COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

WorldFAIR+: Making data work for cross-domain grand challenges

A CODATA workshop as part of the ISC General Assembly, Muscat, Oman

Save the Date!

INTERNATIONAL DATA WEEK

13-16 October 2025
Brisbane, Australia

bitly

HOSTED BY

ARDC Australian Research Data Commons

Australian Government

NCRIS National Research Infrastructure for Australia

ARDC is enabled by NCRIS

CODATA Committee on Data International Science Council

RDA Research Data Alliance

World Data System

OUR PARTNERS

Business Events Australia

Tourism & Events Queensland

brisbane australia

Brisbane Convention & Exhibition Centre

internationaldataweek.org/idw-2025/

SciDataCon 2025, part of International Data Week: Call open for sessions and papers!

- Propose a session with speakers or a workshop; or an individual paper.
- **Data for Positive Change! https://bit.ly/SDC2025-Call**
  - CAREful Indigenous Data Governance
  - Rigorous, responsible, and reproducible science in the era of FAIR data and AI
  - Open research through Interconnected, Interoperable, and Interdisciplinary Data
  - Empowering the global data community for impact, equity, and inclusion
  - Infrastructures to Support Data-Intensive Research – Local to Global
  - The Transformative Role of Data in Sustainable Development Goals and Disaster Resilience
- SciDataCon persistent themes of data in science, data science, data stewardship.
- Deadline 15 April.

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

WORLD DATA SYSTEM

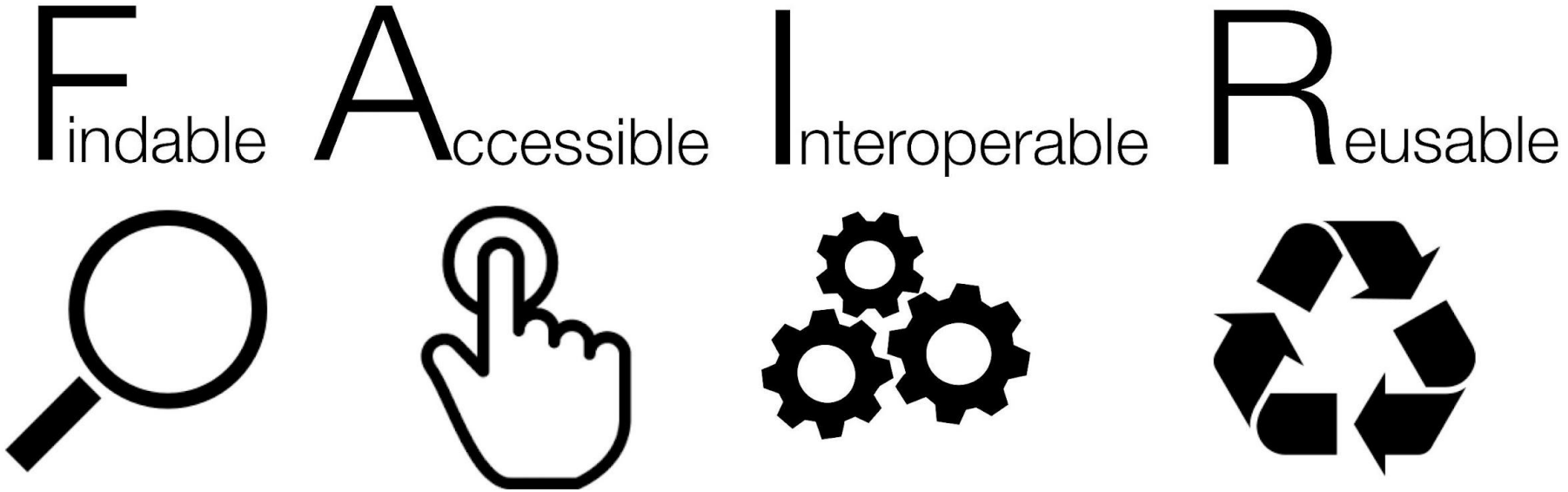
Programme

1. Making data work for cross-domain grand challenges: WorldFAIR+, Simon Hodson, Executive Director, CODATA (25 mins)
2. Discussion: what are the interdisciplinary challenges in your discipline / country? (20 mins)
3. Chemistry and WorldFAIR+ - IUPAC making the central science available to other disciplines, Richard Hartshorn, Professor, School of Physical and Chemical Sciences, University of Canterbury, Christchurch, New Zealand and CODATA Vice-President (10 mins)
4. Combining social science data for interdisciplinary research, Steve McEachern, Director, UK Data Service and CODATA Officer (10 mins)
5. Addressing transdisciplinary approach for tackling urban heat as a science mission, Shaily Gandhi, Senior Post-Doctoral Researcher, Geosocial Artificial Intelligence Research Group, Interdisciplinary Transformational University, Linz, Austria and ISC Fellow (10 mins)
6. WorldFAIR+: How to get involved, Simon Hodson, Executive Director, CODATA (5 mins)
7. Discussion: how to get involved? (15 mins)

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

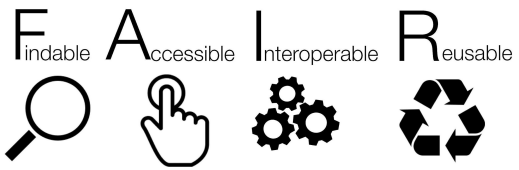
Making data work for cross-domain grand challenges: WorldFAIR+

Simon Hodson, CODATA Executive Director



(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)

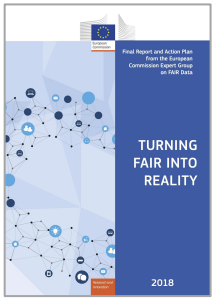
FAIR Principles



- **FAIR: encompasses in an easy communicable acronym, high level principles of good data stewardship**
  - Increases the usability and utility of data, metadata, code.
  - Extremely influential (6519 citations Nature; 15595 citations Google Scholar).
- **Emphasis of the benefits of machine-actionability: network of FAIR data**
  - FAIR principles designed to support the use of data at scale, by machines, harnessing technological potential, better enabling AI.
  - **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'.**

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.



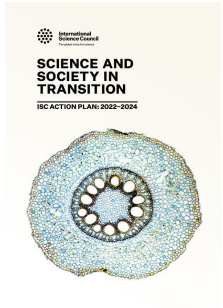
Turning FAIR Into Reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data, 2018, Hodson (chair of working group/lead author), et al., https://doi.org/10.2777/1524

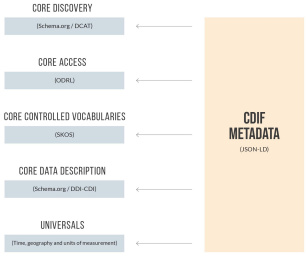
Making Data Work – WorldFAIR – WorldFAIR+

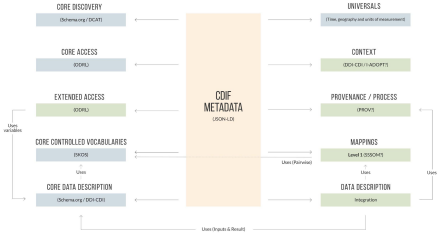
Making Data Work (2018-2022)

WorldFAIR (2022-2024)

WorldFAIR+ (2024+)

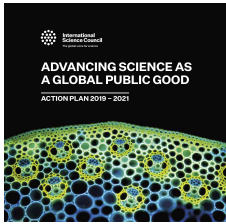






Making Data Work for Cross-Domain Grand Challenges

- The digital and data revolution presents us with huge opportunities and significant challenges.
- **Premise:** The major, pressing global scientific and human challenges 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.
- **Addressing the SDGs or DRR in a science-informed way, requires a fundamentally interdisciplinary approach, the interaction of natural and social sciences… And lots of FAIR data!**
- 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/



DDI®
DATA DOCUMENTATION INITIATIVE

WorldFAIR: Global cooperation on FAIR data policy and practice

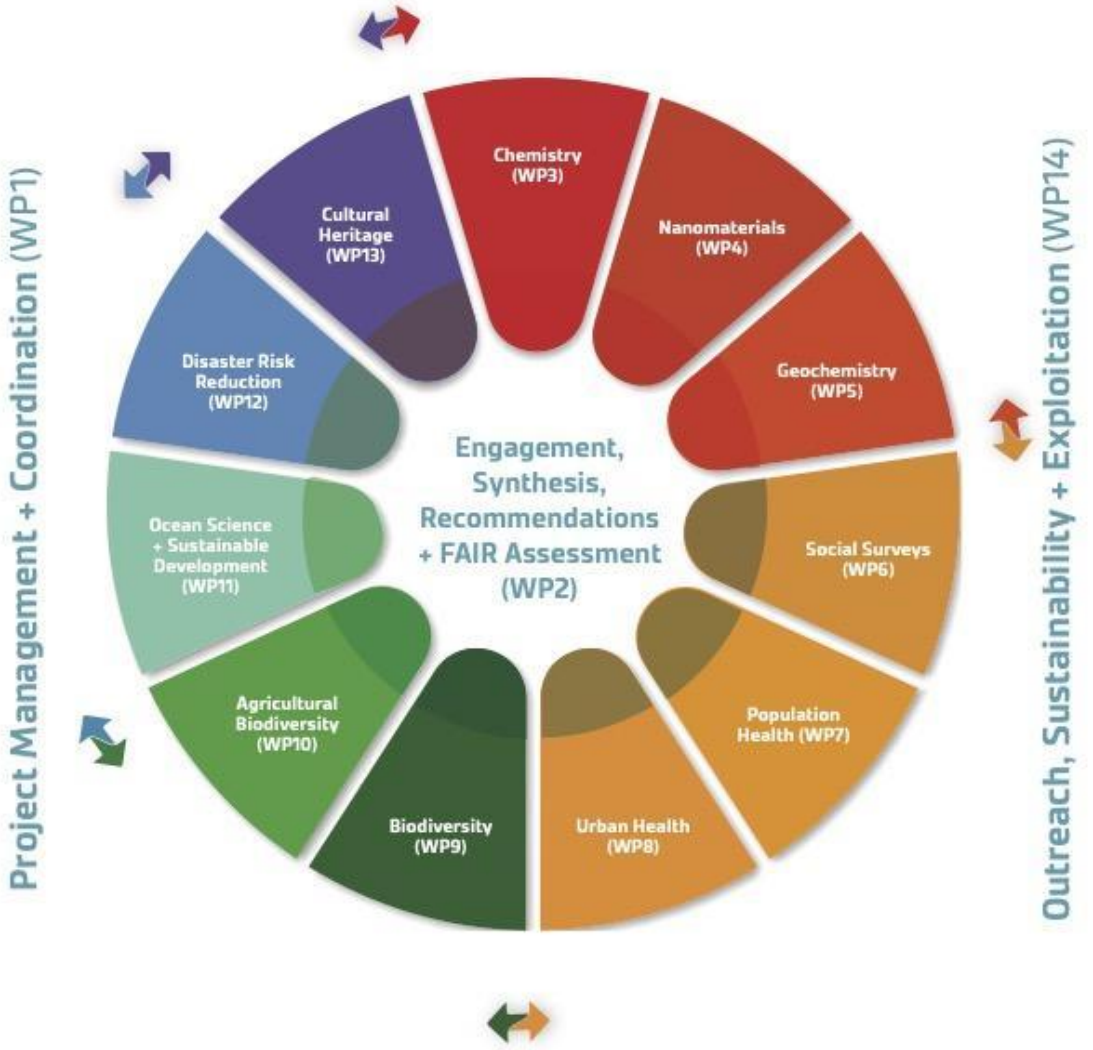
- Advances in FAIR implementation in **cross-domain** scenarios, in 11 specific **disciplines** and **globally**.
- **Global** in approach, because research domains, data and metadata standards and specifications need to be global. Leveraged **CODATA and RDA networks** to achieve this.
- Includes **authoritative 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 Partners

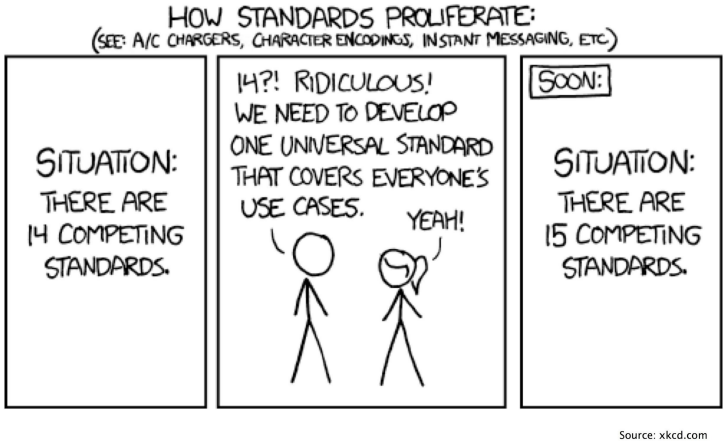
WorldFAIR Case Studies

- **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)

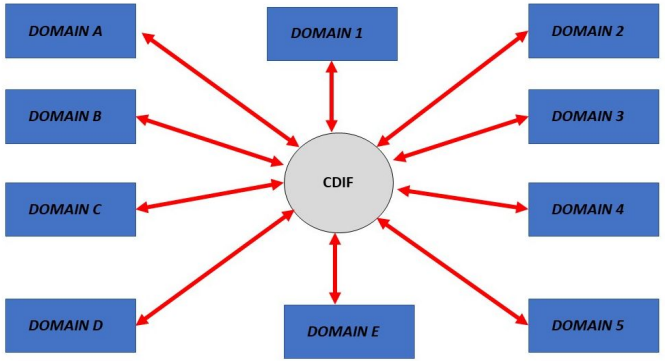


What is the CDIF (Cross-Domain Interoperability Framework)?

- **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…)
- 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, to provide a *lingua franca*.

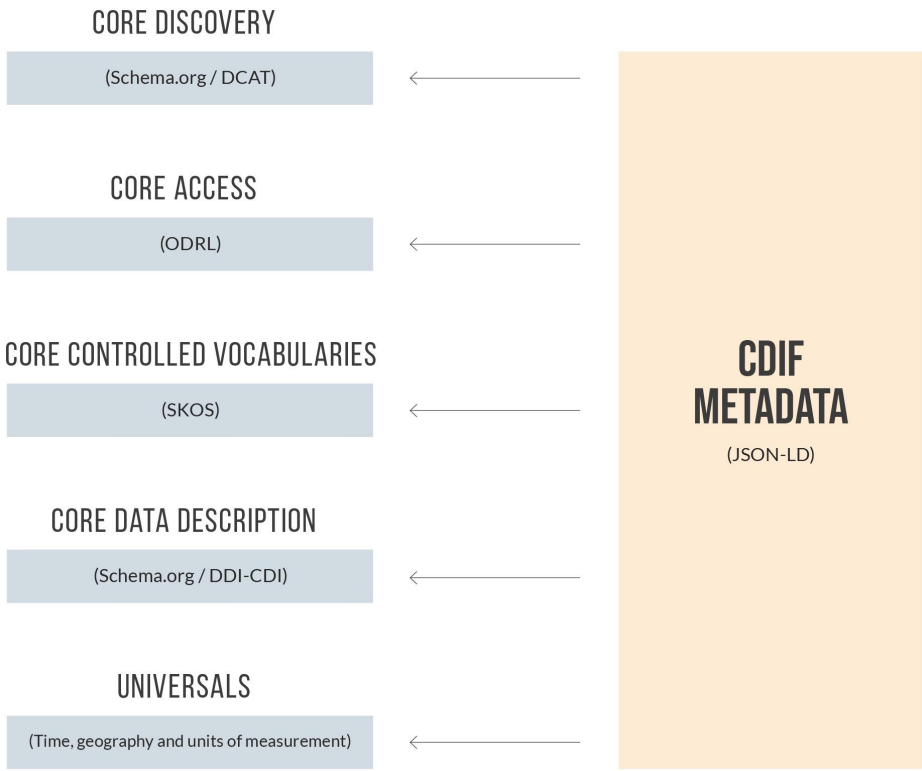


Source: xkcd.com

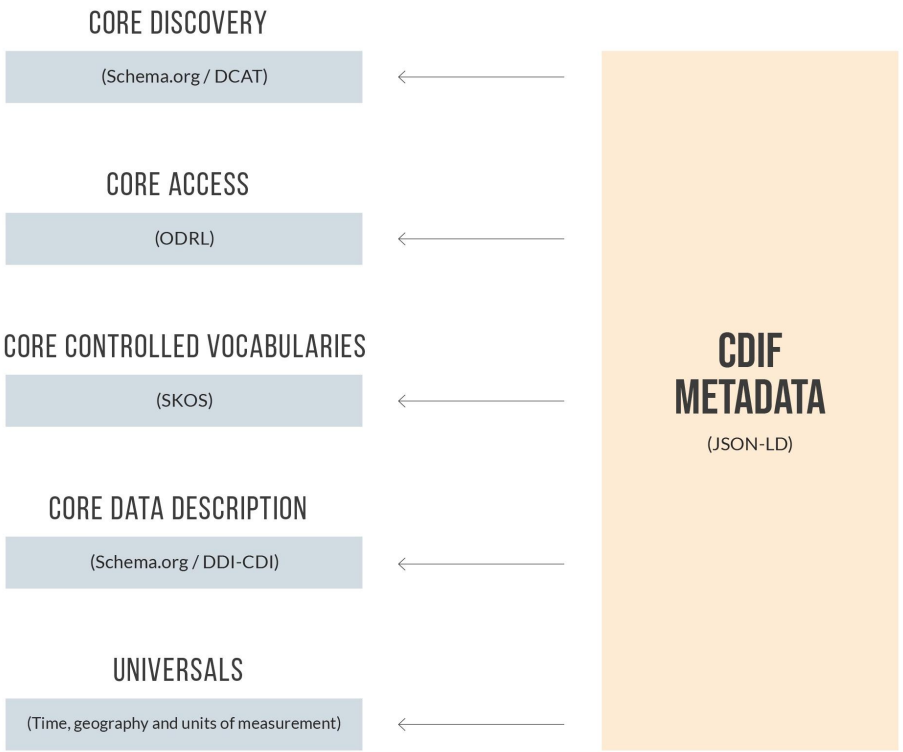


What is CDIF?

- 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.
- **About CDIF: https://bit.ly/CDIF-Intro**
- 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
- Very good feedback from implementers.



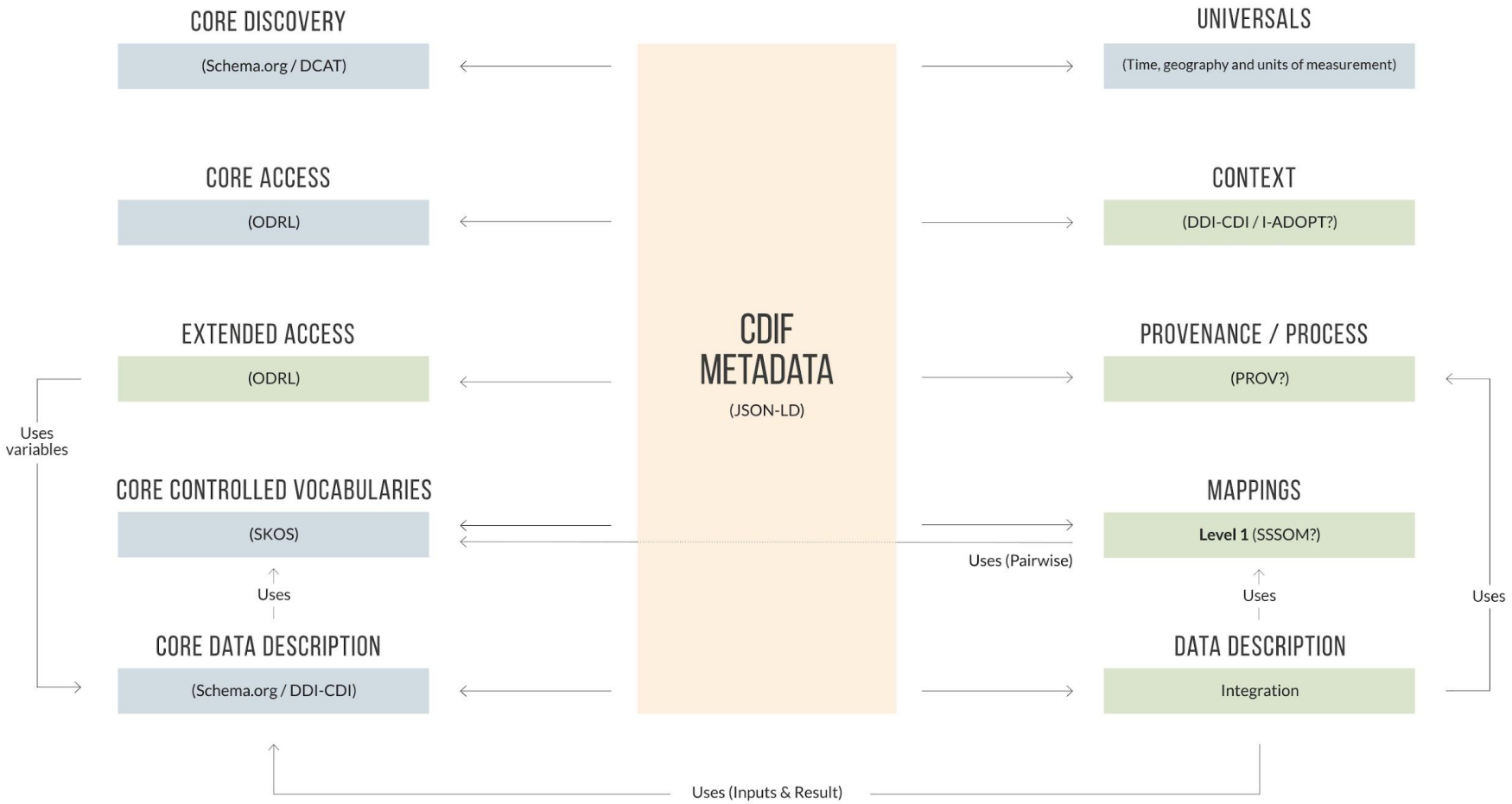
CDIF, the Cross-Domain Interoperability Framework



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…".** 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



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. Three years. Underway. Africa.
    - 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).
    - Combining **population health / statistical data, clinical outcome data, phylogenetic data, environmental data**.
    - Exploring issues of AI assisted cohort analysis and fine-grained privacy management.
2. **“FAIR Data and Emergencies:** CODATA will coordinate a project funded by ISC applying the WorldFAIR methodology and implementing CDIF components for **disaster preparedness, response and recovery**. 18 months. Started 1 September 2024. Türkiye and Malawi.
    - Case study on Turkish earthquake. Scoping underway. Will likely focus on applying CDIF to a **national disaster data platform**.
    - Case study on Malawi floods and cholera (cascading hazards). Scoping underway. Will likely focus on **CDIF-assisted data combination from multiple sources** (MoH, WB, UNICEF, universities, health centres and social media).

WorldFAIR

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

WorldFAIR+, CDIF Implementation Projects

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. Started 1 October 2024. Europe (Germany and UK, but with global relevance).

    - Partners include STFC (UKRI), Cardiff (UK Catalysis Hub), Helmholtz Zentrum Berlin, Helmholtz Metadata Consortium.

    Outputs

    - Overview of standards, vocabularies (and ontologies), data formats and practices within the XAS area (landscape analysis)**: underway, examining standards and workflows.**
    - Semantic description of at least two XAS community standards using a CDIF profile (XAS-CDIF)
    - Implementation plan and method for using XAS-CDIF
    - Prototypes for using XAS-CDIF
    - Recommendations and guidelines for the use of CDIF for XAS data (XAS-CDIF) and in other domains.

WorldFAIR

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

WorldFAIR+, CDIF Implementation Projects

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. Europe (Greece, Portugal, France).
    - Multiple data types and standards.
    - Includes work on organizational and legal interoperability, as well as CDIF implementation.
        - Organizational interoperability framework.
        - Legal interoperability framework.
        - CDIF-based approach to Climate-Adapt data integration.
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.
    - Enabling the alignment of IUGS CGI and other geology standards with CDIF.
    - Envisage a similar methodology to the OSCARS project with XAS data.

WorldFAIR

COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL

WorldFAIR+, CDIF Implementation Projects

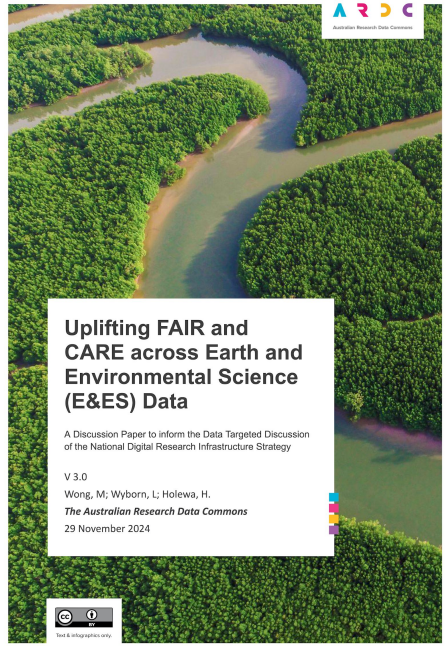
6. **“JUSTSAFE”:** CODATA is a member of the consortium for a three-year, EC-funded project looking at data services for Disaster Risk Reduction and Climate Resilience. Case studies focus on marginalized groups, **extreme heat events, and floods.** Significant co-design and transdisciplinary element. Three years. Starts later in 2025. Europe, multiple countries. Funding from EC.
    - Multiple data types and standards.
    - Includes work on organizational and legal interoperability, as well as CDIF implementation.
7. **“Citizen-Driven Living Labs for Urban Heat Island Mitigation”:** CODATA is part of a consortium for a pilot International Science Council ‘Science Mission’ taking a transdisciplinary approach for health, equity and sustainability to urban heat in India and SE Asia. Multiple data sources (including in situ sensor data, participant and citizen science data. Working with women led NGO, Mahila Housing Trust as well as civic governments and multiple data experts. Eighteen months. Starts later in 2025. Funding from ISC.
    - CDIF core to data integration.
8. **ARDC Planet Commons:** Planning underway for an initiative with ARDC Planet Commons taking the WorldFAIR approach and implementing CDIF.
    - See ‘Uplifting FAIR and CARE across Earth and Environmental Science (E&ES) Data’: https://doi.org/10.5281/zenodo.14241825

WorldFAIR

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

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



COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

Discussion:
What are the interdisciplinary challenges in your discipline / country?

Chemistry and WorldFAIR+: IUPAC making the central science available to other disciplines

Professor Richard Hartshorn
School of Physical and Chemical Sciences
University of Canterbury, Christchurch, New Zealand

FAIR-enabling chemistry data standards/tools

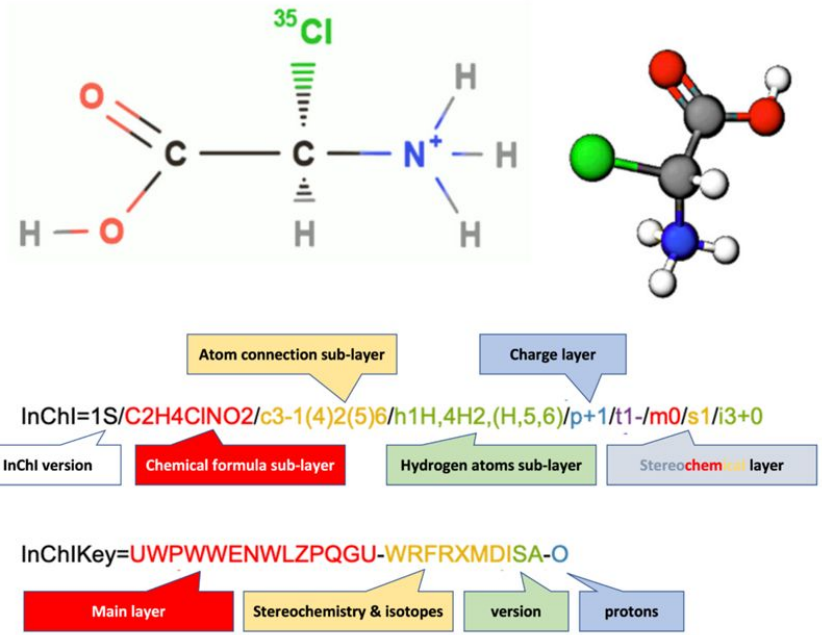
WorldFAIR

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

FAIR-enabling chemistry data standards/tools

WorldFAIR



InChI - the International Chemical Identifier

Adapted from GO FAIR

FAIR-enabling chemistry data standards/tools

WorldFAIR

InChI
InChIKey
Smiles string

Databases

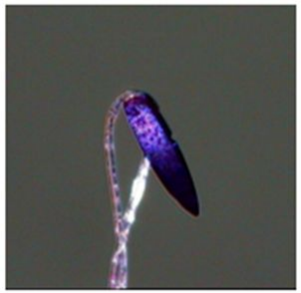
Chemical Structure
Properties/Data

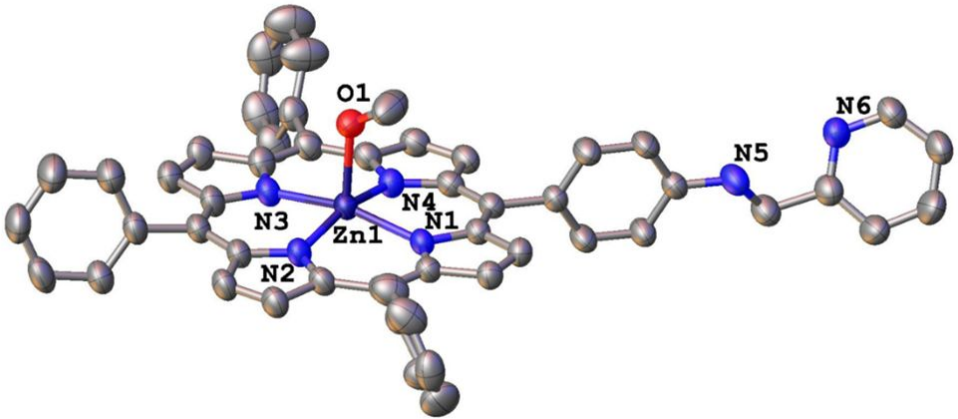
Global Resolver - a web service interface to manage this process

Adapted from GO FAIR

FAIR-enabling chemistry data standards/tools

WorldFAIR





Crystallography leading the way

Adapted from GO FAIR

FAIR-enabling chemistry data standards/tools

WorldFAIR

```
_space_group_crystal_system     'monoclinic'
_space_group_IT_number          14
_space_group_name_H-M_alt       'P 1 21/c 1'
_space_group_name_Hall          '-P 2ybc'
loop_
 _space_group_symop_operation_xyz
 'x, y, z'
 '-x, y+1/2, -z+1/2'
 '-x, -y, -z'
 'x, -y-1/2, z-1/2'

_cell_length_a                  10.3655(3)
_cell_length_b                  10.6196(4)
_cell_length_c                  38.2201(10)
_cell_angle_alpha               90
_cell_angle_beta                97.340(2)
_cell_angle_gamma               90
_cell_volume                    4172.7(2)
_cell_formula_units_Z           4
_cell_measurement_reflns_used   8917
_cell_measurement_temperature   293(2)
_cell_measurement_theta_max     72.7990
_cell_measurement_theta_min     4.2120
_shelx_estimated_absorpt_T_max  0.966
_shelx_estimated_absorpt_T_min  0.525
_exptl_absorpt_coefficient_mu   1.174
_exptl_absorpt_correction_T_max 1.000
_exptl_absorpt_correction_T_min 0.291
_exptl_absorpt_correction_type  gaussian
_exptl_absorpt_process_details
;
CrysAlisPro 1.171.41.93a (Rigaku Oxford Diffraction, 2020)
Numerical absorption correction based on gaussian
integration over
a multifaceted crystal model
Empirical absorption correction using spherical harmonics,
implemented in SCALE3 ABSPACK scaling algorithm.
;
_exptl_absorpt_special_details  ?
_exptl_crystal_colour           'clear dark violet'
_exptl_crystal_colour_lustre    clear
_exptl_crystal_colour_modifier  dark
_exptl_crystal_colour_primary   violet
_exptl_crystal_density_diffrn   1.299
_exptl_crystal_density_meas     ?
_exptl_crystal_density_method   ?
_exptl_crystal_description      needle
_exptl_crystal_F_000            1696
```

The CIF Standard and CheckCIF

Adapted from GO FAIR

Structure Validator

WorldFAIR

Machine readability…

Adapted from GO FAIR

Structure Validator

WorldFAIR

Machine readability…
Interpretation of Standard Formats may differ

Adapted from GO FAIR

Structure Validator

WorldFAIR

Machine readability…
Interpretation of Standard Formats may differ

Does an automatically generated structure match mine?

Adapted from GO FAIR

Structure Validator

WorldFAIR

Machine readability…
Interpretation of Standard Formats may differ

Does an automatically generated structure match mine?

How do we validate chemical structures?

- Correct valences?
- Implicit hydrogen atoms?
- Stereocentres represented appropriately?

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

Envisioning IUPAC at the World's FAIR hub

WorldFAIR

D3.1 FAIR Chemistry Data Guidance

D3.2 FAIR Chemistry Training Cookbook

D3.3 FAIR Chemistry Protocol Services

Acknowledgements

Dr Rajika Munasinghe
InChI Trust

Leah McEwen

A legion of volunteers…

WorldFAIR “Global cooperation on FAIR data policy and practice” is funded by the EC HORIZON-WIDERA-2021-ERA-01-41 Coordination and Support Action under Grant Agreement No. 101058393.

WP6 – Social Surveys

Steve McEachern
Director, UK Data Service, University of Essex
And Visitor, Australian Data Archive, Australian National University

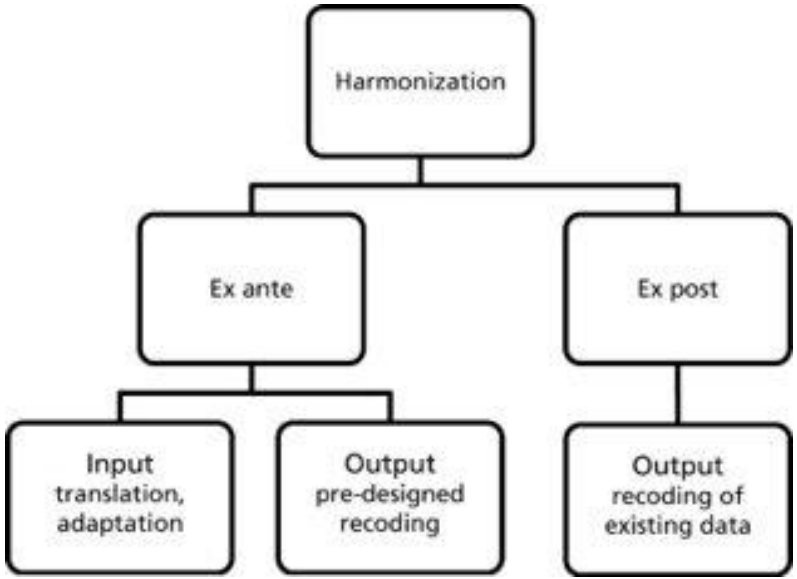
What we are doing

- Comparative study of the data management, harmonization and integration practices of one of the satellite countries – Australia, through the AUSSI-ESS – and the core ESS, an ERIC social science infrastructure.
- The project will examine both administrative procedures, data and metadata management, and technical environments.
- It will then leverage the DDI metadata standards to understand how such multi-national collections could be made increasingly interoperable and reusable through shared procedural and technical development, and
- Establish a set of guidelines and tools for the development of cross-national collections into the future

Data Harmonisation

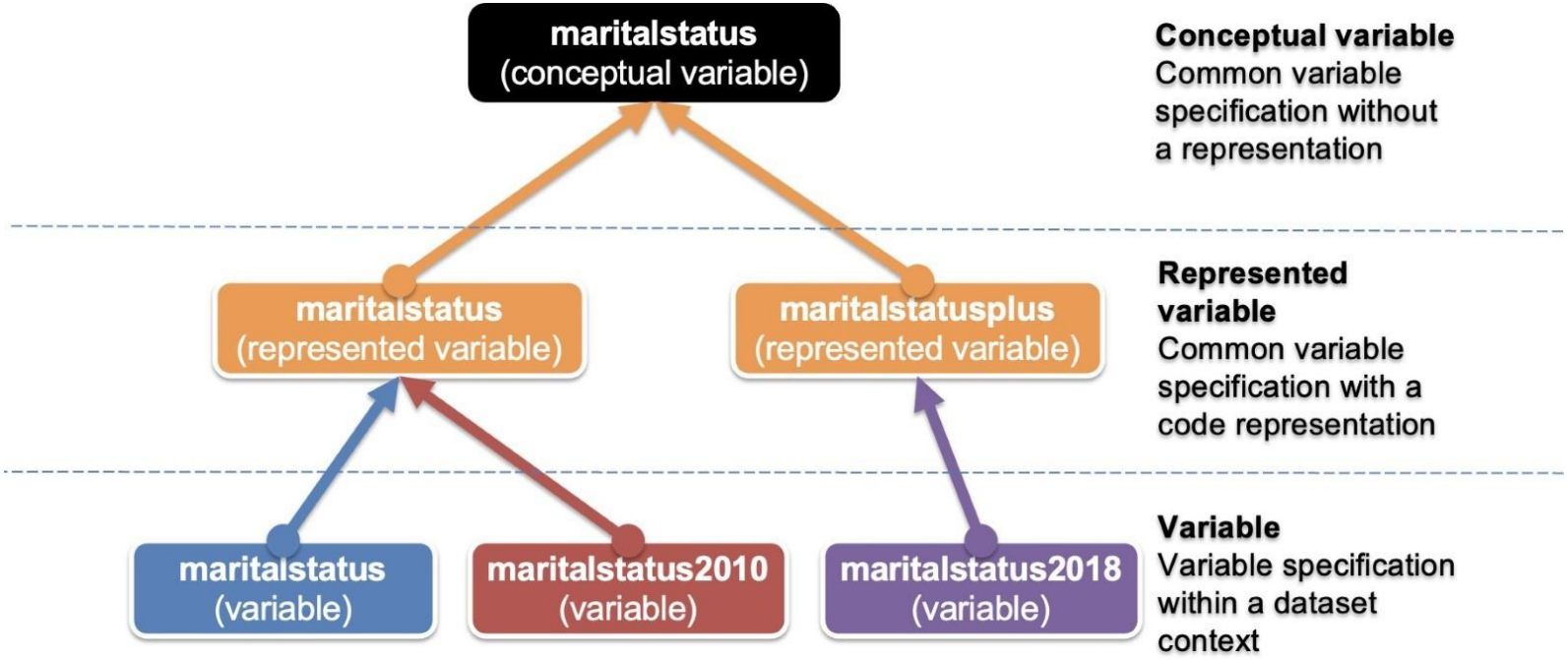
Three broad approaches:

1. Input translation and adaptation: use of the same survey questionnaire in all contexts, along with the same survey methodology (sampling, data collection procedures, etc.)
2. Ex-ante output harmonisation: collection of data using the same survey questions, but with pre-designed re-coding
3. Ex-post output harmonisation: import "existing data and build an integrated database with variables following a common definition" (Wolf et al., 2017).

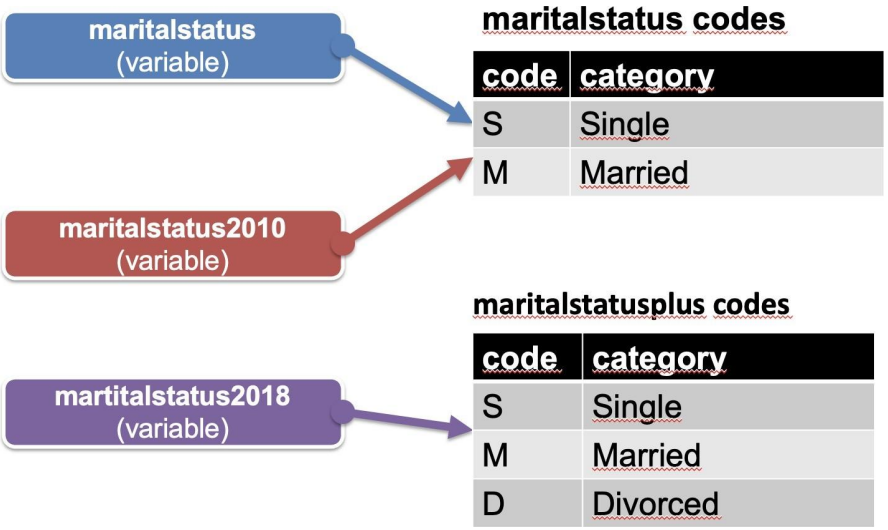


Harmonisation and Metadata standards

- SDMX: Statistical Data and Metadata Exchange
- DDI: Data Documentation Initiative

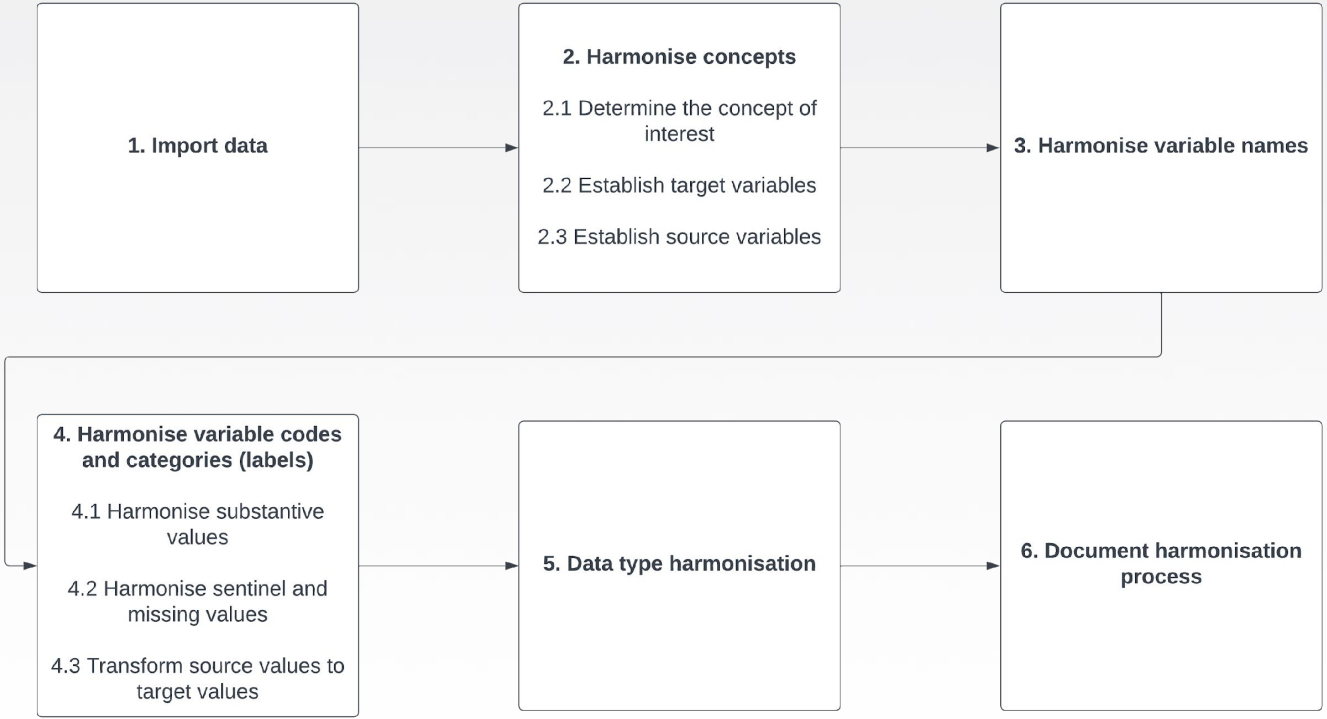


Survey data elements



- The **variable** (including the variable **name** and **label**);
- The **representation** (text, numeric, categories);
- **Properties** of the representations:
  - For categorical representations, the **categories** and **codes**
  - For numerical representations, the **units of measurement**
  - For text representations, the text **properties;**
- The **encoding** of the values in the dataset (e.g. character encoding, data types, etc.)

The cross-cultural survey harmonisation workflow



Case study: International Social Survey Program (https://issp.org)

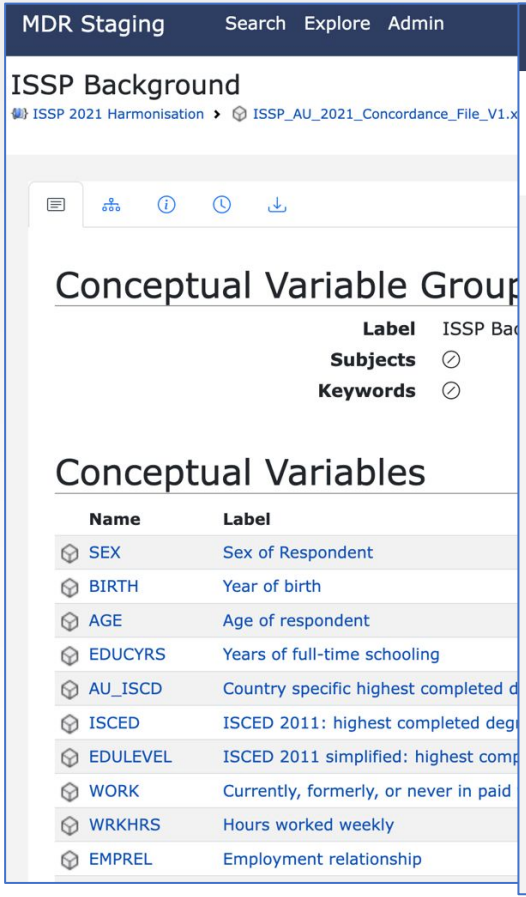
1. Importing

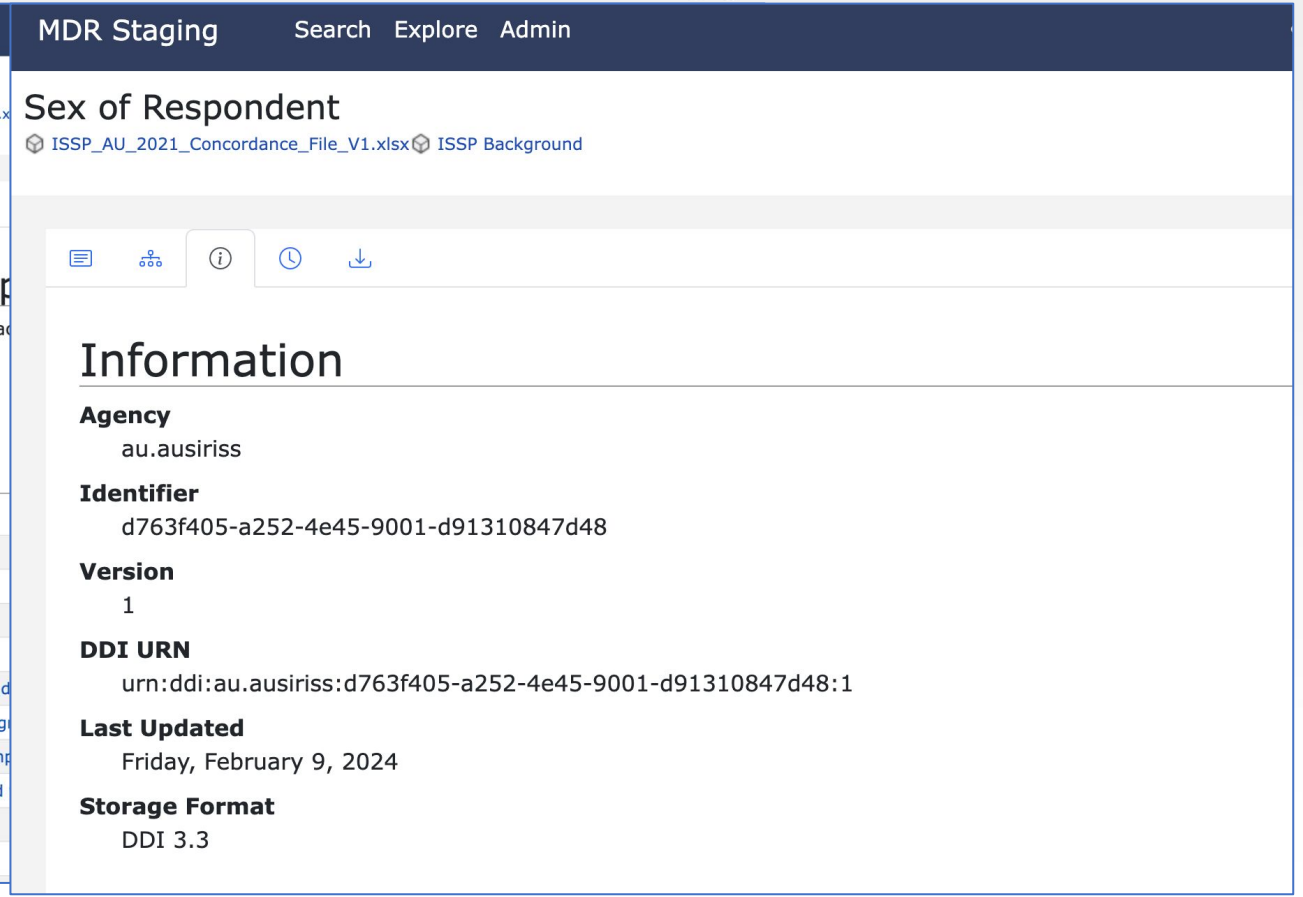
- *Aim: direct data and metadata access, using API calls to registry*
- Result:
  - Data: successful – call to ADA Dataverse repository
  - Metadata: partial
    - Metadata is Findable, Accessible and Reusable (available from the ISSP coordinating repository, online)
    - Excellent documentation at concept, variable and code and category level
    - Not Interoperable: hidden behind a login page, with no API access, and stored inside a syntax (code) file rather than a registry
    - Syntax can be used to generate metadata in an SPSS format file, which can be published to a registry

2. Harmonisation of concepts

- *Aim: Establish the set of target conceptual variable(s) (DDI Conceptual Variable)*
  - Detail the concept that is being defined and measured
  - Locate the agreed target content in project documentation or (ideally) in a variable registry. If this is not available, establish an agreed target variable
  - Establish the source content
- Result:
  - Concepts are fully documented in the ISSP project documentation. But not published in an external registry
  - To generate the target variables for use in the ISSP pilot, the template SPSS script was used to generate an empty SPSS dataset which was then ingested into the Colectica registry.
  - The instance variables in the empty ingested empty dataset can then be used to generate conceptual variables and variable groups.
  - For source variables, in many instances, there is a one-to-one correspondence of source and target variables, or a planned recoding of source content into a target variable *ex ante*.

Pilot Colectica registry





3. Harmonise variable names

- *Aim: execute transformation code to change the variable names in the source dataset(s) to the target names in the destination dataset.*
- Source-to-target concordance file was established in , the generation of the The transformation of source variable names to target variable names is a standard activity in most statistical packages. A script for transforming variable names was generated using a simple Excel concatenation function, and could be automatically generated from a concordance file using standard text functions in the relevant programming language .

4. Harmonise variable codes and categories (labels)

- Aim:
  - *Sub-task 4.1: align the codes and categories associated with the substantive categories in the target variable and source variables*
  - *Sub-task 4.2: align the codes and categories used to manage non-sentinel content (e.g. categories denoting missing information, ineligibility to respond or non-applicability of the question) to target sentinel values*
- Alignment of the codes and categories in a variable can be completed using the creation of a simple matrix, with target codes and categories along the rows, and equivalent codes from the source variables in subsequent columns
- Two core challenges however were identified in the process of creating these harmonisation matrices:
  - a) The variation in processing of missing and sentinel values between data formats required significant iteration to identify and resolve format errors
  - b) The accumulation of small variations in both codes and categories (e.g. Capital letters) created large matrices which were both challenging to read and difficult to automatically process
    - Established a pilot tool to streamline this from Colectica registry content

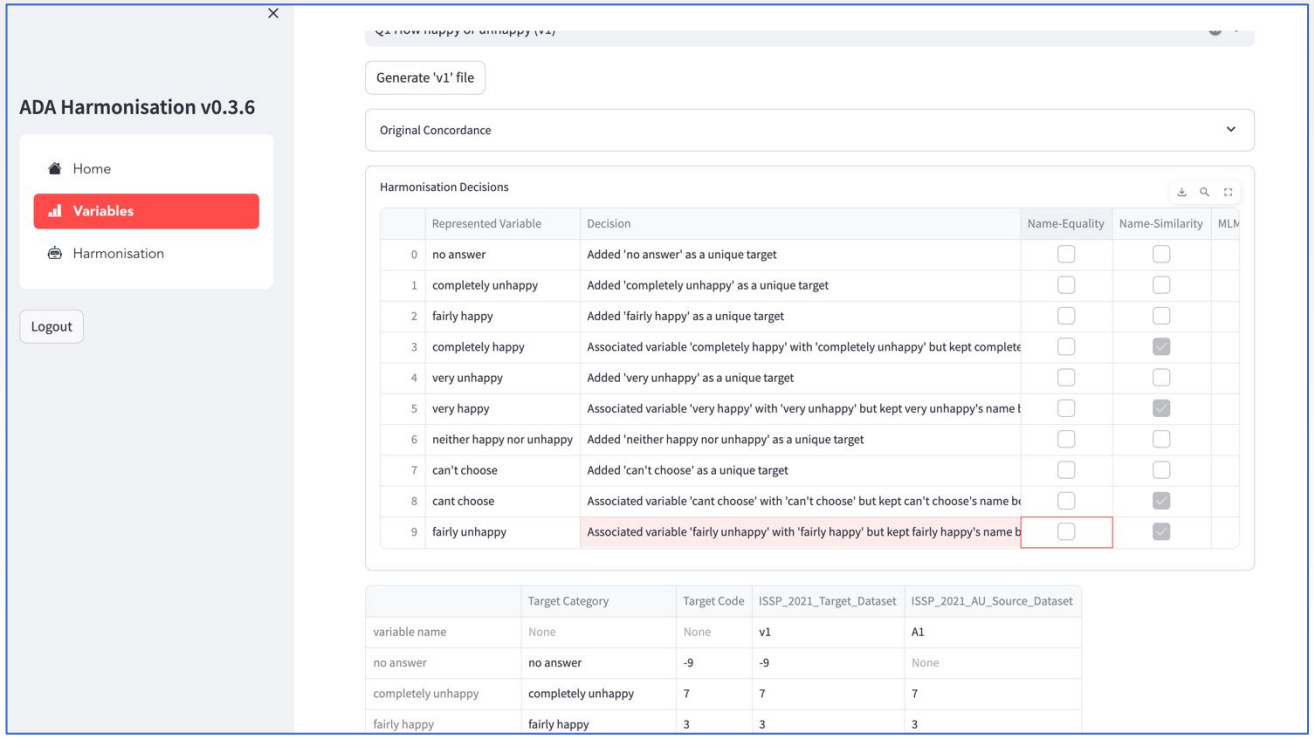
5. Data type harmonisation

- Aim: align the data type of newly generated target variables with the intended data type
- Not tested due to time constraints

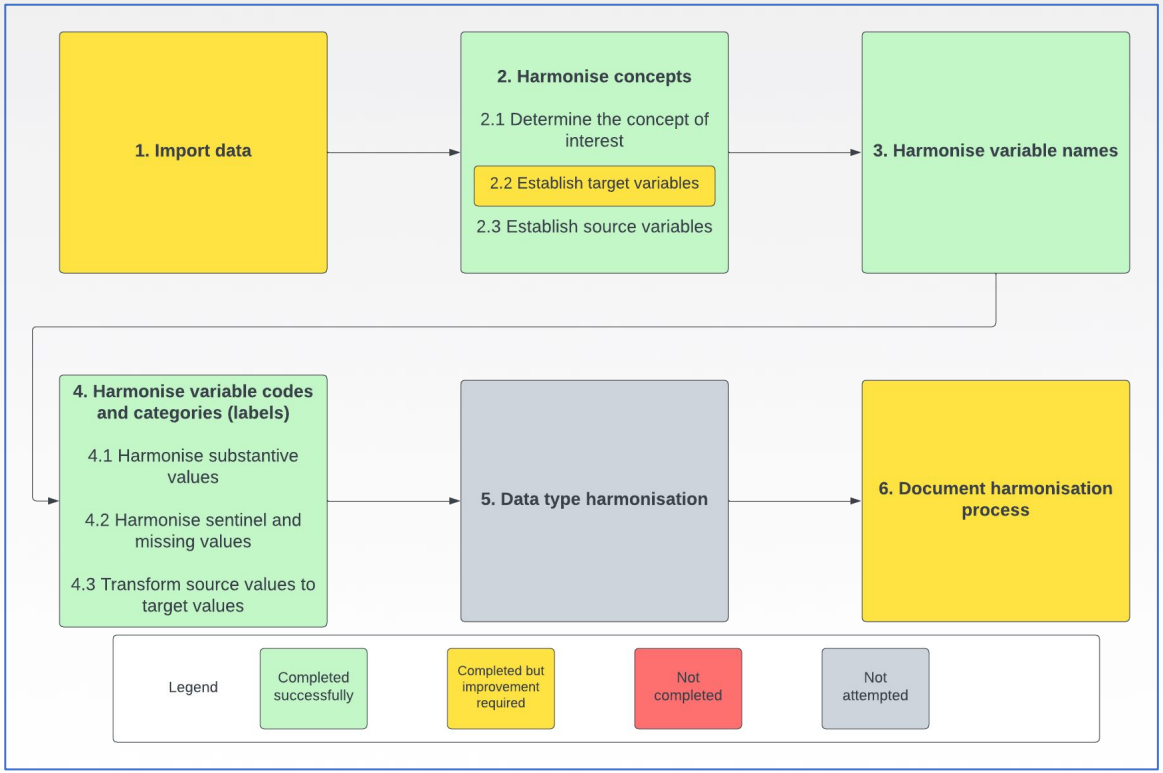
6. Documentation of harmonisation

- *Aim: Capture information on the execution of the harmonisation workflow to capture the process that has been executed.*
- The execution of the CCSH pilot has been able to generate some degree of documentation as a byproduct of executing the process itself.
- The use of software scripts and external registries in the piloting of this harmonisation process resulted in artifacts which reflect the workflow process undertaken.
- This documentation could however be further enhanced through the automated generation of log outputs in the execution of scripts or automation tools.
  - A first attempt to do this has been incorporated in the python Harmonisation tool, which generates a “Harmonisation Decision” table when it generates a machine-learning generated Target Harmonisation Matrix

Example of automated documentation



Summarising the process



Recommendations

- The following recommendations are proposed coming out of this Phase 3 work:

1. Establishment of standardised access controls both to data and metadata registries, to limit the need for less technical users to navigate access control systems
2. Establishment of a code repository for interaction with social science metadata repositories.
3. Establishment of mechanisms for reuse of conceptual variable and other reference metadata across the DDI standards ecosystem. (It was not clear for example how to use or reference a conceptual variable in the Sikt ESS metadata registry within the ADA harmonisation tool)
4. Standardised practices and code libraries for the creation of DDI resource packages for external reuse (to facilitate the reuse in Recommendation 3)

Pilot Science Mission

Citizen-Driven Living Labs for Urban Heat Island Mitigation; a women-led transdisciplinary approach for health, equity and sustainability

Dr. Shaily Gandhi
Senior Post-Doctoral Researcher
Geosocial AI Research Group
Interdisciplinary Transformation University
Austria

MAHILA HOUSING TRUST
Towards Responsible Urban Development

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

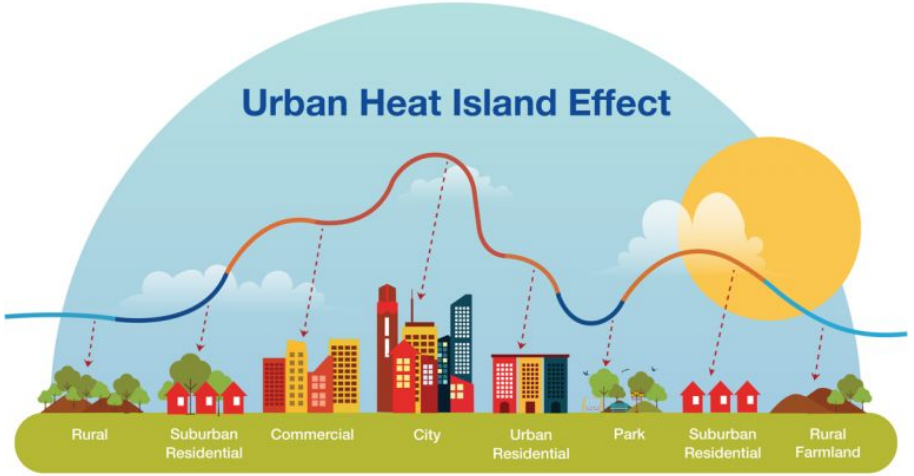
COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

civic data lab

Addressing transdisciplinary approach for tackling urban heat as a science mission

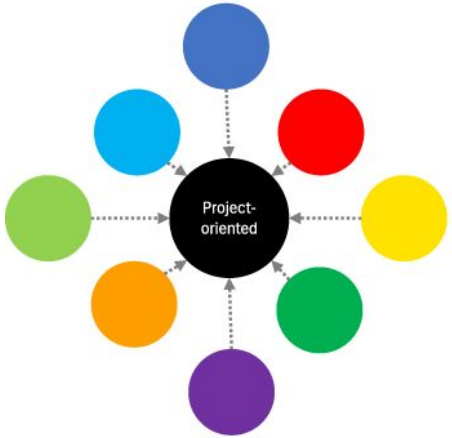
- Urban Heat Islands (UHI)
- Threaten millions across Asia
- Heat action plans
- Non-traditional data sources
- GeoAI technology

Source: https://ecoworks-asia.com/what-are-urban-heat-islands



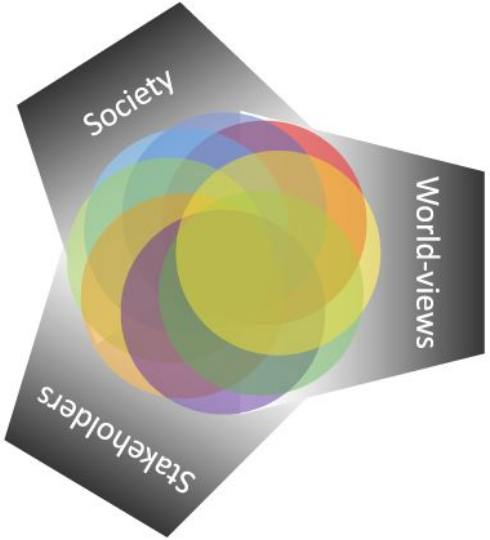
Source: https://gosmartbricks.com/urban-heat-island

Multi- ⇨ Inter- ⇨ Transdisciplinary



Multidisciplinary

Interdisciplinary



Transdisciplinary

Source: Awan, W. Z. (2022). Differentiation between interdisciplinary and transdisciplinary concepts. IMAQ Press. https://imaqpress.com/2678/differentiation-between-interdisciplinary-and-transdisciplinary-concepts/ Google Scholar

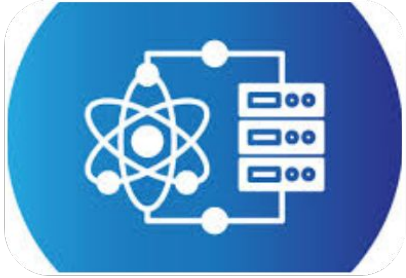
Intersection with multiple SDGs

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

Capacities of the consortium

Transdisciplinary

CDL, MHT, BSC, IT:U



Scientific networks

CODATA, BSC and IT:U

Policy actors

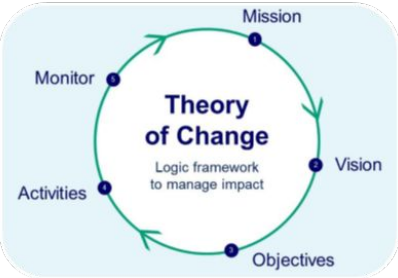
MHT

Computational Infrastructure

BSC

Open Science

CDL, IT:U and BSC



Theories of change

MHT, BSC, IT:U

Stakeholder engagement

MHT , CDL

Community led implementation model:

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

MHT APPROACH

MHT provides the urban poor with

- Knowledge to undertake vulnerability assessments.
- Equips them with available technologies.
- Builds their capacities to devise locally relevant, pro-poor climate resilient solutions.

MHT empowers community-based organizations to

- Implement their own resilience action plans.
- Influence city planning and governance on pro-poor adaptation and resilience action.

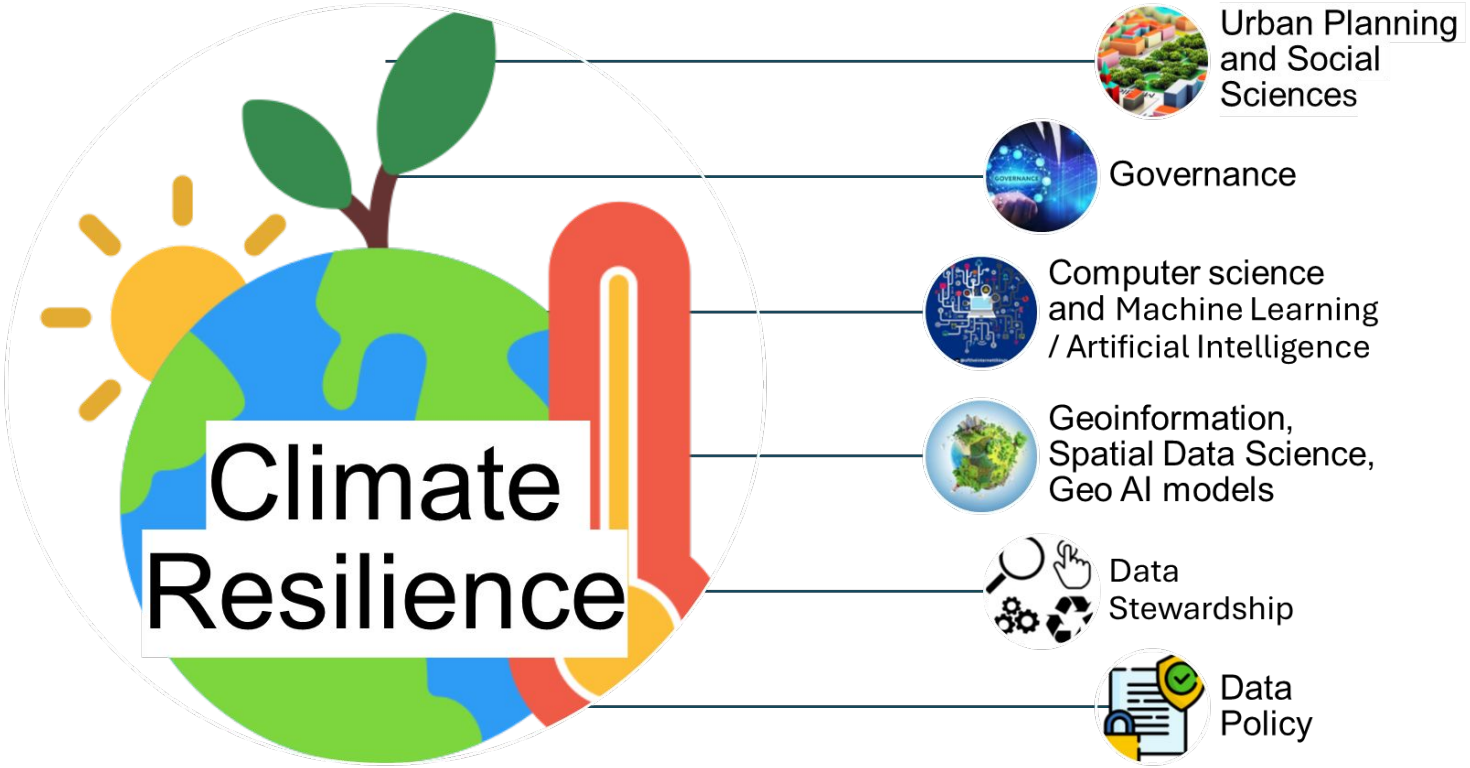
Vision for impact

- Our Pilot Science Mission will seek to:

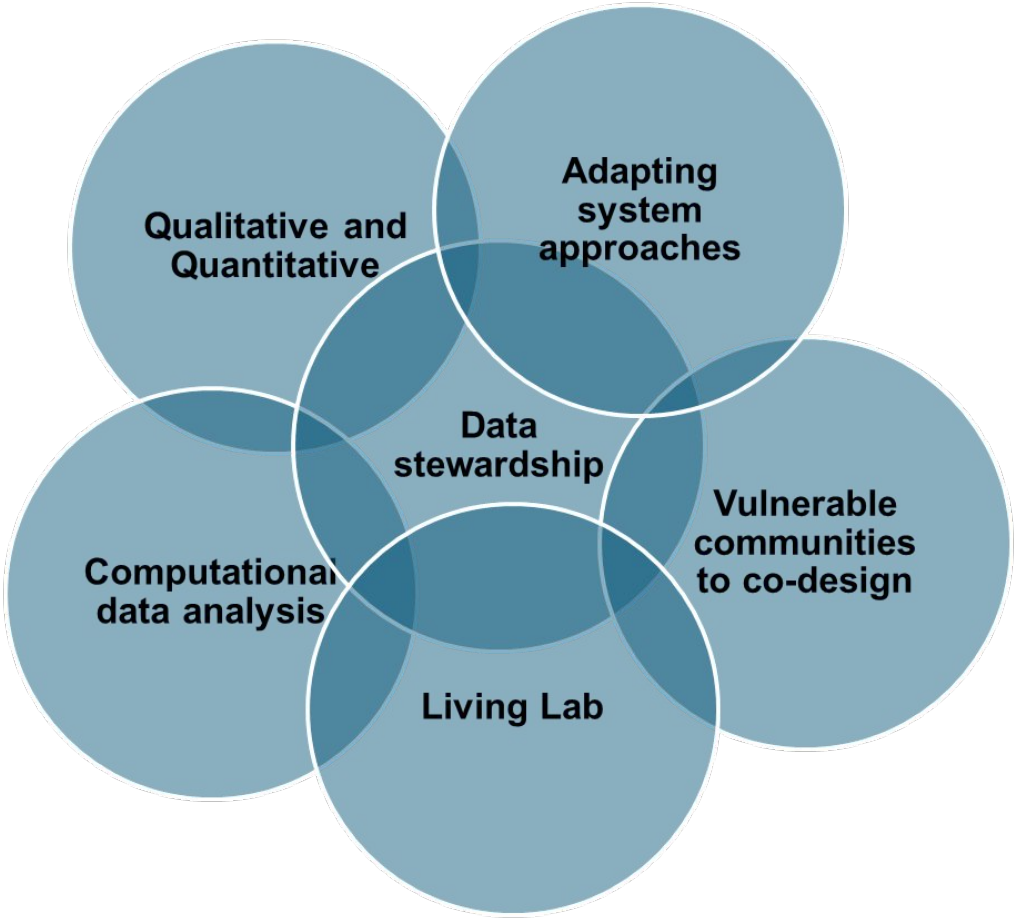
Sustainable Solution Approach

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

Our consortium brings together the following disciplines



Methods


Qualitative and Quantitative
Adapting system approaches
Data stewardship
Computational data analysis
Vulnerable communities to co-design
Living Lab

Multi-modal Datasets

Satellite

Social media

Geospatial

Sensor

Survey and Interview

Citizen's science

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

Next Steps…

WP	Activity	Q1	Q2	Q3	Q4	Q5	Q6
1	Project Initiation and Coordination						
2	Community Engagement and Data Collection						
3	Multi-Modal Data Analysis of the Urban Heat Island Effect						
4	Piloting the mitigation and adaptive strategies						
5	Policy Engagement and Advocacy						
6	Capacity Building and Knowledge Sharing						
7	Monitoring, Evaluation, and Learning						

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

Discussion:
How to get involved with WorldFAIR+?

WorldFAIR+

Vision:

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

Potential Case Studies and partnerships:

- Do you have a potential (project, initiative) case study needing FAIRification, data engineering, metadata uplift?
- Would you be interested in CDIF implementation?
- Keen to discuss potential case studies!

WorldFAIR+ how to get involved?

- ISC has approved WorldFAIR+ as part of its portfolio of activities: https://bit.ly/ISC-WorldFAIR-PLUS
- Vision and approach for WorldFAIR+: https://bit.ly/worldfair-plus
- Will put in place lightweight MoU / LoA for case studies and partner projects.
- **Contact simon@codata.org**

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

WorldFAIR