CODATA:
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.**

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

COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL

International Science Council

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

DRUM

- DRUM TG sent a position paper to the 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.**

  - **Address conversions of non-SI units, digitally referenceable units and conversions.**

Digital SI

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

Open Science for a Global Transformation

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.
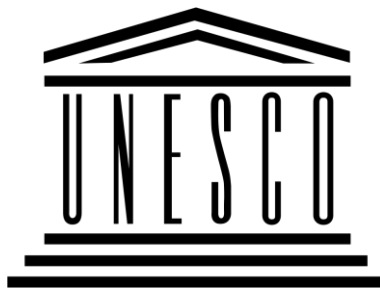
Open Science for a Global Transformation

- Science is a **global public good**. Open Science aims to maximise the benefit of science for society and the engagement of society with science.

- Open Science aims to maintain and promote good practice and **scientific reproducibility** by maximising access to robustly described data, code and methods underpinning scientific conclusions.

- **Open Science, through responsible governance, allows and requires necessary and proportionate protection of data, its sources, and derived information**. *It categorically does not mean unrestricted openness.*

- **Open by default. As open as possible, and only as closed as necessary.**

- Recommend the adoption **of Open Science frameworks**, including governance, ethics (CARE principles), core resources and means of communication.

- **Recommendations should lead to a set of Open Science Goals.**
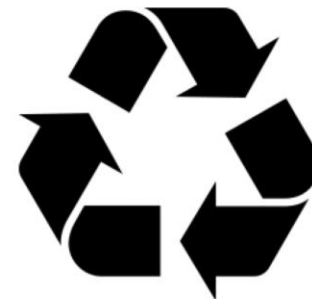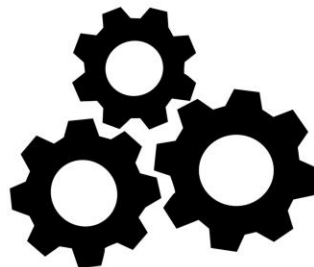
F indable   A ccessible   I nteroperable   R eusable

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)

COMMITTEE ON DATA
CODATA
INTERNATIONAL
SCIENCE COUNCIL

International
Science Council

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

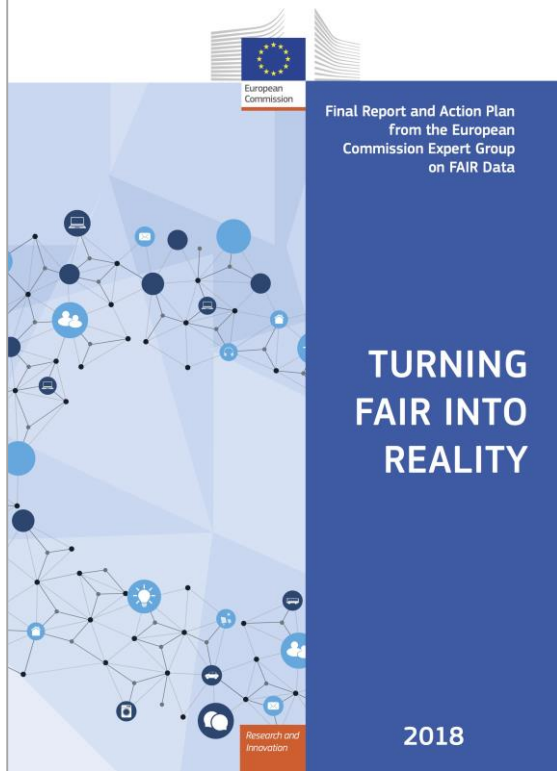- A2. metadata are accessible, even when the data are no longer available

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)

European Commission Expert Group, Chaired by Simon Hodson, Turning FAIR into Reality (2018)
https://doi.org/10.2777/1524

- **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.

African Open Science Platform

- Three year pilot project funded by DST and NRF in South Africa, delivered by ASSAf, directed by CODATA.

- Advocacy and community building.

- **Landscape survey of Open Science and data initiatives in Africa.**

- **Framework documents on: data policies, incentives, training, RDM in institutions and data infrastructure.**

- **Vision and Strategy Document: https://doi.org/10.5281/zenodo.2222418**

1. Federated network and cloud facilities.

2. Open science, FAIR and RDM tools.

3. Data Science and AI Institute.

4. Inter-disciplinary global challenge projects

5. Community network for education and skills in FAIR data and Open Science.

6. Network for societal engagement and participation.

- **Initial secretariat will be hosted for five years by NRF South Africa. Recently advertised for an Executive Director.**

THE **AFRICAN OPEN SCIENCE PLATFORM**

The Future of Science and Science for the Future

science & technology
Department
Science and Technology
REPUBLIC OF SOUTH AFRICA

ASSAf

International Science Council
Regional Office for Africa

CODATA

NRF
National Research Foundation

COMMITTEE ON DATA
CODATA
INTERNATIONAL SCIENCE COUNCIL

International Science Council

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: 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
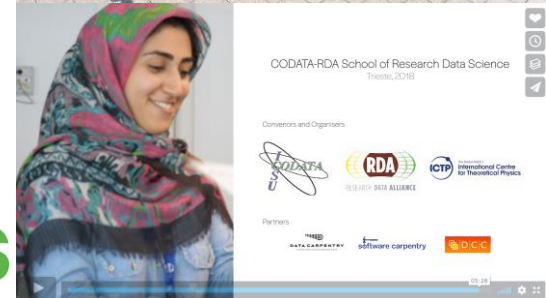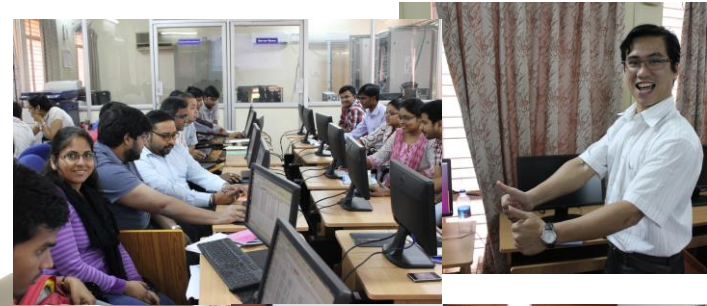
Data Skills and Training

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: currently being judged

  - Podcast series in preparation.

- 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.

Home   About   Contact   Content   Research Integrity

Search...   🔍   Log in   Register

CO   DATA SCIENCE JOURNAL

Start Submission   **Become a Reviewer**

Ju[ubiquity press open access

Karen Stocks et al.

Geoscientists' Perspectives on Cyberinfrastructure Needs: A Collection of User...

🐦 Follow on Twitter | f 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 🐦 g+ in

Teaching Research Data Management for Students
Wiljes & Cimiano   —   13 Aug 2019
Share: f 🐦 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 🐦 g+ in

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

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: f ⧫ 8+ in

Analysis of Several Years of DI Magnetometer Comparison Results by the Geomagnetic Network of China and IAGA

He et al.   — 01 Oct 2019

Share: f ⧫ 8+ in

Data Sharing at Scale: A Heuristic for Affirming Data Cultures

Poirier & Costelloe-Kuehn   — 30 Sep 2019

Share: f ⧫ 8+ in

Building Infrastructure for African Human Genomic Data Management

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.

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

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

Thank you for your attention

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