Building Foundation for a World of Open Data and Open Science
ISC, Institutional and Partner Members

- **ISC Members:** International Scientific Unions, Associations and other bodies related to ISC.
  - [https://codata.org/membership/international-science-council-bodies/](https://codata.org/membership/international-science-council-bodies/)
  - Improved engagement with Unions and Associations: DRUM, DDI-CDI and Decadal Programme.

- **Institutional Members:** Commercial and NFP organisations.
  - Members and Benefits Documents: [https://codata.org/membership/institutional-members/](https://codata.org/membership/institutional-members/)

- **Partner Organisations:** Scientific and data organisations with which it is important to partner (e.g. Data Together Organisations, DDI Alliance, IIASA).
  - Members and Template MoUs: [https://codata.org/membership/partner-organisations/](https://codata.org/membership/partner-organisations/)
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.
CODATA National Committees

- **Engage**: point of contact with CODATA;
- **Influence**: contribute to CODATA strategy;
- **Coordinate**: forum by which national stakeholders may advance data agenda in step with international developments;
- **Collaborate**: propose Task Groups, host or participate in international workshop series, engage with Early Career Data Professionals Group;
- **Partner**: undertake activities with other National Committees, bilaterally or in groups.
- **Data Together**: where possible and appropriate engage with the other Data Together Organisations
- **New Member**: New Zealand; other discussions underway with potential national members.
- Guidance for National Committees: [https://codata.org/membership/national-members/](https://codata.org/membership/national-members/)
CODATA Task Groups

- Advanced mathematical tools for data-driven **applied systems analysis**: led to joint IIASA-CODATA Working Group.
- Applying Data Integration and Data Science Tools toward Research of Urban Life and **Smart Cities**: supporting decadal programme.
- Digital Representation of Units of Measure (DRUM)
- Improving Data Access and Reusability (IDAR-TG): particularly for data at risk...
- **Agriculture Data**, Knowledge for Learning and Innovation: training and analysis platform, Kenya, East Africa
- **Citizen Science for the SDGs** – Aligning Citizen Science outcomes to the UN Sustainable Development Goals: article, fieldwork by two interns.
- **FAIR Data for Disaster Risk Reduction**: white papers, policy briefs and webinars.
- Preservation of and Access to Scientific and Technical Data in/for/with Developing Countries (PASTD): training and knowledge bases for LMICs

For a larger image, please visit [https://tinyurl.com/DisasterLossData](https://tinyurl.com/DisasterLossData)

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

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

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

- 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).
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**.
**FAIR Guiding Principles**

**To be Findable:**
- F1. (meta)data are assigned a globally unique and persistent identifier
- F2. data are described with rich metadata (defined by R1 below)
- F3. metadata clearly and explicitly include the identifier of the data it describes
- F4. (meta)data are registered or indexed in a searchable resource

**To be Accessible:**
- A1. (meta)data are retrievable by their identifier using a standardized communications protocol
  - A1.1 the protocol is open, free, and universally implementable
  - A1.2 the protocol allows for an authentication and authorization procedure, where necessary
- A2. metadata are accessible, even when the data are no longer available

**To be Interoperable:**
- I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (meta)data use vocabularies that follow FAIR principles
- I3. (meta)data include qualified references to other (meta)data

**To be Reusable:**
- R1. (meta)data are richly described with a plurality of accurate and relevant attributes
  - R1.1. (meta)data are released with a clear and accessible data usage license
  - R1.2. (meta)data are associated with detailed provenance
  - R1.3. (meta)data meet domain-relevant community standards

(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](http://dx.doi.org/10.1038/sdata.2016.18))
▪ **Findable**: have sufficiently rich metadata and a unique and persistent identifier, to enable discovery.

▪ **Accessible**: retrievable by humans and machines through a standard protocol; authentication and authorization where necessary.
  - Allows programmatic access for analysis.

▪ **Interoperable**: metadata use a ‘formal, accessible, shared, and broadly applicable language for knowledge representation’.
  - The descriptions of variables etc follow a shared specification and are commensurable.

▪ **Reusable**: metadata provide rich and accurate information; clear usage license; detailed provenance.
  - Both humans and their analytical tools know what can be done with the data (license) and can assess its provenance.

---

European Commission Expert Group, Chaired by Simon Hodson, Turning FAIR into Reality (2018)

[https://doi.org/10.2777/1524](https://doi.org/10.2777/1524)
Making Data Work: programme design

1: Making Data Work
Programme of technical and consensus building activities for interoperability

2: Mobilising Domains and Breaking Down Silos

3: Advancing Interoperability through Cross Domain Case Studies

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

- Current case studies in: resilient cities, disaster risk reduction and infectious diseases. More planned and invited!
- Working with domain and cross-domain areas, semantic solutions and machine learning.
Initial Pilot Activities

Initial Working Groups / Activities

1. Digital Representation of Units of Measure (TG is a key contribution to the decadal programme)
2. Semantic Interoperability and Conceptual Framework (good practice for semantic resources)
3. Supporting further refinement of the DDI-Cross Domain Integration specification
4. Policy Monitoring Indicators (SDGs, Sendai etc)
5. Infectious Diseases: projects looking at data integration in HIV and COVID
6. Resilient and Healthy Cities: large group with a number of cities and projects, identifying shared themes.
DDI-Cross Domain Integration Collaboration

- DDI-CDI (Cross Domain Integration) is designed to interface with other standards and to help interoperability between different data types, standards, formats.
- Invite participation of International Scientific Unions and domain experts in a series of virtual workshops to identify use cases and further refine the specification.
- EOSC funding to consult on DDI-CDI with European Research Infrastructures and further refine the specification.
- Upcoming workshops on representation of units, on provenance, on environmental use cases and health/medical use cases.
Initial Pilot Activities: Combining Social Science and Health Data

- Project with LSHTM combining HDSS (health and demographic surveillance system) data with clinical data.
- Secure data system.
- Consistent semantics and metadata.
- DDI-CDI also playing a role.
- Aim to apply the same approach to COVID Data.
- Diagram courtesy Chufundo Kanjala, INSPIRE Project.
Initial Pilot Activities: Resilient Cities (Data-Knowledge-Action System)

- Decadal Programme Pilot Working Group on Resilient and Healthy Cities.
- Partnership with the ISC Programme on Urban Health and Wellbeing.
- Developing a conceptual model for Data Knowledge Action System.
- Data case studies: application of data audit, FAIR data.
  - E.g. mobility and contact tracing.
- Feeds into Interdisciplinary Collaborative Modelling with both data, community and expert inputs.
- Diagram courtesy of Franz Gatzweiler, UHWB.
Decadal Programme: Delivery Agents / Activities

Delivery Agents
- Coordinating Programme Office / Secretariat
- Cohort of Metadata and Ontology Experts
- Distributed programme offices / nodes
- Partner projects and working groups
- Preparing call for Management Committee and for Programme Offices and Partners.

Delivery Activities / outputs
- Consensus workshops
- Regular intensive sprint workshops (Dagstuhl model)
- Identification and description of issues in domains / across domains
- Alignment and harmonisation of metadata specifications, refinement of ontologies and taxonomic systems, development of cross-walks between vocabularies, and the application of automation and machine learning to assist data linking and integration.
Making Data Work for Cross-Domain Challenges

Aim to launch the Decadal Programme at the ISC GA and associated events ‘Global Knowledge Forum’ in Oman, 10-14 October 2021: https://council.science/about-us/governance/general-assembly/muscatassembly
FAIR Convergence Symposium

- Entirely virtual event.
- 30 November-4 December.
- Keynotes, interactive sessions, posters.
- Preparatory workshops on key themes: Sept-Nov.
- Call for Sessions, Posters and Lightning Talks: https://conference.codata.org/FAIRconvergence2020/
- Deadline for Session Proposals 30 September.
- Strongly encourage session proposals from CODATA China.
INTERNATIONAL DATA WEEK 2021
Data to Improve our World

8-11 NOVEMBER 2021
SEOUL, REPUBLIC OF KOREA
Data Skills and Training

**FAIRsFAIR Project: Major EU H2020 Project**

- Contributing to WP3 on Good Practices; WP5 on Synthesis and Synchronisation across initiatives; WP6 on FAIR Competence and Training.
- Laura Molloy employed by CODATA full-time on FAIRsFAIR.

**#terms4FAIRskills**

- Initiative to develop a community recognised terminology for FAIR data skills and competencies, EOSC Funding obtained for short project.

**Regular Beijing Data Science Training Workshops**

- Most recently in Sept 2019
- Other training workshops in Bangalore and Jakarta (2015).
- Helped scope the approach of the CODATA-RDA Data Schools.
CODATA-RDA Schools of Research Data Science

- Film: [https://vimeo.com/299263596](https://vimeo.com/299263596)
- New website for the initiative (under construction): [https://codata-rda-datascienceschools.github.io/](https://codata-rda-datascienceschools.github.io/)
- 2020: Pretoria... virtual school for alumni in September...
- 2018: Brisbane, Trieste, Trieste Advanced Workshops, Kigali, São Paulo
- 2017: Trieste, Trieste Advanced Workshops, São Paulo
- 2016: Trieste
CODATA Connect: Early Career and Alumni Group

CODATA Connect: [https://codata.org/initiatives/strategic-programme/codata-connect/](https://codata.org/initiatives/strategic-programme/codata-connect/)

- Initial Leads are Shaily Gandhi (India) and Felix Emeka Anyam (Nigeria).
  - Webinar Series on Resilient Cities
  - Webinar Series on Research Skills
  - Essay Competition and Datathon.

- Both are alumni of CODATA Data Schools; Shaily and Felix then organised a school on urban data science [https://sws.cept.ac.in/course-detail/urban-data-science-S19FT001](https://sws.cept.ac.in/course-detail/urban-data-science-S19FT001)

CODATA-RDA Data Schools Alumni

- Students > Helpers > Instructors > Directors...
- Alumni Sara El Jadid, Marcela Alfaro and Bianca Peterson are now co-chairs of the Data Schools.
- Virtual Alumni School in September.
Follow CODATA!

- CODATA Website: [http://www.codata.org/](http://www.codata.org/)
- CODATA on Twitter: @CODATANews and @simonhodson99
- Facebook: [https://www.facebook.com/codata.org/](https://www.facebook.com/codata.org/)
- Insta: [https://www.instagram.com/codatainternational/](https://www.instagram.com/codatainternational/)
Thank you for your attention

Alena Rybkina, CODATA
www.codata.org
a.rybkina@gcras.ru