COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL

International Science Council

CODATA Virtual General Assembly: International Collaboration for FAIR Data and Open Science

Simon Hodson
Executive Director
CODATA
www.codata.org

CODATA Virtual General Assembly
25 June 2020

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.**

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).

COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL

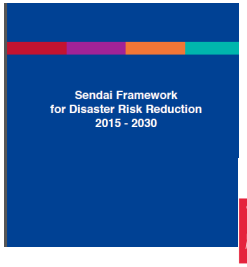
International
Science Council

Data 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**.

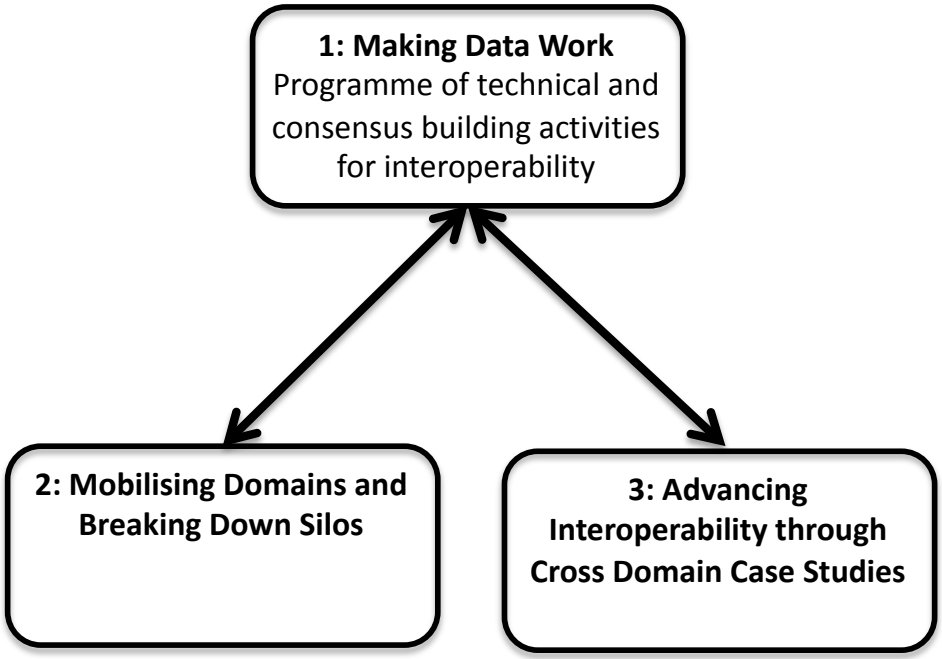
futureearth
research for global sustainability

IRDR
Integrated Research on Disaster Risk



URBAN HEALTH AND WELLBEING
A SYSTEMS APPROACH

Making Data Work: programme design



- 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. Semantic Interoperability and Conceptual Framework
2. Policy Monitoring Indicators (SDGs, Sendai etc)
3. Infectious Diseases: projects looking at data integration in HIV and COVID
4. Resilient and Healthy Cities: large group with a number of cities and projects, identifying shared themes.

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.

Initial Pilot Activities

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.
- Series of webinars to assist review of specification: https://bit.ly/DDI-CDI-Webinars
- Invite participation of International Scientific Unions and domain experts in a series of virtual workshops to identify use cases and further refine the specification.

Decadal Programme: Delivery Agents / Activities

Delivery Agents

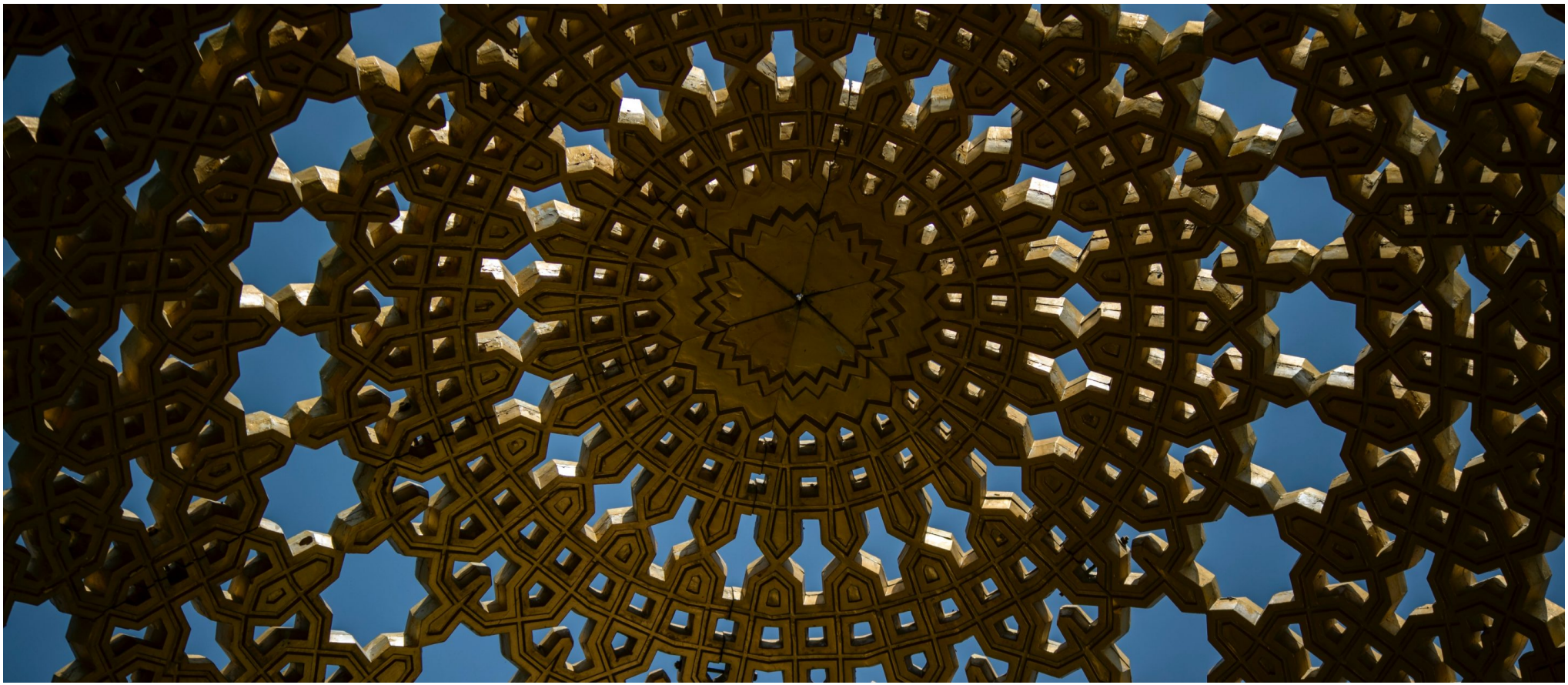
- Coordinating Programme Office / Secretariat
- Cohort of Metadata and Ontology Experts
- Distributed programme offices / nodes
- Partner projects and working groups

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



Data Policy Committee

Major Reports Since 2015

- **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 (with OECD, 2017) https://doi.org/10.1787/302b12bb-en
- Guidelines for the Legal Interoperability of Research Data (with RDA, 2016) https://doi.org/10.5281/zenodo.162241
- The Value of Open Data Sharing (for GEO, 2015) http://dx.doi.org/10.5281/zenodo.33830
- The Science International Accord on Open Data in a Big Data World (for ICSU, 2015) http://www.science-international.org/#accord
- Current Best Practice for Research Data Management Policies (for Danish e-Infrastructure Cooperation, 2015) http://dx.doi.org/10.5281/zenodo.27872

CODATA Twenty-Year Review of GBIF: http://doi.org/10.35035/ctzm-hz97

GBIF (Global Biodiversity Information Facility), major global data infrastructure and information facility.

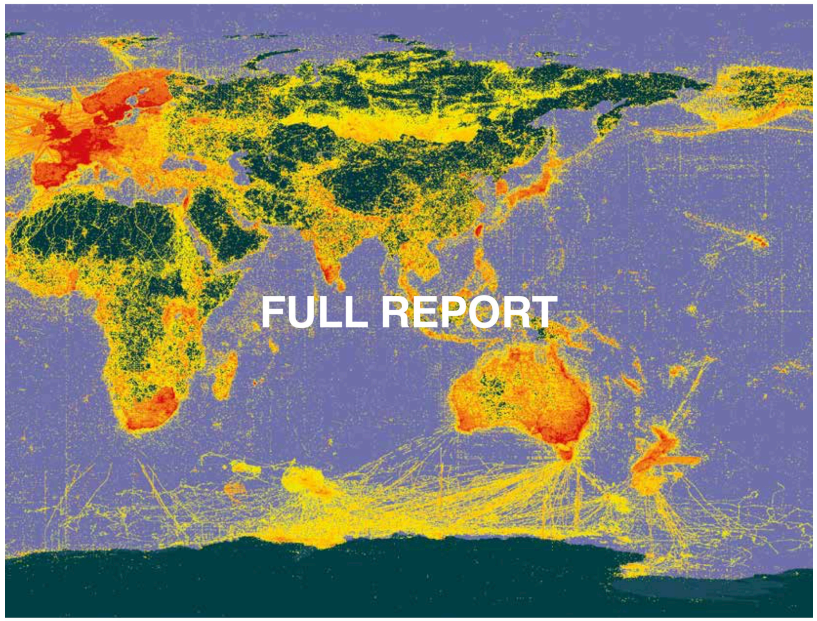
Important for CODATA to have a role in providing strategic overview: an area in which through the DPC we would like to do further work.

Excellent and thorough job done by Paul Uhlir and Hans Pfeiffenberger, supported by interviewers Aitong Li, Joseph Muliaro Wafula, Liliana Ballesteros-Mejia, William Michener, Simon Hodson and by community and CODATA reviewers.

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

TWENTY-YEAR REVIEW OF GBIF

GBIF Global Biodiversity Information Facility



Beijing Declaration on Research Data, 8 Nov 2019: http://bit.ly/Beijing-Declaration-Research-Data

Grand challenges related to the environment, human health, and sustainability confront science and society. Understanding and mitigating these challenges in a rapidly changing environment require data to be FAIR (Findable, Accessible, Interoperable, and Reusable) and as open as possible on a global basis.

Key Concepts

- Maximise (re-)use by humans and machines
- FAIR data and need for a FAIR ecosystem
- As Open as possible, only as closed as necessary
- No proprietary lock-in
- A Global Data Commons
- Address the global divide in scientific production and the UN Landmark Agreements.

Open Science for a Global Transformation

Open Science for a Global Transformation

CODATA coordinated submission to the UNESCO Open Science Review.

Led by the CODATA Executive Director, members of the Data Policy Committee and representatives of GO FAIR, WDS and ICSTI.

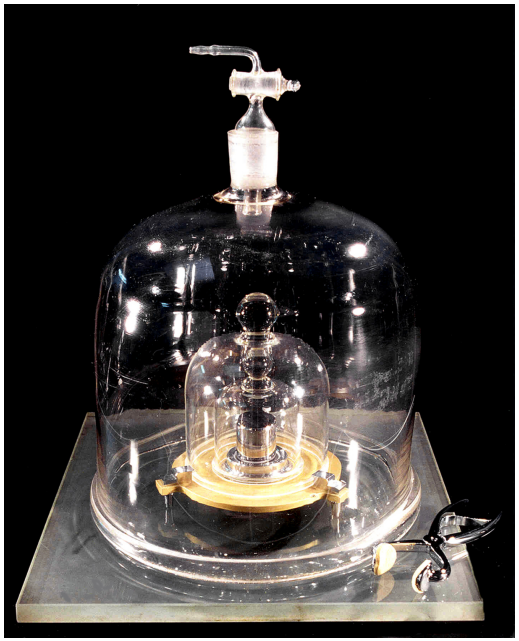
Data Policy Committee

New chair for the Data Policy Committee, will be appointed in July 2020.

Many thanks to the outgoing chair, Paul Uhlir!

Revision of the SI Units and the CODATA Fundamental Physical Constants

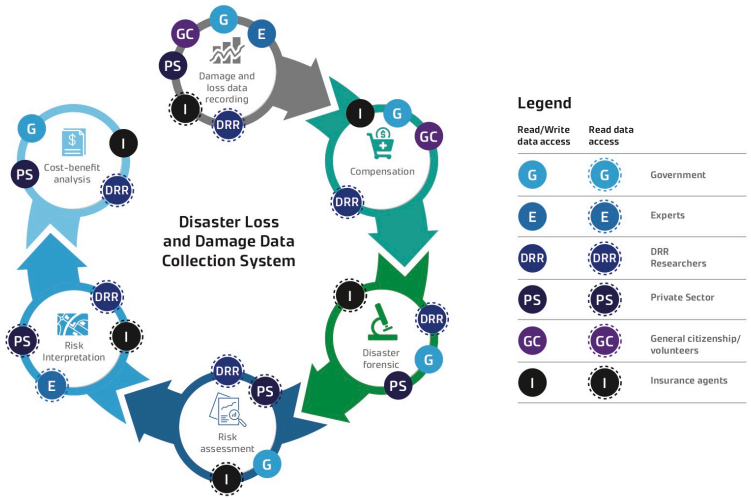
- Major revision of the SI Units agreed on 16 November 2018; came into force on 20 May 2019.
- The kilogram, ampere, kelvin and mole will now be based, respectively on the Planck constant h, the elementary charge e, the Boltzmann constant k, and the Avogadro constant NA.
- See http://bit.ly/codata-fundamental-constants and http://iopscience.iop.org/article/10.1088/1681-7575/aa950a/pdf



COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

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***

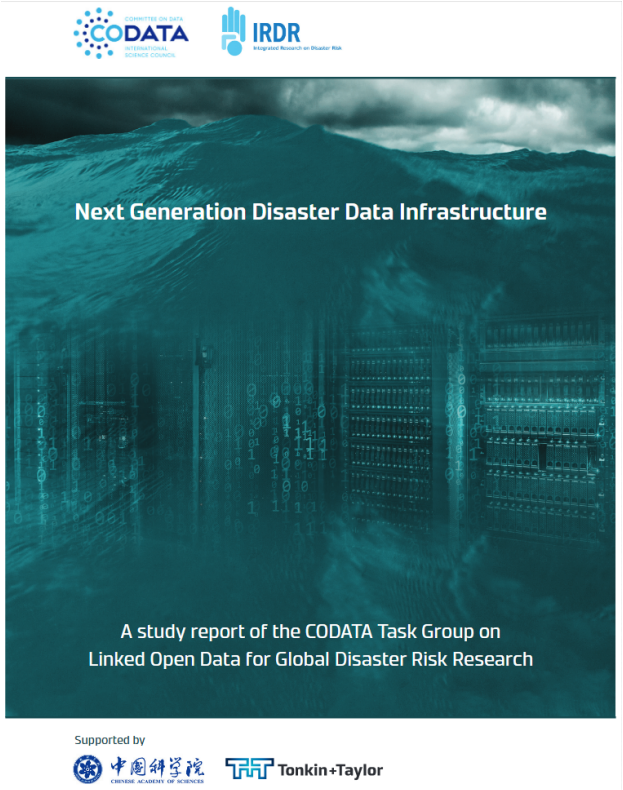


For a larger image, please visit https://tinyurl.com/DisasterLossData

http://bit.ly/TGGDRR-White_Paper_2

FAIR DRR

- Very active task group: White Paper series; new Policy Brief series.
  - White Paper, Next Generation Disaster Data Infrastructure https://doi.org/10.5281/zenodo.3406127
  - Policy Brief, 'Are we there yet? The transition from response to recovery for the COVID-19 pandemic' https://doi.org/10.1016/j.pdisas.2020.100102
  - Webinar series: see El Niño long range warning systems webinar https://bit.ly/el-nino-long-range-warning
- Monthly Disaster Risk Reduction and Open Data Newsletter: https://codata.org/blog/category/drr-and-open-data-newsletter/



DRUM

- DRUM TG preparing a position paper, will shortly communicate with International Scientific Unions, to do the following:
  - Make the case for the importance of digital units of measure;
  - Invite an ‘ambassador’ from each Union / Association to be the point of contact for DRUM and engage with the TG;
  - Where appropriate, the ‘ambassador’ will be proposed also to be the liaison with BIPM and nominated for an important workshop on the Digital SI;
  - Invite the Union to present use cases that demonstrate the utility and importance of digital representation of units of measure, or illustrate pain points.

Digital SI

Bureau International des Poids et Mesures

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.**
- See: http://www.codata.org/fair-data-training and https://terms4fairskills.github.io/Announcement.html

FAIRSFAIR
Fostering Fair Data Practices in Europe

Regular Beijing Data Science Training Workshops

- Most recently in Sept 2019
- Also in 2017, 2016, 2014, 2012.
- 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

- **CODATA-RDA Schools of Research Data Science: http://bit.ly/CODATA-RDA-data_schools**
- Film: https://vimeo.com/299263596
- New website for the initiative (under construction): https://codata-rda-datascienceschools.github.io/
- **2020: Pretoria… virtual school for alumni in September…**
- 2019: Addis, Trieste, Trieste Advanced Workshops, Costa Rica.
- 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/

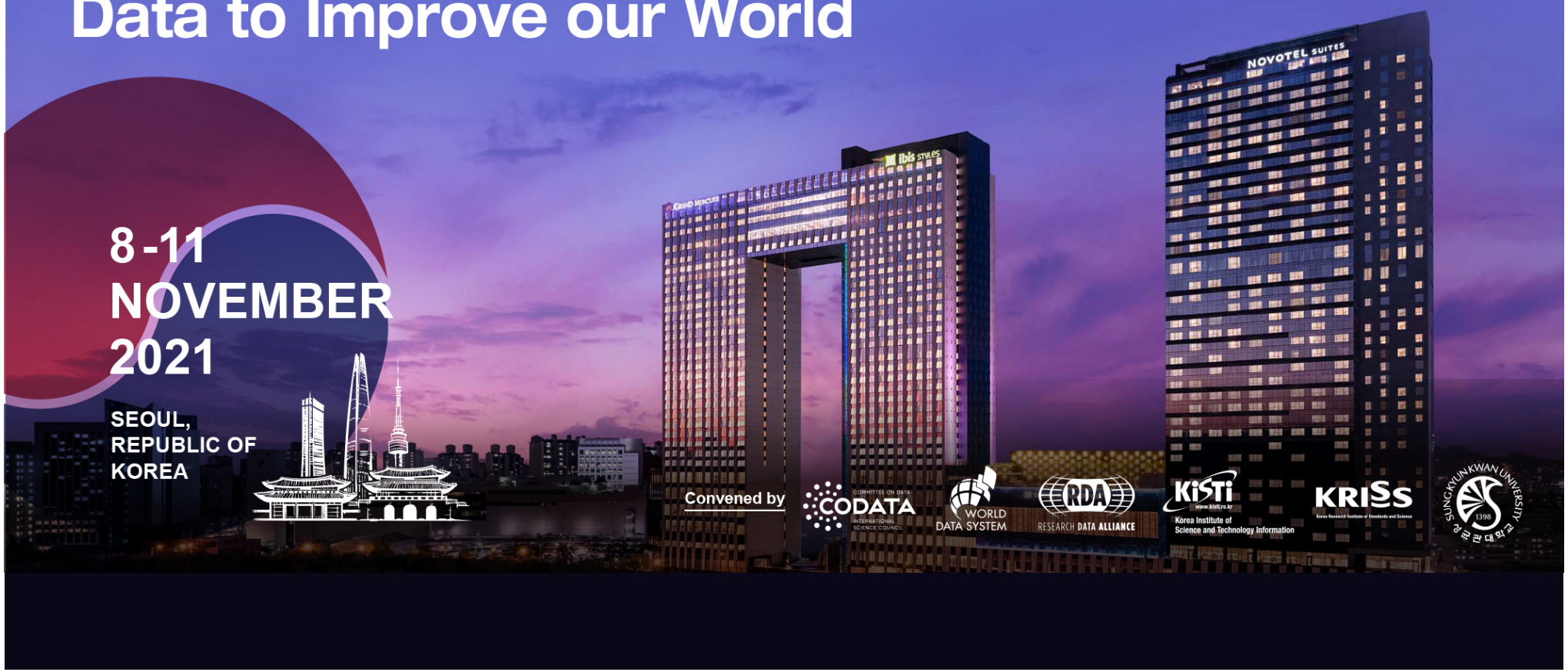
- Initial Leads are Shaily Gandhi (India) and Felix Emeka Anyam (Nigeria).
  - Webinar Series on Resilient Cities
  - Webinar Series on Research Skills
  - Essay Competition and Datathon.
- Both are alumni of CODATA Data Schools; Shaily and Felix then organised a school on urban data science https://sws.cept.ac.in/course-detail/urban-data-science-S19FT001

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.

INTERNATIONAL DATA WEEK 2021

Data to Improve our World



8 -11 NOVEMBER 2021

SEOUL, REPUBLIC OF KOREA

Convened by

CODATA COMMITTEE ON DATA INTERNATIONAL SCIENCE COUNCIL

WORLD DATA SYSTEM

RDA RESEARCH DATA ALLIANCE

KISTI www.kisti.re.kr Korea Institute of Science and Technology Information

KRISS Korea Research Institute of Standards and Science

SUNGKYUNKWAN UNIVERSITY 1398

FAIR Convergence Symposium

- Mixed and virtual event.
- 23-24 October 2020, Paris and Virtual.
- Call for Sessions, Posters and Lightning Talks: https://conference.codata.org/FAIRconvergence2020/
- Deadline 20 July.

Education for Data Science

- Workshop in partnership with the Israel Academy of Science and Humanities.
- Planned for February 2021, may change date or go virtual, **but it will happen!**
- Data Science education in the general curriculum, for domains, for all researchers, for specialists (masters and beyond).
- International comparisons.
- Invitation to all National Committees and to other members: submit abstract by 31 July 2020.

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

האקדמיה הלאומית הישראלית למדעים
المجمع الوطني الإسرائيلي للعلوم والآداب
THE ISRAEL ACADEMY OF SCIENCES AND HUMANITIES

ACADEMIA SCIENTIARUM ISRAELITICA

IDW 2023, A Festival of Data, Salzburg, 23-26 October 2023

CODATA Conferences 2019

CODATA Conference Series

- CODATA 2019 Beijing: Towards next-generation data-driven science: policies, practices and platforms: https://conference.codata.org/CODATA_2019/
- Major successful conference of over 300 participants; excellent keynotes and sessions.

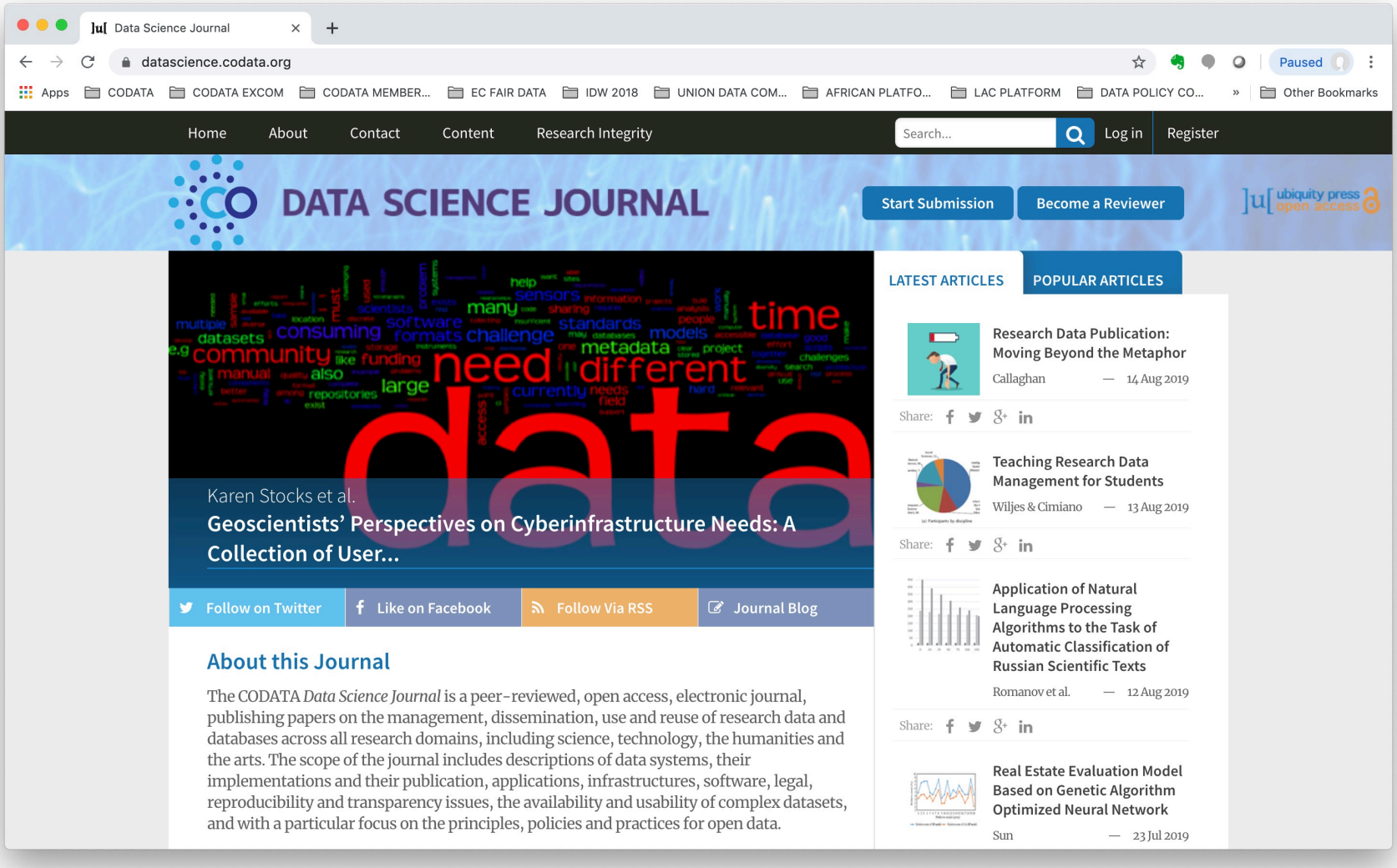
FAIR RDM in Institutions

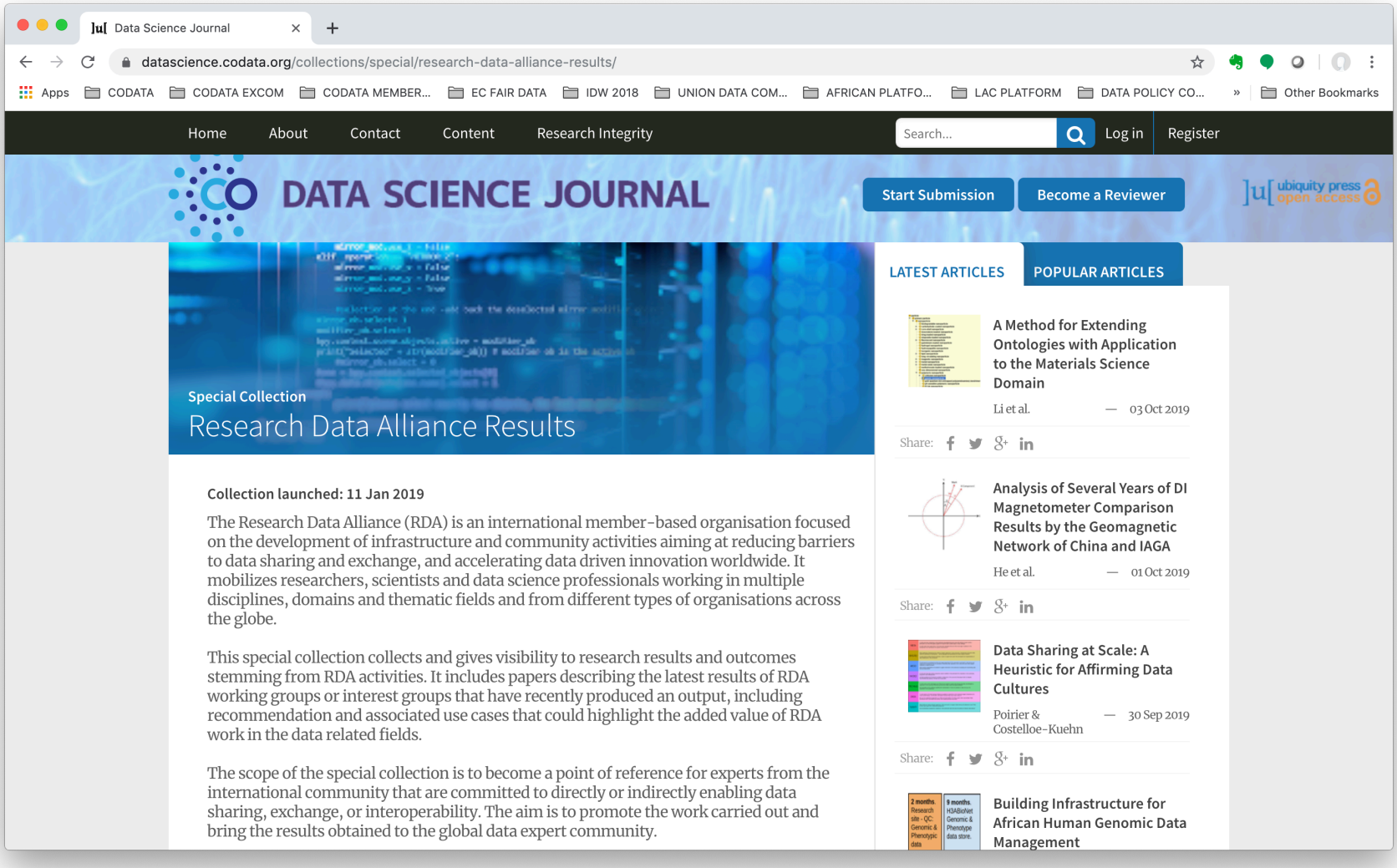
- CODATA-Drexel Workshop on FAIR RDM in Institutions 2019, Drexel University, Philadelphia, 31 March-1 April: https://conference.codata.org/Drexel_CODATA_2019/
- CODATA-Helsinki Workshop on FAIR RDM in Institutions 2019, National Archives of Finland, Helsinki, 20-21 October: https://conference.codata.org/Helsinki-CODATA-2019/

Photos: The National Archives of Finland/Marko Oja

VizAfrica Series

- VizAfrica 2019 Botswana: University of Botswana, Gaborone, 18-19 November: https://vizafrica.codata.org/2019-Botswana/


Data Science Journal
datascience.codata.org
Paused
Apps
CODATA
CODATA EXCOM
CODATA MEMBER...
EC FAIR DATA
IDW 2018
UNION DATA COM...
AFRICAN PLATFO...
LAC PLATFORM
DATA POLICY CO...
Other Bookmarks
Home
About
Contact
Content
Research Integrity
Search...
Log in
Register
DATA SCIENCE JOURNAL
Start Submission
Become a Reviewer
ubiquity press
open access
Karen Stocks et al.
Geoscientists' Perspectives on Cyberinfrastructure Needs: A Collection of User...
Follow on Twitter
Like on Facebook
Follow Via RSS
Journal Blog
About this Journal
The CODATA Data Science Journal is a peer-reviewed, open access, electronic journal, publishing papers on the management, dissemination, use and reuse of research data and databases across all research domains, including science, technology, the humanities and the arts. The scope of the journal includes descriptions of data systems, their implementations and their publication, applications, infrastructures, software, legal, reproducibility and transparency issues, the availability and usability of complex datasets, and with a particular focus on the principles, policies and practices for open data.
LATEST ARTICLES
POPULAR ARTICLES
Research Data Publication: Moving Beyond the Metaphor
Callaghan — 14 Aug 2019
Share:
Teaching Research Data Management for Students
Wiljes & Cimiano — 13 Aug 2019
Share:
Application of Natural Language Processing Algorithms to the Task of Automatic Classification of Russian Scientific Texts
Romanov et al. — 12 Aug 2019
Share:
Real Estate Evaluation Model Based on Genetic Algorithm Optimized Neural Network
Sun — 23 Jul 2019


Data Science Journal
datascience.codata.org/collections/special/research-data-alliance-results/
Apps
CODATA
CODATA EXCOM
CODATA MEMBER...
EC FAIR DATA
IDW 2018
UNION DATA COM...
AFRICAN PLATFO...
LAC PLATFORM
DATA POLICY CO...
Other Bookmarks
Home
About
Contact
Content
Research Integrity
Search...
Log in
Register
DATA SCIENCE JOURNAL
Start Submission
Become a Reviewer
ubiquity press
open access
Special Collection
Research Data Alliance Results
Collection launched: 11 Jan 2019
The Research Data Alliance (RDA) is an international member-based organisation focused on the development of infrastructure and community activities aiming at reducing barriers to data sharing and exchange, and accelerating data driven innovation worldwide. It mobilizes researchers, scientists and data science professionals working in multiple disciplines, domains and thematic fields and from different types of organisations across the globe.
This special collection collects and gives visibility to research results and outcomes stemming from RDA activities. It includes papers describing the latest results of RDA working groups or interest groups that have recently produced an output, including recommendation and associated use cases that could highlight the added value of RDA work in the data related fields.
The scope of the special collection is to become a point of reference for experts from the international community that are committed to directly or indirectly enabling data sharing, exchange, or interoperability. The aim is to promote the work carried out and bring the results obtained to the global data expert community.
LATEST ARTICLES
POPULAR ARTICLES
A Method for Extending Ontologies with Application to the Materials Science Domain
Li et al. — 03 Oct 2019
Share:
Analysis of Several Years of DI Magnetometer Comparison Results by the Geomagnetic Network of China and IAGA
He et al. — 01 Oct 2019
Share:
Data Sharing at Scale: A Heuristic for Affirming Data Cultures
Poirier & Costelloe-Kuehn — 30 Sep 2019
Share:
Building Infrastructure for African Human Genomic Data Management

Humans of Data

- The Editor-in-Chief of the Data Science Journal is **Human of Data 13**
- **http://codata.org/blog/category/humans-of-data/**
- Humans of Data is an art intervention into the international research data community by Laura Molloy laura@codata.org / @LM_HATII
- Images CC-BY-NC Laura Molloy

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

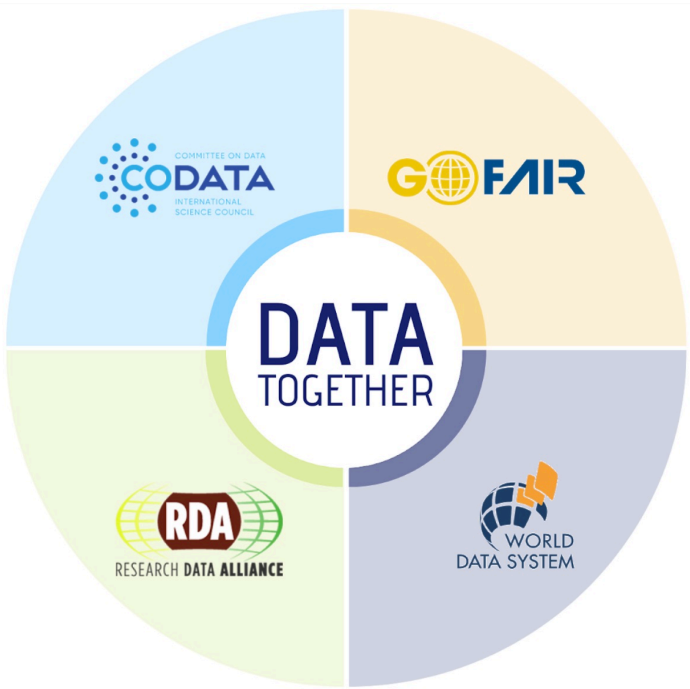
"Science is about discovering that things aren't as you expected. The more I learn, the more I realise I don't know. One of the fun things about what I do just now is that I get to see a lot of different research communities and how they conceive of and represent data, and what data mean to them.

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/

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/
  - 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/
- **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/



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

COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL

International Science Council

Thank you for your attention

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