

Interoperability and Reusability for Cross Domain Data: the next challenge for FAIR

CODATA's mission and operation

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



Data Policies



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

Data Science



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

Data Skills



- CODATA-RDA School of Research Data Science.
- CODATA China, PASTD and other training activities.
- #terms4FAIRskills and FAIRsFAIR Competence Centres.

Data to Improve our World



- **Decadal Programme:** Making Data Work for Cross Domain Grand Challenges
- Promoting Good Data Practices
- Regional Open Science Platforms

F
indable

A
ccessible

I
nteroperable

R
eusable

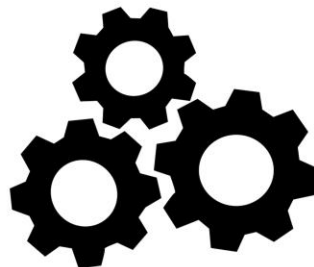


Image CC-BY-SA by [SangyaPundir](#)

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

Making Data Work for Cross-Domain Challenges: the Premise

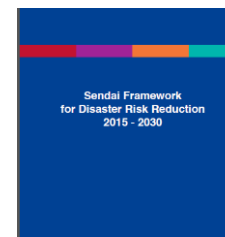
- The major, pressing global scientific and human issues of the 21st century can **ONLY** be addressed through **research that works across disciplines to understand complex systems**, and which uses a **transdisciplinary** approach to turn data into knowledge and then into action.
- The digital and data revolution presents us with huge opportunities and significant challenges.
- Major challenges for many scientific domains – requires work on data specifications, semantics, infrastructures, etc.
 - **80% of effort used on data wrangling; conservative estimate of 10.2 Bn Euro opportunity cost from sub-optimal data stewardship.**
- Open Science and FAIR data provide solutions.
- Considerable global interest in data platforms (EOSC etc).



Data and Science for Global Grand Challenges

- Addressing global grand challenges requires cross-domain collaboration.
- Needs the ability to gather data from many sources, to combine them and extract information from complex and heterogeneous data.
 - Combining data for SDG indicators is challenging.
 - Combining data for the scientific contribution to understanding of SDGs is very challenging!
- **ISC and ISC members (particularly Unions and Associations), and ISC programmes have a role to play.**
- Addressing how to access and combine data (issues of data interoperability) need input from domain experts and definitions agreed by communities.
- Major challenge of fundamental importance to science – **the work of a global decadal programme.**

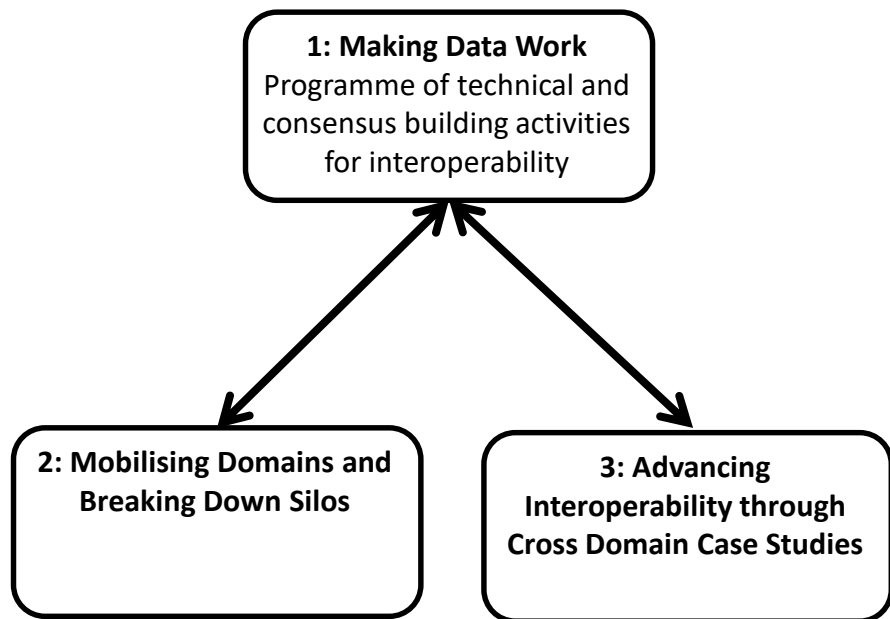
futurearth
research for global sustainability



URBAN HEALTH
AND WELLBEING
A SYSTEMS APPROACH



Making Data Work: programme design



- Programme design comprises three broad work areas.
 1. Consensus and technical solutions for data interoperability (terminologies, ontologies, metadata, machine learning);
 2. Mobilising domains and breaking down silos (working with Unions, Associations and other domain organisations);
 3. Advancing solutions through cross-domain case studies.

Current Pilot Activities

Technically Focused Groups

1. Digital Representation of Units of Measure (DRUM) Task Group:

- Addressing the way units are described, represented, referenced.
- Encouraging alignment with Digital SI and other initiatives.
- Engaging with domains and international Scientific Unions.

2. Good practice for semantic resources and vocabularies

- Dagstuhl Group produced '10 Simple Rule for Making a Vocabulary FAIR'
<https://arxiv.org/abs/2012.02325>
- IUSSP-CODATA WG on FAIR Vocabularies in Population Science
<http://bit.ly/IUSSP-CODATA-FAIR-Vocabs>
- Preparing Working Group on representation, governance and sustainability of vocabularies.

3. Supporting further refinement of the DDI-Cross Domain Integration specification.

- Includes EOSC-funded co-creation project looking at what DDI-CDI can do for EOSC.



Current Pilot Activities

Cross-Domain Case Studies

- 4. Policy Monitoring Indicators** (UN agenda, SDGs, Sendai etc)
 - Exploring data for indicators
 - UNDRR Hazards list
- 5. Infectious Diseases:** projects looking at data integration in infectious disease research, surveillance etc.
 - Current pilot (INSPIRE looking at HIV and COVID).
- 6. Resilient and Healthy Cities:** large group with a number of cities and projects, identifying shared themes and theoretical and technical approaches.
 - Workshop on Data-Knowledge-Action for Quality of Life and Green Spaces in June (partnership with Programme on Urban Health and Well-Being)

Collaboration Initiatives

- 7. Global Open Science Cloud:** CAS-funded project as the first project under the Decadal Programme.
- 8. Collaboration with GO FAIR:** FAIR DOs, FAIR Implementation Profiles (FIPs)
- 9. Data Together Collaborations...**



The Role of DDI-CDI in EOSC: Possible Uses and Applications



The Role of DDI-CDI in EOSC: Possible Uses and Applications

Authors

Arofan Gregory, Simon Hodson, and Joachim Wackerow

Please cite this document as: Gregory, A., Hodson, S., Wackerow, J., 2021, 'The Role of DDI-CDI in EOSC: Possible Uses and Applications', <https://doi.org/10.5281/zenodo.4707263>

Acknowledgment

This work was supported by the EOSC Secretariat.

[EOSCsecretariat.eu](https://www.eosc-secretariat.eu) has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644.

- DDI-CDI (Cross Domain Integration) is designed to interface with other standards and to help interoperability and integration of data between different data types, standards, formats.
- EOSC co-creation project to explore uses and applications and make recommendations for EOSC and the specification.
- **The Role of DDI-CDI in EOSC: Possible Uses and Applications, final report** <https://doi.org/10.5281/zenodo.4707263>
- Report examines the challenge for EOSC: issues of scale and the challenge of cross-domain data reuse.
- Presents examples of how DDI-CDI can be used in data integration and cross domain use cases; describes how DDI-CDI can interact with other standards, notably DCAT; describes how DDI-CDI fits into a FAIR Ecosystem of FAIR Digital Objects; describes how DDI-CDI can be implemented in the Dataverse platform.
- Series of webinars to assist review of specification: <http://bit.ly/DDI-CDI-webinar-series>

DDI-CDI: A new type of standard

- DDI is known for domain standards in the Social, Behavioural, and Economic (SBE) sciences
 - DDI Codebook and DDI Lifecycle
 - Detailed, machine-actionable XML standards for data archiving, production, and management
- DDI Cross-Domain Integration (DDI-CDI) is different
 - Domain-independent – focus is on structural metadata (not semantic)
 - Model-based (UML), not technology-specific
 - Designed to supplement domain standards with metadata for data integration and reuse
 - Designed to be machine-actionable
- DDI-CDI has two functions:
 - Describe a variety of data structures (wide, long, key-value, dimensional) at a granular “datum-centric” level (think “variables” not “data sets”): i.e. structures and variable cascade.
 - Describe the processes at a granular level to describe how datums relate between data sets and structures: i.e. provenance and process.



Slide credit: Arofan Gregory

DDI-CDI Report: Examples and Use Cases



The Role of DDI-CDI in EOSC: Possible Uses and Applications

Authors

Arofan Gregory, Simon Hodson, and Joachim Wackerow

Please cite this document as: Gregory, A., Hodson, S., Wackerow, J., 2021, 'The Role of DDI-CDI in EOSC: Possible Uses and Applications', <https://doi.org/10.5281/zenodo.4707263>

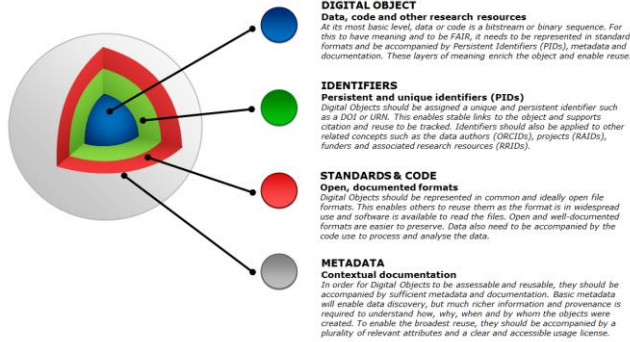
Acknowledgment

This work was supported by the EOSC Secretariat.

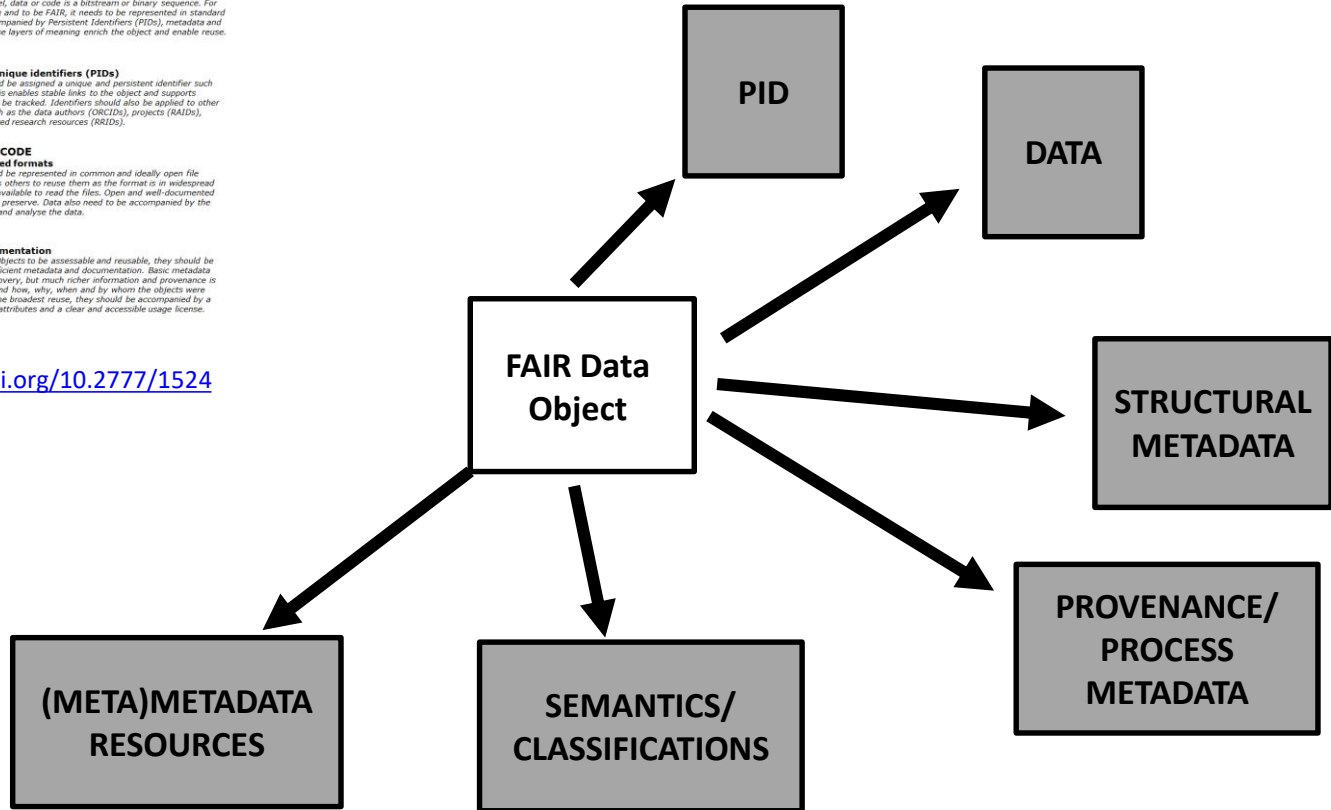
[EOSCsecretariat.eu](https://www.eosc-secretariat.eu) has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-2018-4, Grant Agreement number 831644.

- UK Data Archive: Granular metadata for cross-domain integration (climate data, meter data, survey data)
- Dataverse: Data repository capturing and providing metadata in a lingua franca to support cross-domain integration
- European Social Survey Multilevel Application: Efficiency improvements for survey data integrated with national and regional data from other domains
- The ALPHA Network and INSPIRE: Integration of data from clinical systems and questionnaires highlight differences in domains which require context/provenance information
- DDI-CDI and the FAIR Ecosystem (FAIR Digital Objects and FAIR Implementation Profiles)
- DDI-CDI and the EOSC Interoperability Framework, and the FAIRsFAIR Proposal on Integration of Metadata Catalogues

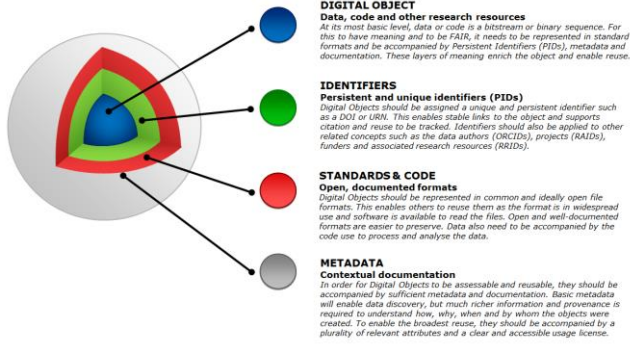
FAIR Digital Objects: a way of thinking about the information needed for reusability / machine actionability



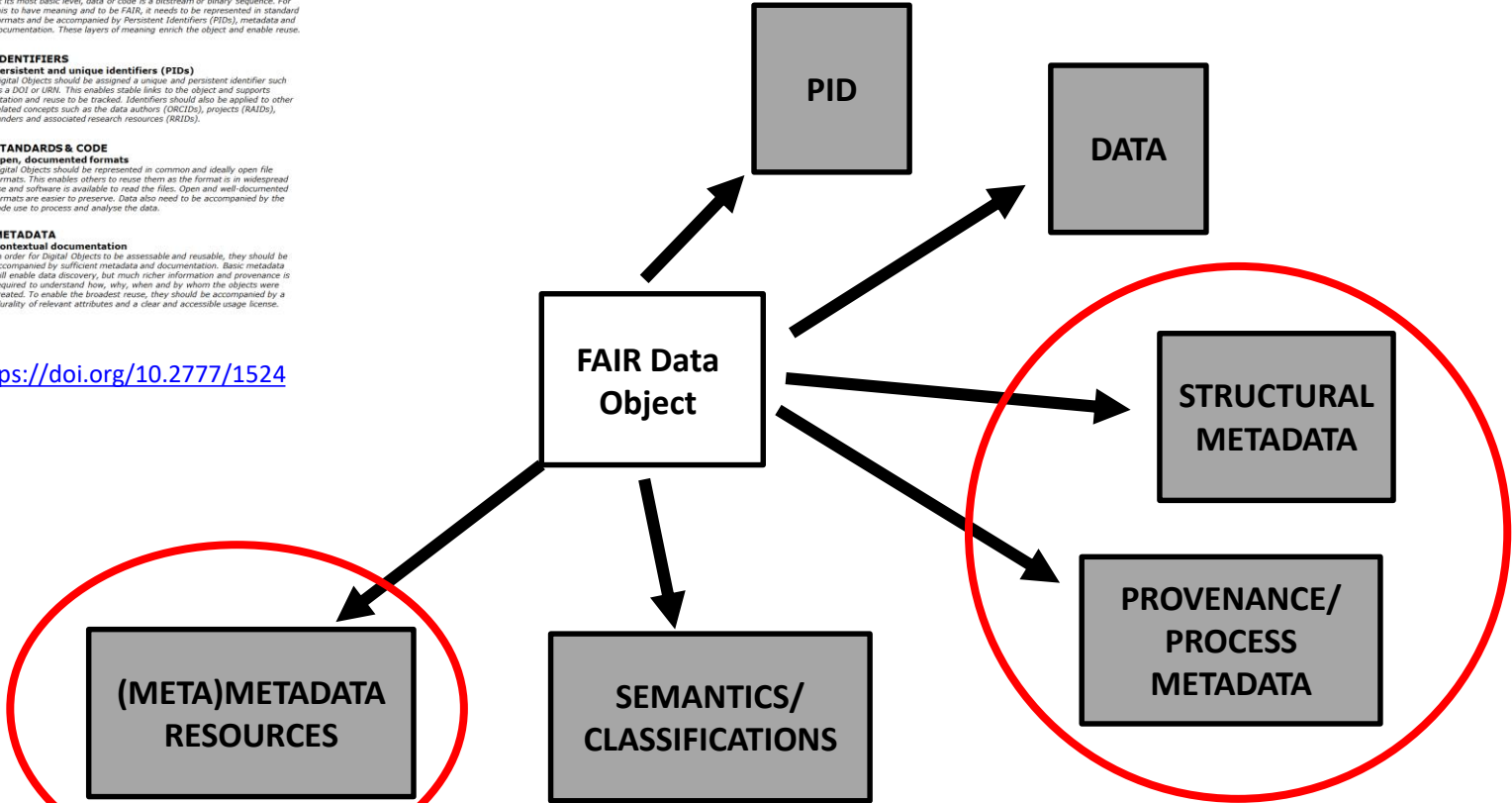
Turning FAIR into Reality <https://doi.org/10.2777/1524>



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



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



EOSC Interoperability Framework (1)

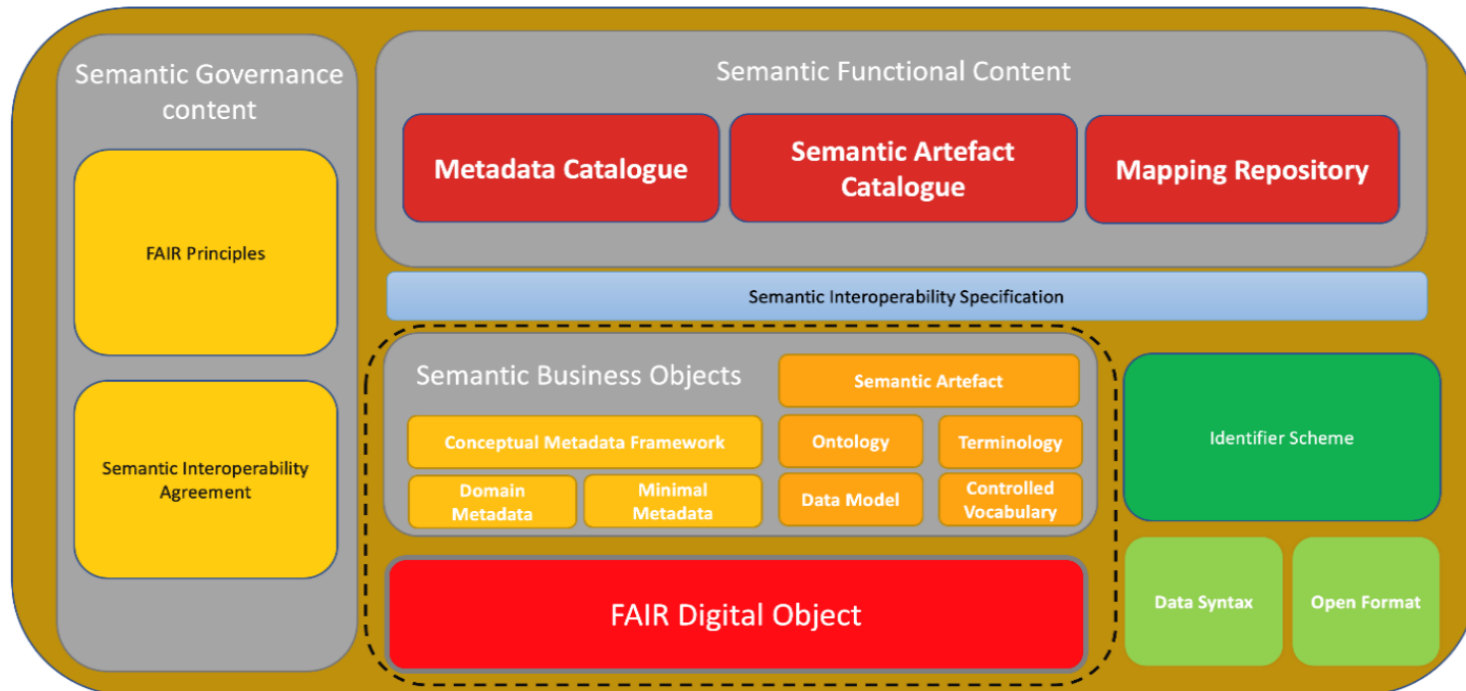


Figure 10. EOSC-IF Semantic view - Objects.

EOSC Interoperability Framework (2)

Example Business Objects Metadata

What collection of digital objects does this service provide?

DCAT, DataCite etc

Minimal Metadata

← **DCAT**

What (semantic) structure are used to describe data to be exchanged?

ISO11179, Prov-O, GSIM,
DDI-CDI etc

Conceptual Metadata FW

← **DDI-CDI**

What digital objects do the service provide on a more granular level?

DDI, HL7FHIR etc

Domain Metadata

← **DDI domain standards**

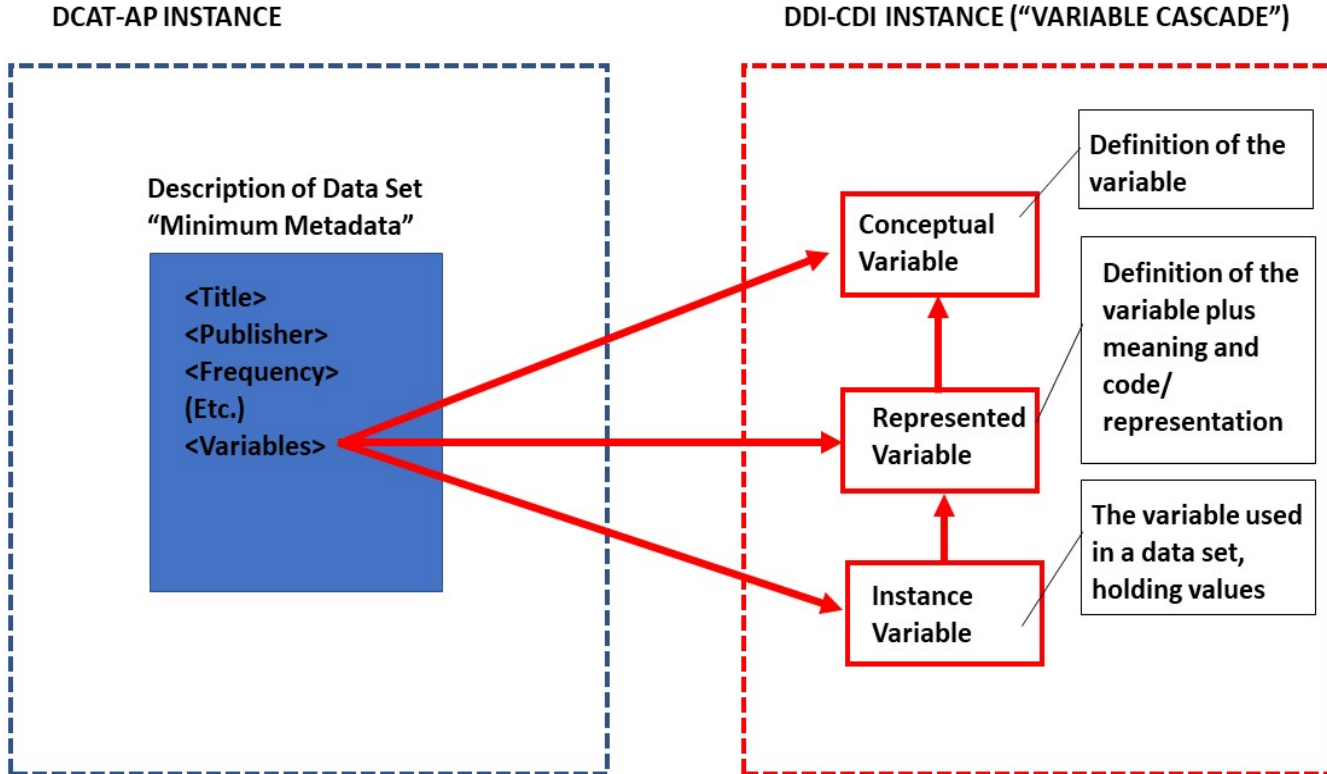
What is the identifier metadata scheme?

DOI, ORCID etc

Identifier Scheme

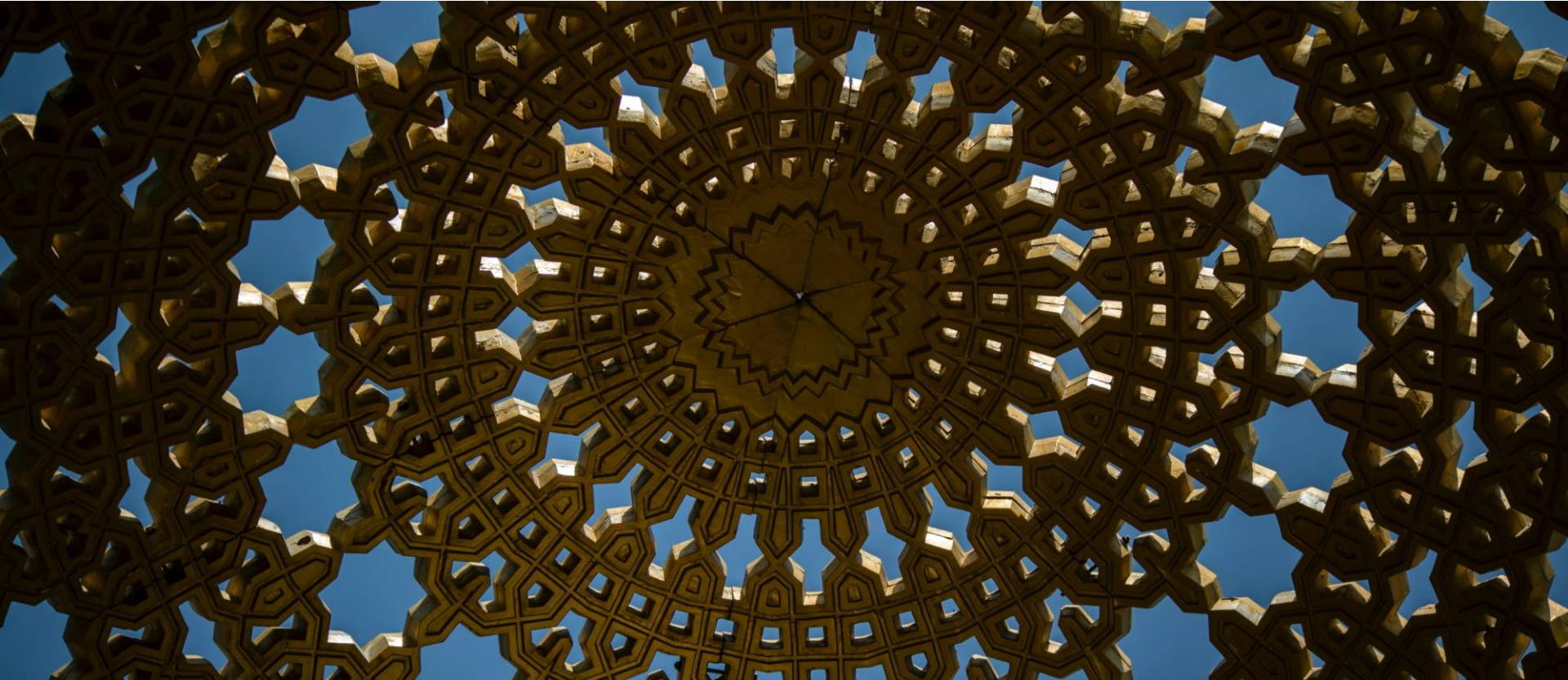
How do these different types of metadata connect?

Example: (Envisioned) DCAT-AP and DDI-CDI



Making Data Work for Cross-Domain Challenges

- Aim to launch the Decadal Programme at the ISC GA and associated events 'Global Knowledge Forum' in **Fully Virtual**, 10-14 October 2021: <https://council.science/about-us/governance/general-assembly/muscatassembly>



INTERNATIONAL DATA WEEK 2021

Data to Improve our World

8-11
NOVEMBER
2021

SEOUL,
REPUBLIC OF
KOREA



Convened by



Call for presentations/papers will be released on 1 June, deadline 31 July <https://www.scidatacon.org/IDW2021/>

Follow CODATA!



- CODATA Website: <http://www.codata.org/>
- CODATA Blog: <http://codata.org/blog/>
- CODATA International News and Discussion List: <http://bit.ly/CODATA-International-List>
- CODATA Data Science and Data Stewardship Careers List: http://bit.ly/CODATA_Careers_List
- CODATA on Twitter: @CODATANews and @simonhodson99
- Facebook: <https://www.facebook.com/codata.org/>
- Insta: <https://www.instagram.com/codatainternational/>

Thank you for your attention

Simon Hodson, CODATA

www.codata.org

simon@codata.org

@simonhodson99 ; @CODATAnews