

# Connect Your Research to the International Community through CODATA: Making Data Work for Cross-Domain Challenges

# 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

# Making Data Work for Cross-Domain Challenges: the Premise

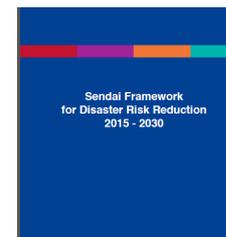
- The major, pressing global scientific and human issues of the 21st century can **ONLY** be addressed through **research that works across disciplines to understand complex systems**, and which uses a **transdisciplinary** approach to turn data into knowledge and then into action.
- The digital and 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 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.**

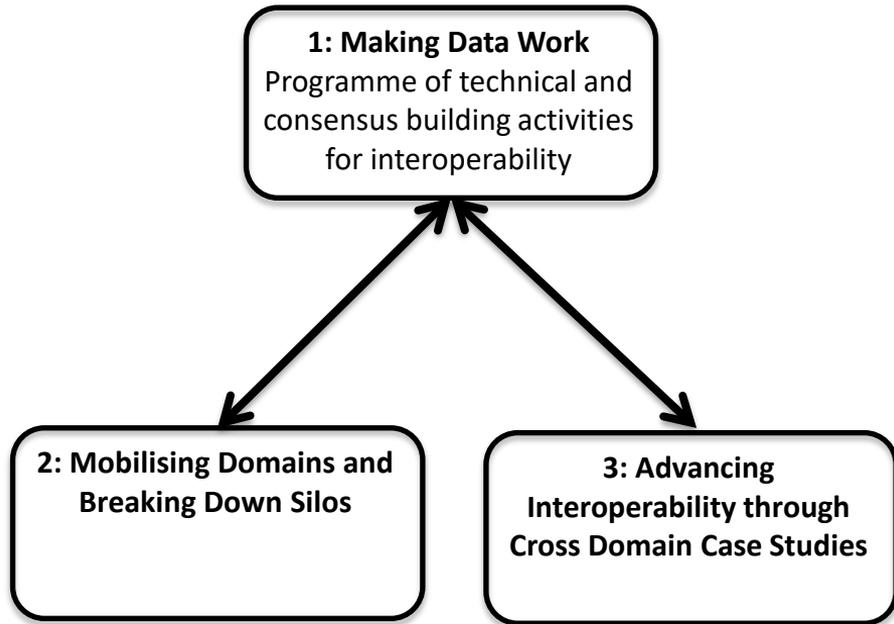
futurearth  
research for global sustainability



URBAN HEALTH  
AND WELLBEING  
A SYSTEMS APPROACH



# Making Data Work: programme design



- Programme 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.



# Initial Pilot Activities: DDI-CDI

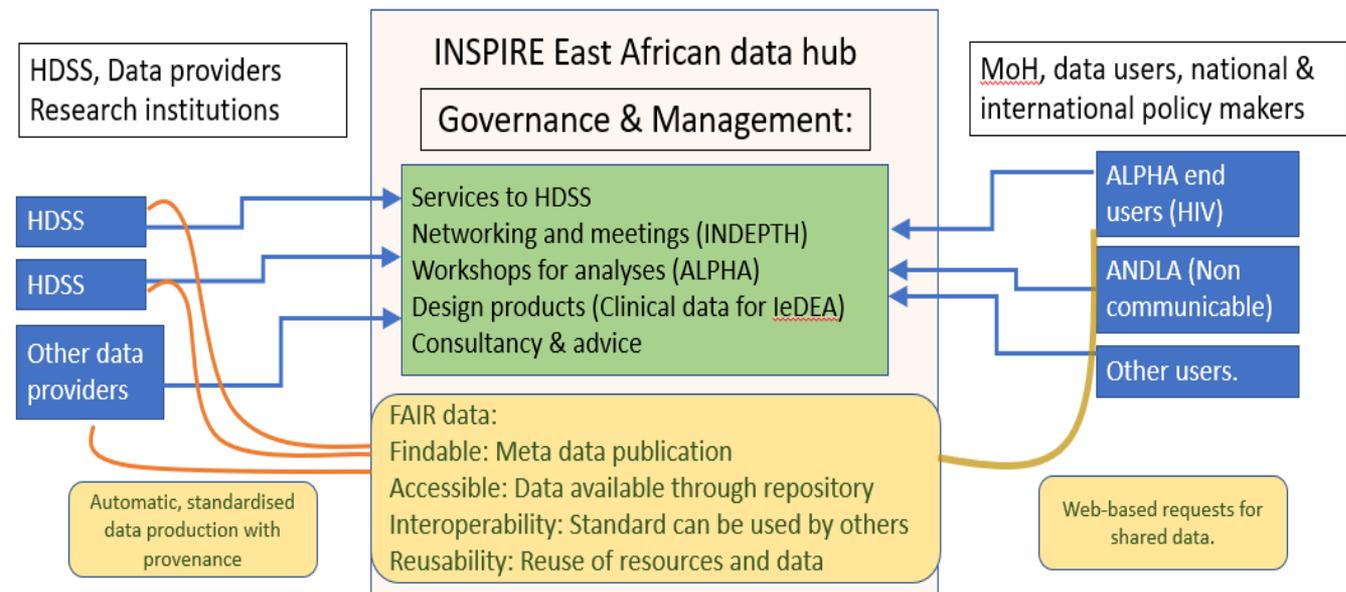


DATA DOCUMENTATION INITIATIVE

## DDI-Cross Domain Integration Collaboration

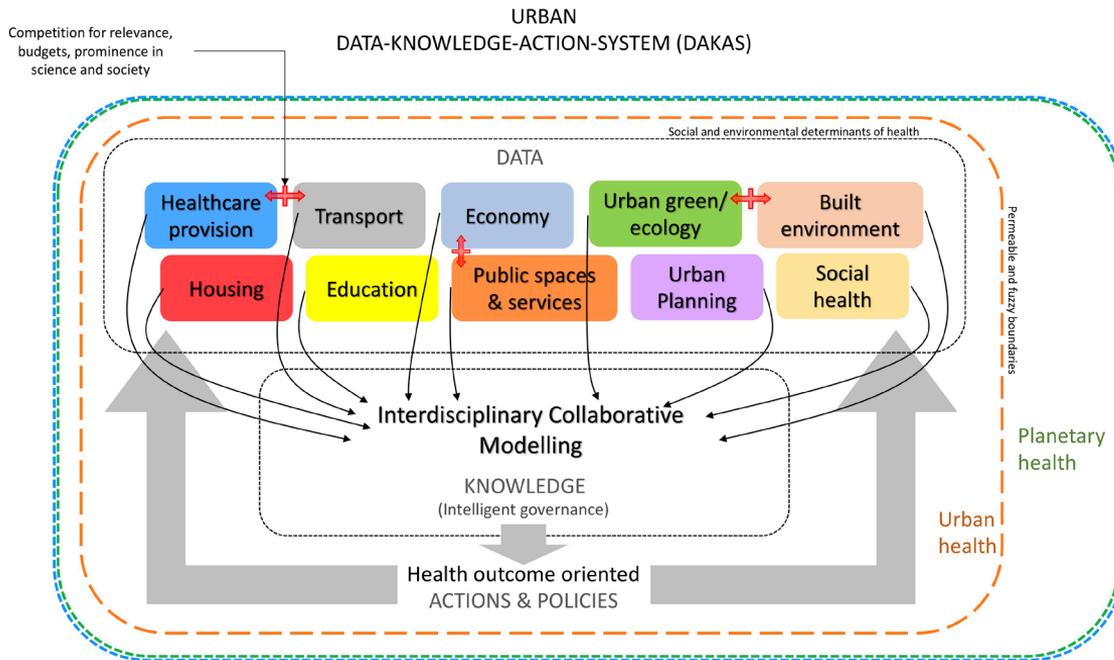
- DDI-CDI (Cross Domain Integration) is designed to interface with other standards and to help interoperability between different data types, standards, formats.
- Series of webinars to assist review of specification: <https://bit.ly/DDI-CDI-Webinars>
- Invite participation of International Scientific Unions and domain experts in a series of virtual workshops to identify use cases and further refine the specification.
- 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.**

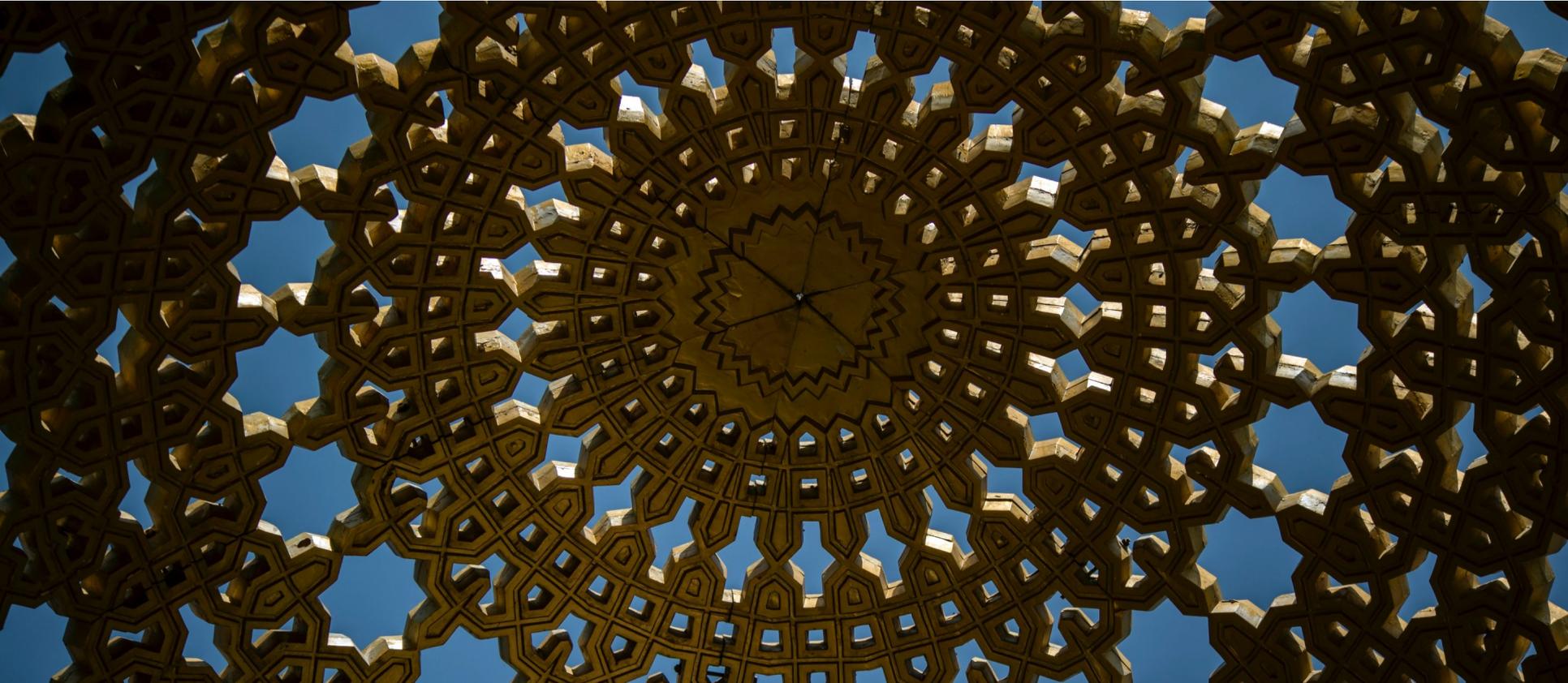


## Delivery Activities / outputs

- 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.
- 27 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 Posters and Lightning Talks is 31 October.**



# INTERNATIONAL DATA WEEK 2021

Data to Improve our World

8-11  
NOVEMBER  
2021

SEOUL,  
REPUBLIC OF  
KOREA



Convened by



# CODATA at UA: Current Projects

- Spacecraft to Statecraft
  - *Developing a data diplomacy framework*
- Data Diplomacy Training & Certification
  - *Developing a curriculum and credentials for graduate students*
- Flexible Regulations Framework
  - *Building a toolbox for policymakers to keep up with the rapid pace of data technology development and impact*
- Trusted Intermediaries Detecting Bias in Technology
  - *Designing Society-in-the-Loop interventions to advance fairness in AI technologies*
- Fluid Commons Collaboratory
  - *Co-creating a novel theory of self-assembling data trusts*
- Video as a Service Global Review
  - *Surveying the landscape of video communication platform users, services, infrastructure, and capacity across the globe*

# 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>





## Open Science for a Global Transformation

<b>Open Science for a Global Transformation</b>	1
<b>Key aspects of a transition to Open Science: Summary as input towards the UNESCO Recommendation</b>	2
Introduction: why is Open Science important and timely?	5
Data Together Organisations and Open Science	7
What are the objectives and benefits of Open Science?	7
Neglected aspects of Open Science	10
Open Science Infrastructures	12
Capacity Building for Open Science	18
Negative Impacts of Open Science and How to Address Them	20
A Global Consensus on Open Science: is it important and urgent?	22
What are the obstacles to reaching global consensus on Open Science and how can they be addressed?	23
Open Science and COVID-19	25
<b>Appendix 1: the Data Together Organizations</b>	28
<b>Appendix 2: Members of the Expert Group</b>	29

# Open Science for a Global Transformation

CODATA coordinated submission to the UNESCO Open Science

Review: <https://bit.ly/UNESCO-CODATA-Submission> and

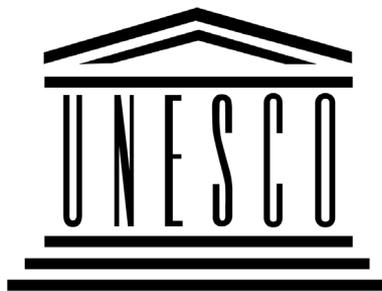
<https://doi.org/10.5281/zenodo.3935461>

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

## UNESCO Recommendation

Simon Hodson is Vice Chair of the International Advisory Group, currently preparing the draft Recommendation.

Further consultation Oct 2020-Jan 2021.



# 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  $N_A$ .
- See <http://bit.ly/codata-fundamental-constants> and <http://iopscience.iop.org/article/10.1088/1681-7575/aa950a/pdf>



# DRUM (Digital Representation of Units of Measure) TG

DRUM TG building engagement and collaboration between various stakeholders:

1. **Promote cooperation and coordination across initiatives**, and in particular mobilising the input of the various scientific domains, as represented by the ISUs/ISAs.
2. **Manifesto**, endorsed by the ISC and the ISUs/ISAs, calling for greater action and investment on the issue of units of measure (their definition, digital representation and conversion).
3. **Publications describing the landscape of units** and progress that can be made to facilitate coordination and machine applicability.
4. **Collection of case studies** from different domains highlighting challenges, making the case for coordination and presenting possible solutions.
5. **Recommendations** for the application of units of measure in metadata schema and code, building on existing good practices (starting with the DDI-Cross Domain Integration schema).

## Digital SI

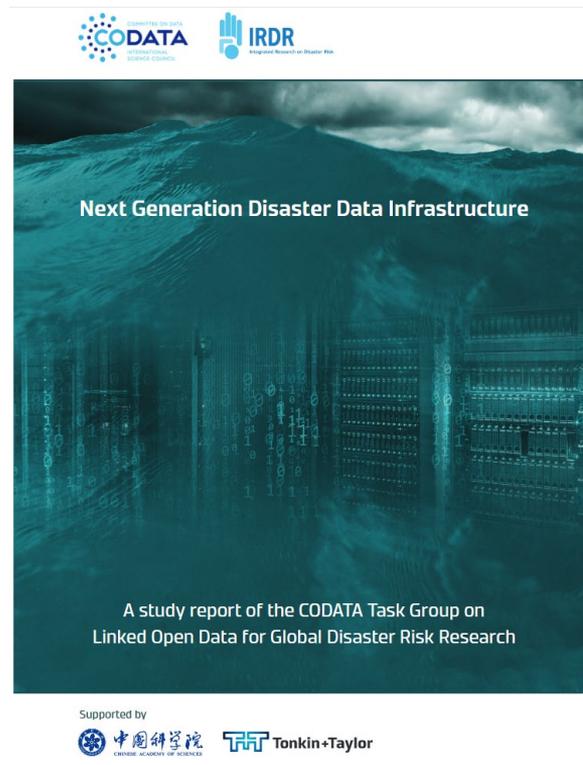
<https://www.bipm.org/en/conference-centre/bipm-workshops/digital-si/>



**Bureau**  
**International des**  
**Poids et**  
**Mesures**

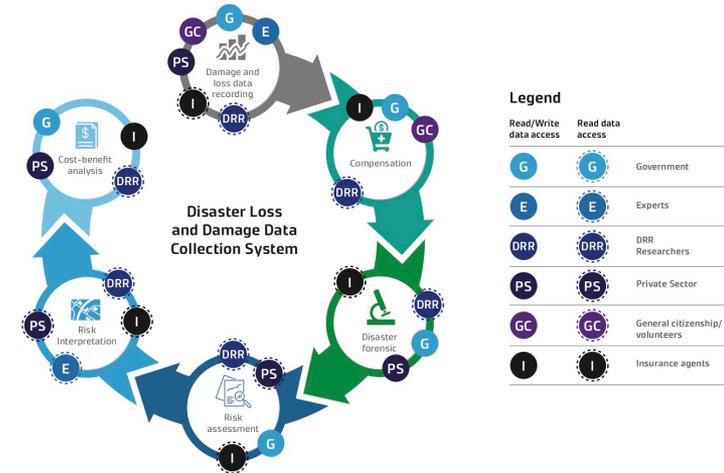
# FAIR DRR TG

- 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/>



# CODATA Task Groups

- Advanced mathematical tools for data-driven applied systems analysis
- Applying Data Integration and Data Science Tools toward Research of Urban Life and Smart Cities
- 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 – Kenyan led.
- Citizen Science for the SDGs – Aligning Citizen Science outcomes to the UN Sustainable Development Goals
- Linked Open Data for Global Disaster Risk Research – very active in DRR, white paper currently out for review.
- Preservation of and Access to Scientific and Technical Data in/for/with Developing Countries (PASTD) – focusing on LMICs



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

[http://bit.ly/TGGDRR-White\\_Paper\\_2](http://bit.ly/TGGDRR-White_Paper_2)

# 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 competencies, **EOSC Funding obtained for short project.**
- See: <http://www.codata.org/fair-data-training> and <https://terms4fairskills.github.io/Announcement.html>



# CODATA-RDA Schools of Research Data Science

- CODATA-RDA Schools of Research Data Science: [http://bit.ly/CODATA-RDA-data\\_schools](http://bit.ly/CODATA-RDA-data_schools)
- Film: <https://vimeo.com/299263596>
- New website for the initiative: <https://www.datascienceschools.org/>
- 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 - RDA

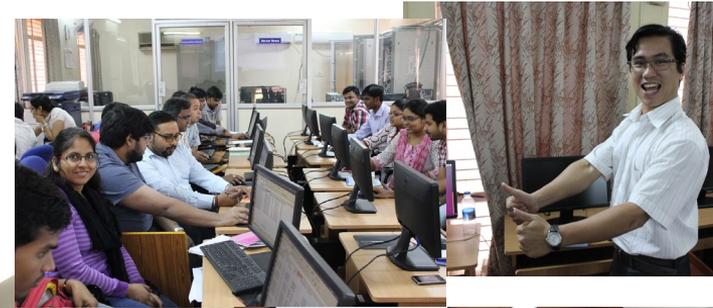
**Data**  
**Schools**



# Data Skills and Training

## Regular Beijing Data Science Training Workshops

- Most recently in Sept 2019
- Also in 2017, 2016, 2014, 2012.
- Other training workshops in Antananarivo, Madagascar (2016); Bangalore, India and Jakarta, Indonesia (2015); Nairobi, Kenya (2014).
- Helped scope the approach of the CODATA-RDA Data Schools.
- **Hopefully more in 2021!**



# 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: ongoing
  - Webinar Series on Research Skills: ongoing
  - Essay Competition: winner will be announced soon – likely to be run again next year.
  - Podcast series launched 'Data for Resilient Cities' <https://crdf.org.in/podcast/data-for-resilient-cities>; <https://soundcloud.com/dataforresilientcities>
- 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 2020.





Start Submission Become a Reviewer



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: f t g+ in

**Teaching Research Data Management for Students**  
 Wiljes & Cimiano — 13 Aug 2019

Share: f t g+ in

**Application of Natural Language Processing Algorithms to the Task of Automatic Classification of Russian Scientific Texts**  
 Romanov et al. — 12 Aug 2019

Share: f t g+ in

**Real Estate Evaluation Model Based on Genetic Algorithm Optimized Neural Network**  
 Sun — 23 Jul 2019



# 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



“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.



# 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](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](http://www.codata.org)**

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

**@simonhodson99 ; @CODATANews**