **CODATA's mission: to connect data and people to advance science and address global challenges.**
- **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.**



# Making Data Work...

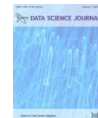


- **Making Data Work for Cross Domain Grand Challenges**
- **World FAIR Project and World FAIR+**
- Recommendations for the Cross-Domain Interoperability Framework (CDIF)
- FAIR Vocabularies with ISUs
- Cross-Domain Case Studies
- Global Open Science Cloud initiative
- Regional Open Science Platforms

# Promoting Data Policy Data science and AI for science



- 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

CODATA mission: <https://codata.org/about-codata/our-mission/>

# Data skills and ECR activities

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)

- Recent podcast series on Open Geo AI: <https://bit.ly/OpenGeoAI-podcast>
- **New co-chairs:** Pragya Chaube, PRIA / UNESCO Chair Coordinator, India; Louis Mapatagane, Walter Sisulu University, South Africa; Christopher ZHU, Imperial College, UK.
- **Sessions at SciDataCon / IDW. Get in touch:** [codataconnect@codata.org](mailto:codataconnect@codata.org)

### ▪ CODATA-RDA Schools of Research Data Science.

- **Various schools, 10th edition in July-August 2025.**
- **Looking to extend offering of Advanced Schools for specific research areas.**

### ▪ CODATA-CNIC Training Workshops: normally annual event (no 2025 edition).

- Training activities with CODATA members and partners: e.g. Ocean Data Management workshop with UNESCO-IOC, Mombasa, Sept 2024; Training workshop on open science infrastructures for disaster management, 14-15 Nov, Ulaanbaatar, Mongolia and online.

## ▪ International list: [https://lists.codata.org/mailman/listinfo/codata-international\\_lists.codata.org](https://lists.codata.org/mailman/listinfo/codata-international_lists.codata.org)

## ▪ Training Opportunities list:

[https://lists.codata.org/mailman/listinfo/data\\_science\\_training\\_lists.codata.org](https://lists.codata.org/mailman/listinfo/data_science_training_lists.codata.org)



# CODATA Research Data Management Terminology

- CODATA was asked to curate the CASRAI RDM terminology when CASRAI ceased functioning.
- Substantial revision process; tighter focus on RDM terms; increased quality of definitions, in line with ISO recommendations.
- CODATA Research Data Management Terminology: <https://bit.ly/CODATA-RDM-Terminology>
- CODATA RDM Terminology as a FAIR vocabulary on ARDC Vocabulary Service <https://vocabs.ardc.edu.au/viewById/685>
  - 'Human-readable' version in Zenodo at <https://doi.org/10.5281/zenodo.10626170>
- **CODATA RDM Terminology Working Group 2025: completing its work.**
- **Updated version will be available in October for IDW / SciDataCon.**



# Promoting 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](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.



## Some CODATA Data Policy Contributions

- **UNESCO-CODATA Data Policy for Times of Crisis (DPTC) Toolkit:** <https://bit.ly/UNESCO-CODATA-DPTC-Toolkit>
- **WorldFAIR Policy Brief (2024):** <https://doi.org/10.5281/zenodo.11242702>
- **UNESCO Recommendation on Open Science (2021)**, co-chaired by Simon Hodson: <https://unesdoc.unesco.org/ark:/48223/pf0000379949>
- **CODATA 20-Year Review of GBIF (2020):** <https://doi.org/10.35035/ctzm-hz97>
- **European Commission Expert Group**, Chaired by Simon Hodson, Turning FAIR into Reality (2018) <https://doi.org/10.2777/1524>
- **Business models for sustainable research data repositories** (for OECD GSF, 2017) <https://doi.org/10.1787/302b12bb-en>
- **Guidelines for the Legal Interoperability of Research Data** (CODATA-RDA IG, 2016) <https://doi.org/10.5281/zenodo.162241>



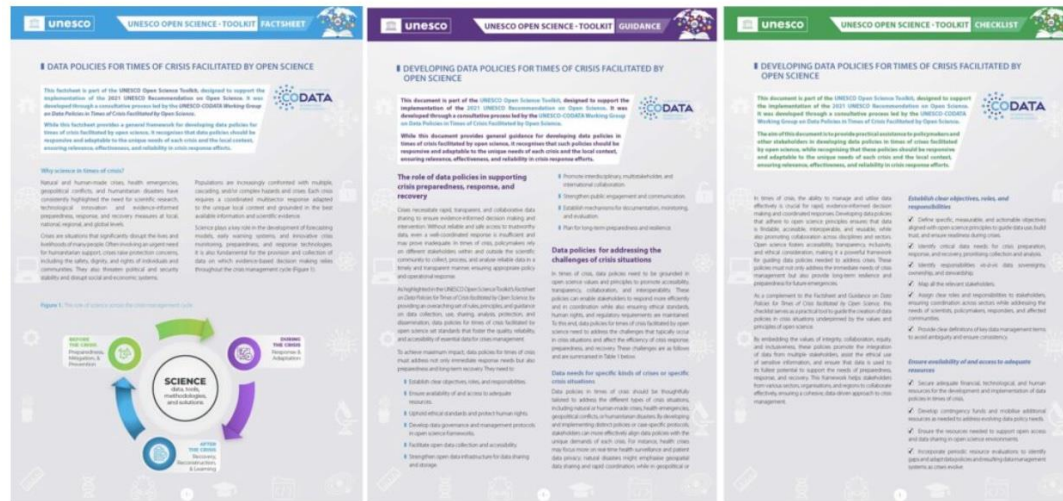
# Data Policy in Times of Crisis

## Joint UNESCO-CODATA Working Group 'Data Policy in Times of Crisis'

- Working Group co-chaired by Ana Persic, UNESCO; Virginia Murray, UKHSA and CODATA; Francis Crawley, chair of CODATA IDPC.
- Prepared a **factsheet**, **guidance** and **checklist** for developing data policies for times of crisis.
- Extensive collaboration and consultation through UNESCO, CODATA, UN Agencies.

## Toolkit now available for implementation!

- Factsheet: <https://unesdoc.unesco.org/ark:/48223/pf0000393829>
- Guidance: <https://unesdoc.unesco.org/ark:/48223/pf0000393830>
- Checklist: <https://unesdoc.unesco.org/ark:/48223/pf0000393831>
- See: <https://bit.ly/UNESCO-CODATA-DPTC-Toolkit>



# CODATA-WorldFAIR Policy Recommendations: Enabling Global FAIR Data

We need to enable Research Infrastructures to support 21st century science.

There is an urgent need for a shift from a 'bibliographic' data stewardship practice to a data engineering practice!

<https://doi.org/10.5281/zenodo.11242702>

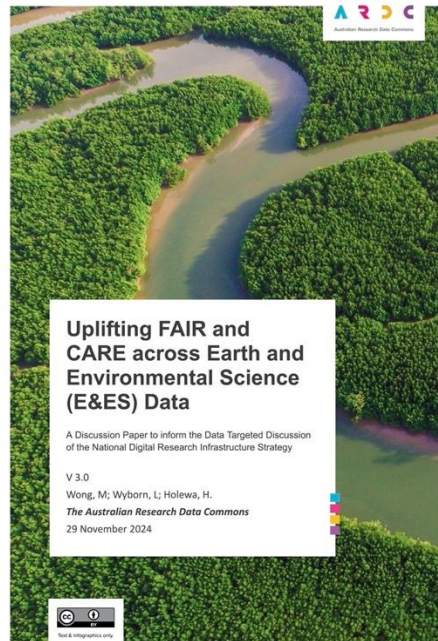
- We need (international) research infrastructures to aggregate and integrate data for key research areas.
- This task is hampered by the persistence of the bibliographic, publication model for thinking about data.
  - Parsons, M.A. and Fox, P.A., 2013. Is Data Publication the Right Metaphor?. Data Science Journal, 12(0), p.WDS32-WDS46.DOI: <https://doi.org/10.2481/dsj.WDS-042>
  - 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 enable Research Infrastructure to furnish integrated **data products** for researchers. This is how we reduce the cost of data wrangling and support cross-domain research.





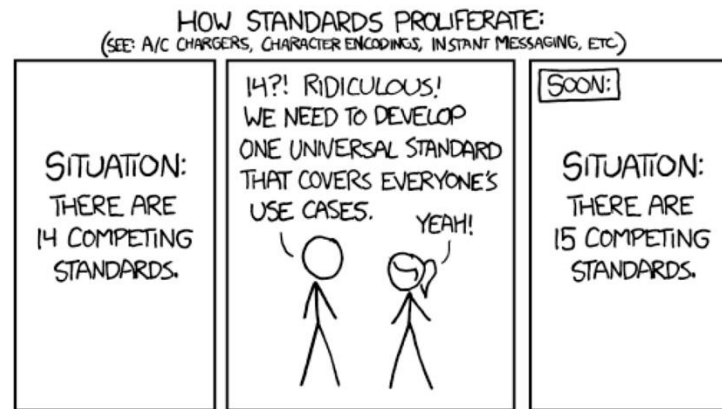
# Uplifting FAIR and CARE: CODATA WorldFAIR Policy and Technical Recommendations promoted by ARDC

- Uplifting FAIR and CARE recommends implementing the WorldFAIR policy and technical recommendations for the Australian Planet Data Commons.
- Endorses the WorldFAIR Policy Recommendations.
- Recommends the implementation and further development of CDIF, the Cross-Domain Interoperability Framework.
- Argues that FAIR and CDIF can assist with the implementation of decisions made by indigenous communities in line with the CARE principles.
- Uplifting FAIR and CARE:  
<https://doi.org/10.5281/zenodo.14241825>

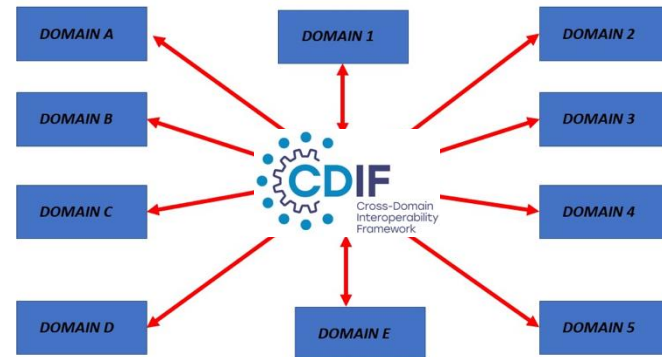


# What is the CDIF (Cross-Domain Interoperability Framework)?

- **Identifies a set of functional requirements for interoperability, particular for steps in data combination, and identifies good practices for each of these requirements.**
- Draws on work with the WorldFAIR case studies and with a number of international initiatives (ODIS, Science on Schema.org, UN Stats KG work, GBIF...)
- **Good web practices:** Significant proportion of CDIF rests on good web practice, domain neutral standards and good practice: disciplines can adopt or map.
- **Use cases:** domain or cross-domain projects or data services that need to combine data for analysis, modelling etc.
- **Directed at implementers:** describes use cases, identifies standards, gives guidance and on how to implement them.
- **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*.**

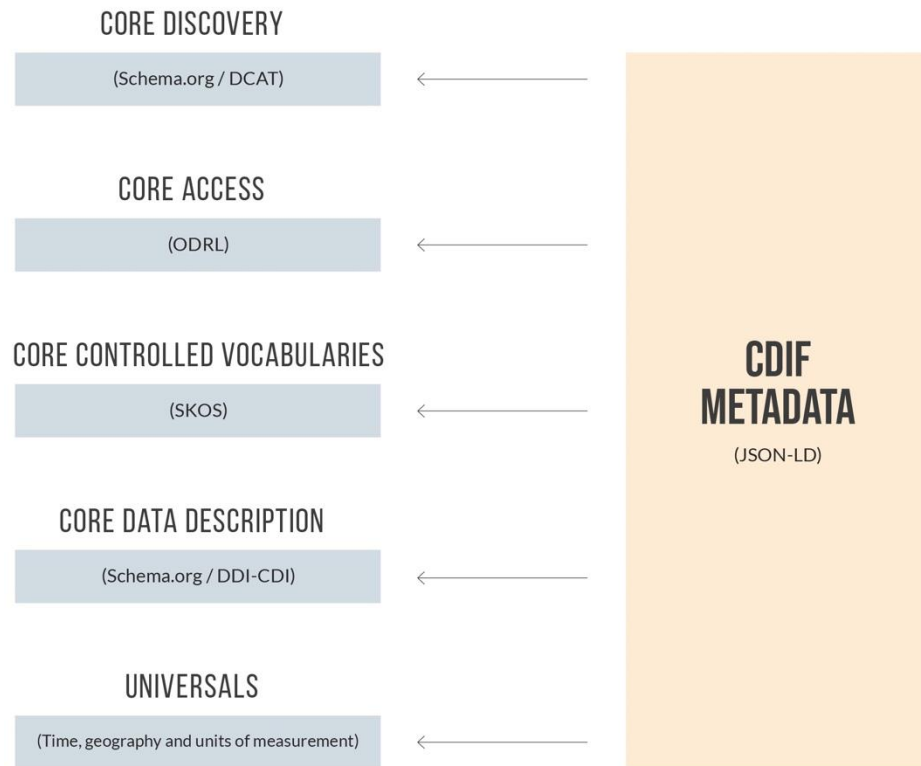


Source: xkcd.com

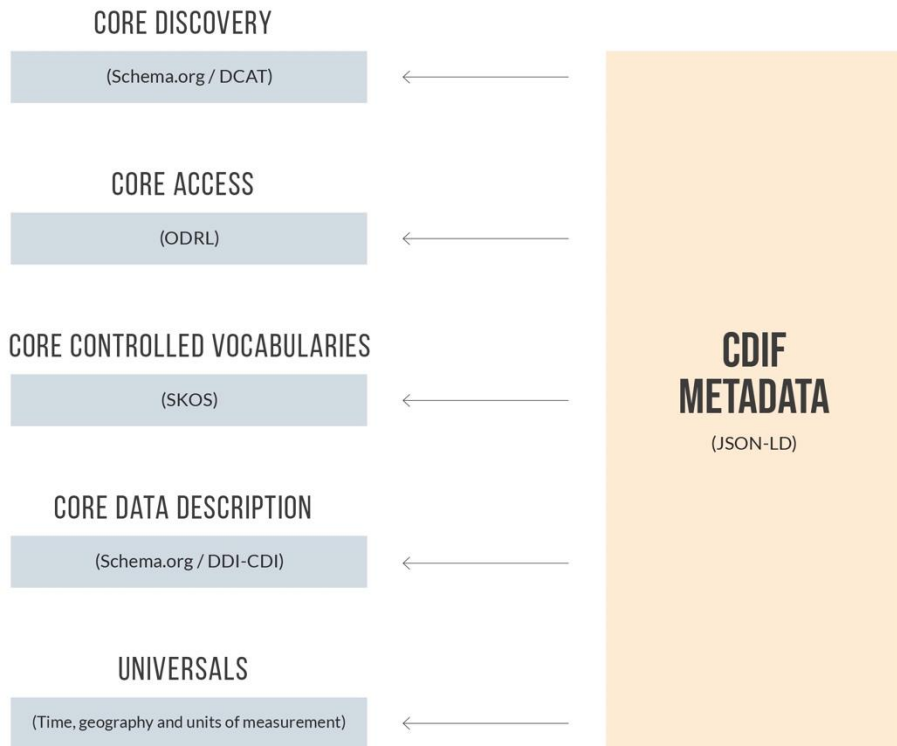


# What is CDIF?

- The Cross Domain Interoperability Framework (CDIF) is a set of practical, implementation-level principles designed to improve data management practices within any community and lower the barriers to cross-domain data reuse. CDIF offers standards and methodologies for achieving different levels of interoperability necessary for reusing data across diverse domains. It is built around five core profiles that address the essential functions for implementing cross-domain FAIR principles.
- CDIF was first released in May 2024 as an output of the WorldFAIR project:  
<https://doi.org/10.5281/zenodo.11236871>
- The point of reference for CDIF and its component profiles is now the CDIF Book: <https://bit.ly/CDIF-Book>
- **Lots of uptake in ongoing and new projects, by Chile Data Observatory, with a number of data infrastructures!**



# CDIF, the Cross-Domain Interoperability Framework

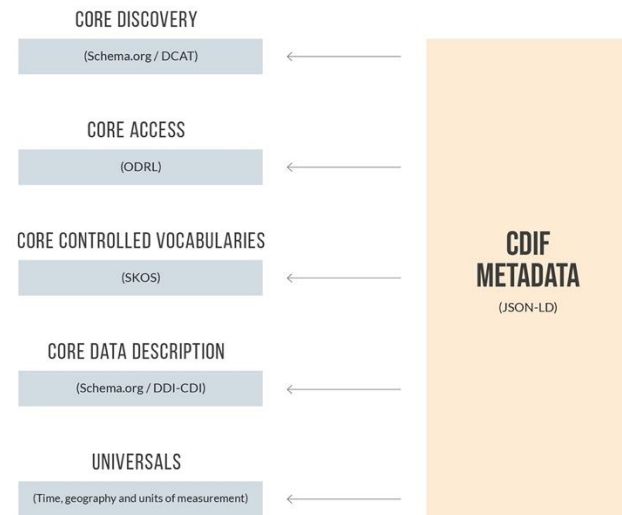


## Feedback received from a colleague at a national data infrastructure

“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...”**”.

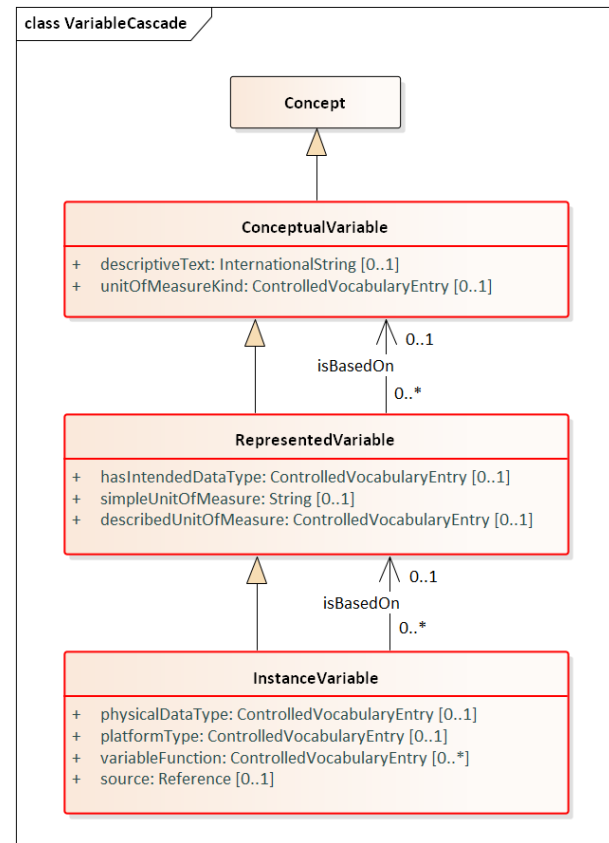
# Discovery Profile

- Discovery profile: <https://bit.ly/cdif-discovery>
  - A [Content model](#) that specifies the information expected to be included in any metadata record, with required, recommended and optional content items.
  - A [JSON-LD serialization](#) for that content using the [Schema.org](#) vocabulary to define the fields in a metadata record, and [an implementation using the DCAT rdf vocabulary](#)
  - [Workflows to publish CDIF metadata](#) so that it can be found and indexed by search providers using standard web technology
- Variable description in the discovery metadata**
  - Name of the variable as it appears in the dataset.
    - Uses schema.org variableMeasured.
  - Text description.
  - propertyID with URI for the represented concept.



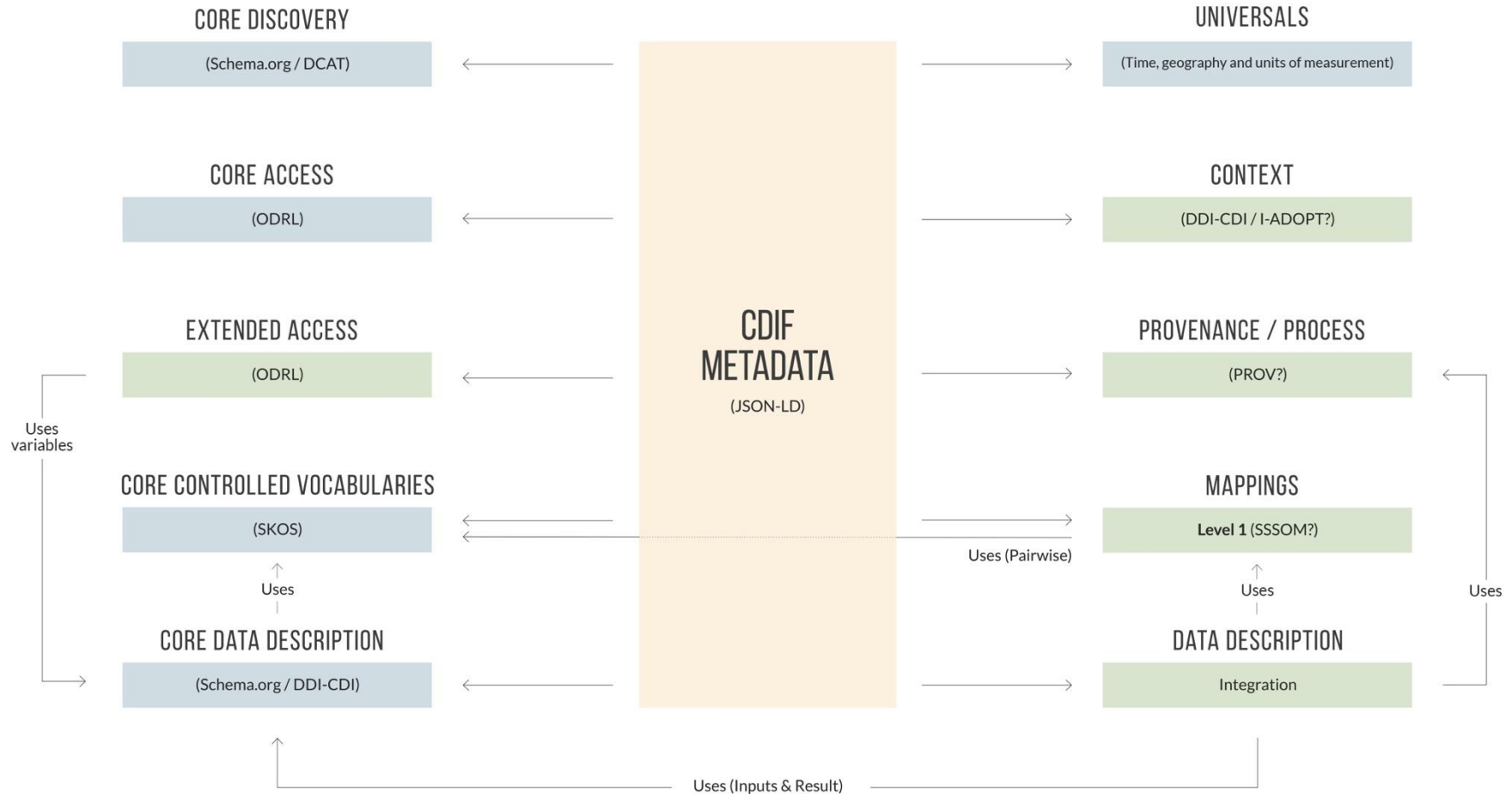
# Description Profile: DDI CDI for Data Structure, Variable Cascade, Provenance...

- 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.
- Now officially released: <https://ddialliance.org/ddi-cdi>





# CDIF WG: ongoing work



# Save the Date!



## INTERNATIONAL DATA WEEK

13-16 October 2025  
Brisbane, Australia



### HOSTED BY



Australian Research Data Commons



Australian Government



NCRIS  
National Research  
Infrastructure for Australia

ARDC is  
enabled  
by NCRIS



COMMITTEE ON DATA

INTERNATIONAL  
SCIENCE COUNCIL



RESEARCH DATA ALLIANCE



WORLD  
DATA SYSTEM

### OUR PARTNERS

BUSINESS EVENTS AUSTRALIA



brisbane  
australia



[internationaldataweek.org/idw-2025/](https://internationaldataweek.org/idw-2025/)

# Thank you for your attention

Simon Hodson, CODATA

[www.codata.org](http://www.codata.org)

[simon@codata.org](mailto:simon@codata.org)

@simonhodson99 ; @CODATAnews