CODATA UPDATE

Bonnie C. Carroll
Secretary General, CODATA
www.codata.org

CODATA Strategy: Mobilizing the Data Revolution

- A 50 year old global organization, founded by the International Council of Science to address data issues.
- General Assembly represents all members
- Executive Committee has wide representation, includes members from 11 countries and 5 continents.
- Secretariat in Paris at ICSU HQ
- Member countries participate in TGs, WGs and other initiatives plus have their national agendas

Three priority areas essential to a coordinated international response to the data revolution.

- Promoting implementation of effective data principles, policies and practices;
- Advancing the frontiers of data science and adaptation to scientific research;
- Building capacity by improving data skills and the functions of science systems needed to support data management (particularly in LMICs)

Major Initiatives for Impact

ISSC international social science council

INTERNATIONAL COUNCIL FOR SCIENCE

ISSC & ICSU Members Vote to Pursue a Merger

World's top bodies representing the social and natural sciences vote to pursue a merger, forming a single organization representing all social and natural sciences by 2018

CODATA Management

- Growth and Sustainability

-New Country Membership
-New Commercial Members and Other Affiliates
-Project Grants and Contracts
-Hosting for CODATA Secretariat/Regional offices
-Secondments and Internships
-Mergers & Acquisitions

- Focus on CODATA Consolidated Network (HQ & Field)
  - Role of National Committees

Role of CODATA National Committies

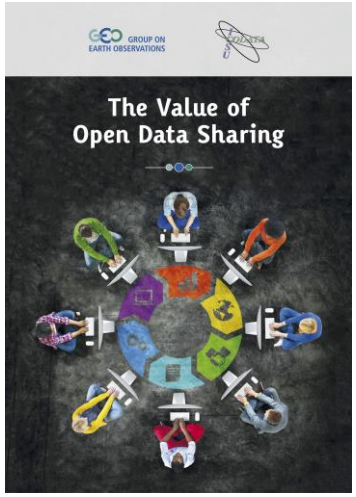
CODATA National Committees are composed of national stakeholders and data experts.

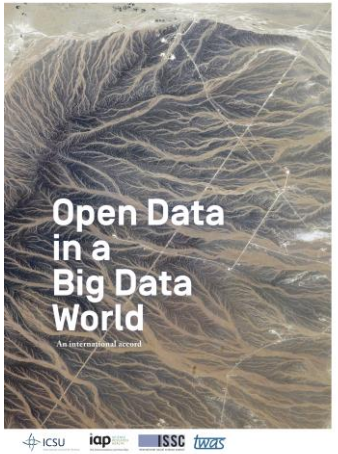
- **Opportunities**
  - Forum by which national stakeholders may advance data agenda in step with international developments;
  - Point of contact with CODATA;
  - Contribute to CODATA strategy;
  - **P**ropose Task Groups, host or participate in international workshop series;
  - **U**ndertake activities with other National Committees, bilaterally or in groups.

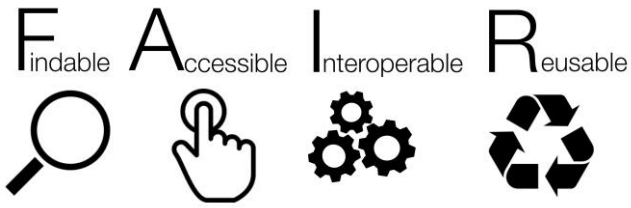
Actions for USNC at End of Presentation

CODATA Data Policy Activities

- Data Policy Committee, chaired by Paul Uhlir, international expert in Data Policies and member of CODATA Executive Committee.
- Current Best Practice for Research Data Management Policies http://dx.doi.org/10.5281/zenodo.27872
- The Value of Open Data Sharing, report for GEO http://dx.doi.org/10.5281/zenodo.33830
- Legal Interoperability, Principles and Implementation Guidelines https://doi.org/10.5281/zenodo.162241
- FAIR Data
  - Chairing the European Commission's Expert Group on FAIR Data: http://bit.ly/FAIR_Data_Expert_Group
- OECD Global Science Forum and CODATA Project on Business Models for Sustainable Data Repositories: http://www.codata.org/working-groups/oecd-gsf-sustainable-business-models

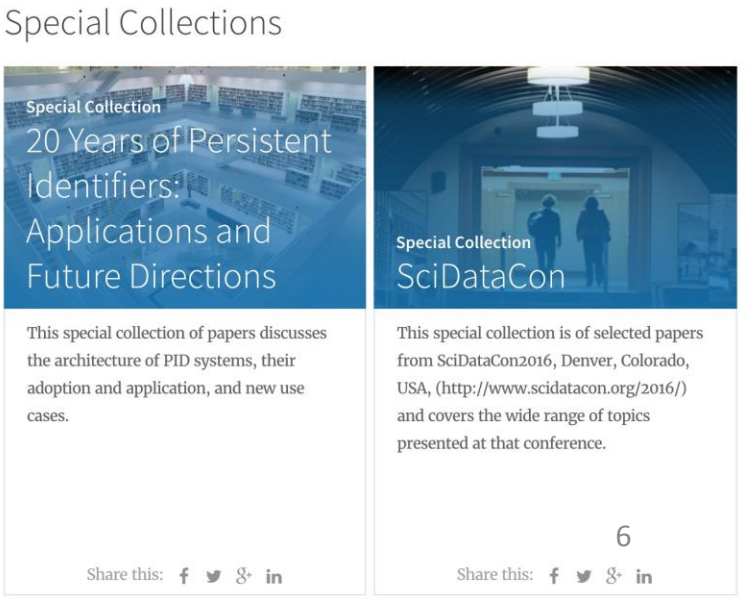






Advancing the Frontiers of Data Science

- Data Science Journal http://datascience.codata.org/
  - Data Science Journal Special Collections
    - Papers from SciDataCon 2016, part of International Data Week, now available
- New conference series:
  - International Data Week, every 'even' year: 2018, 2020…
  - CODATA Conference will happen every 'odd' year: 2017, 2019…
- Commission on Data Standards for Science
- Commission (Task Group) on Fundamental Constants
- Task Groups (elected by General Assembly every two years)
- Working Groups (approved by secretariat and Executive Committee to address strategic issues)



Commission on Data Standards for Science

Major transdisciplinary research issues depend on the integration of data and information from different sources.

- **'Inter-Union Workshop on 21st Century Scientific and Technical Data Developing a roadmap for data integration', Paris, 19-20 June: http://bit.ly/codata_standards_workshop**
  - Representatives of International Scientific Unions: IUCr for CIF; IUPAC for chemical terminologies; IUGS for GeoSciML; etc.
  - Representatives of Standards Organisations: e.g. Darwin Core, for biology, biodiversity; DDI for social science surveys; OGC for geospatial data; W3C for the web.
- Position paper in development.
  - Directory of activities involving international scientific unions.
  - Maturity model for vocabularies and standards.
  - Case studies of applications of vocabularies and standards for transdiciplinary research.
- Larger follow-up workshop 13-15 November, Royal Society, London.
- Vision of a decadal initiative to advance science through integration of data and information.

CODATA WG on Description of Nanomaterials

Convene ISUs, International Stakeholders and data experts → Form Working Group → Draft Framework for Description of Nanomaterials → Refine/validate in FP7 Future Nano Needs Project

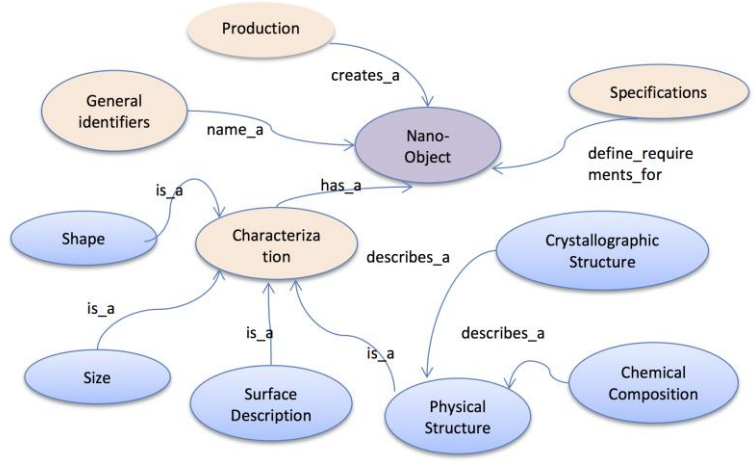


Figure 4. Information categories for describing an individual nano-object

CODATA WG on the Description of Nanomaterials:
http://www.codata.org/nanomaterials

Uniform Description System v.02, May 2016:
http://dx.doi.org/10.5281/zenodo.56720

Future Nano Needs Project:
http://www.futurenanoneeds.eu/

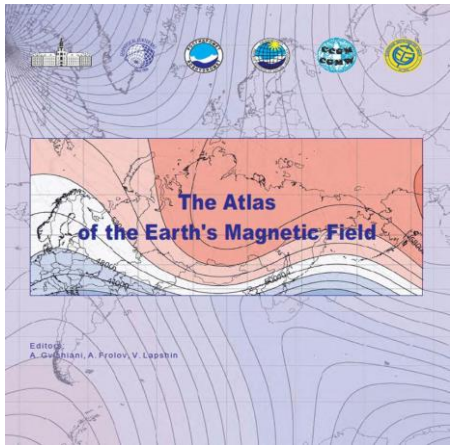
CODATA Task Groups 2016 - 2018

- Agriculture Data, Knowledge for Learning and Innovation
- Building Foundational Training in Research Data Science
- Citizen Science and the Validation, Curation, and Management of Crowdsourced Data (with WDS)
- Coordinating Data Standards amongst Scientific Unions
- Data Citation Standards and Practices
- Earth and Space Science Data Interoperability
- Linked Open Data for Global Disaster Risk Research
- Preservation of and Access to Scientific and Technical Data in/for/with Developing Countries (PASTD)

Challenges in Data Science

TG Earth-Space Science Data Interoperability

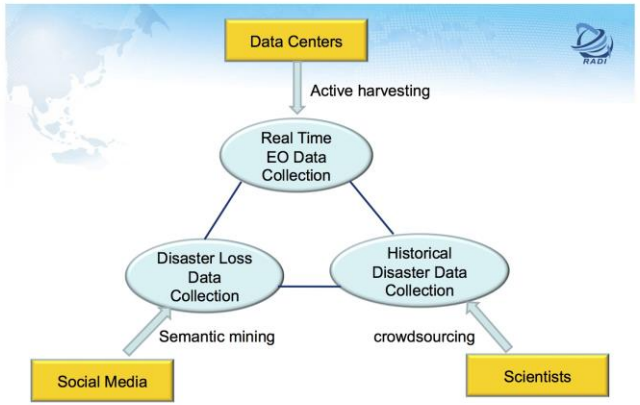
Preparing second edition of Atlas of the Earth's Magnetic Field http://bit.ly/atlas-magnetic-field



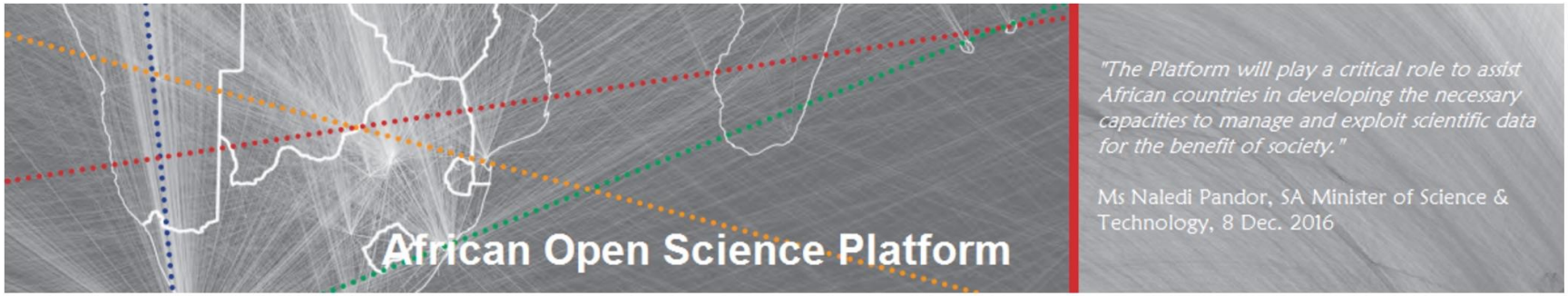
- Contribution to standards for multidisciplinary GIS for geoscience data
- **Increased focus on interoperability and standardisation issues.**
- International collaboration for conferences and training activities (Moscow and Sochi, July 2016; Peterhof, October 2017).

TG LOD Global Disaster Risk Research

- White Paper: 'Gap Analysis on Open Data Interconnectivity for Global Disaster Risk Research' http://bit.ly/White_Paper-LOD_Disaster_Gap_Analysis
- Important response from the perspective of science and data to post-Sendai framework
- **Inviting comments until 30 September.**



Directed by CODATA and delivered by The Academy of Science of South Africa (ASSAf)



African Open Science Platform

- High level advisory council: chaired by Khotso Mokhele, Chancellor, UFS, South Africa
- Technical advisory board: chaired by Professor Joseph Wafula, CODATA ExComm, JKUAT, Kenya.
- Workshops at Science Forum South Africa (Dec 2016) and at Association of African Universities (June 2017). AAU is a key partner. Building partnership with RUFORUM and UbuntuNet (Nov 2017) and others.
- Meetings of National Data Fora in Madagascar (September) and Botswana (October).
- National Forum on Open Data Open Science, hosted at the University of Botswana, 30-31 October 2017. Initiative to engage with issues such as Data Legislation Governance and Policy; Coordinated Research Data Cyber-Infrastructure; Data Awareness & Capacity Building; Data Management and Repositories.

African Open Science Platform
Pilot Project Work Packages

Establish African Open Data Forum / Platform

Co-design African Open Data Policies

Develop Incentives Frameworks

Develop Research Data Science Training

African Research Data Infrastructure Roadmap

Activities require low funding for coordination, secondment, contributions in kind and evaluation.

Activities require higher investment for coordination, co-design implemenatation and evaluation.

Funded Research Data Infrastructure Initiatives

Funded, co-designed transdisciplinary research projects

CODATA-RDA Data Science Training Initiative

- Annual foundational school hosted at ICTP, Trieste (with the objective to build a network of partners, train-the-trainers).
- Advanced workshops, ICTP, Trieste, following the foundational school.
- National or regional schools, organized with local partners.
- Planning at least two pilot schools as part of the African Open Science Platform project.

- Next #DataTrieste Summer School, 6-17 August 2018.
- Next #DataTrieste Advanced Workshops 20-24 August 2018.
- Next regional foundational school 'CODATA-RDA School of Research Data Science', São Paulo, 4-15 December 2017: http://www.ictp-saifr.org/?page_id=15270 - **deadline 22 September**

Foundational Curriculum

Seven components: open science, data management and curation; software carpentry; data carpentry; data infrastructures; statistics and machine learning; visualisation.

Builds on much existing courses to create something more than the sum of its parts:

- **Open Science** – reflection on ethos and requirements of sharing/openness
- **Open Research Data** – Basics of data management, DMPs, RDM life-cycle, data publishing, metadata and annotation
- **Author Carpentry** – Improving research efficiency with command line and OS tools.
- **Software Carpentry** – Introduction the Unix shell and Git (sharing software and data)
- **Data Carpentry** – Introduction to programming in R, and to SQL databases
- **Visualisation** – Tools, Critical Analysis of Visualisation
- **Analysis** – Statistics and Machine Learning (clustering, supervised and unsupervised learning)
- **Computational Infrastructures** – Introduction to cloud computing, launching a Virtual Machine on an IaaS cloud

Building international network of short courses http://bit.ly/first_data_school_trieste

Programme and materials: **http://bit.ly/School_of_Research_Data_Science-Programme** ; **http://bit.ly/first_data_school_materials**

CODATA and Data Science

Capacity Building: Training

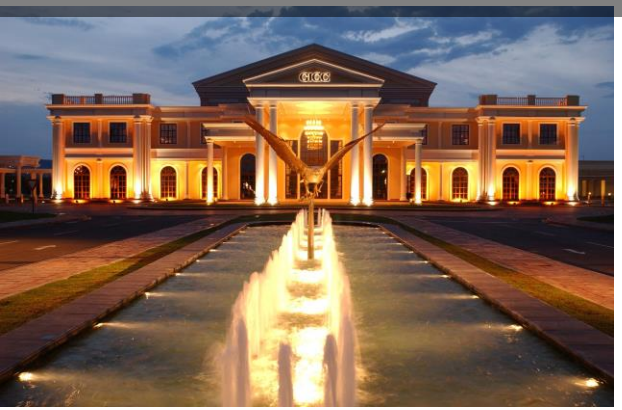
CODATA Training in Big Data for Science
Beijing, 4-20 July 2016
http://bit.ly/CODATA-China_Training_2104-Call

CODATA ISI Workshop on Big Data for Science, Indian Statistical Institute, Bangalore, 9-20 March 2015
http://drtc1.isibang.ac.in/bdworkshop/


INTERNATIONAL DATA WEEK

IDW 2018

Gaborone, Botswana: 22–26 October 2018



CODATA ICSU WORLD DATA SYSTEM

RDA RESEARCH DATA ALLIANCE

UNIVERSITY OF BOTSWANA

USNC/BRDI Business Decisions

- Nominations
  - President (2018-22)
  - 2 Vice Presidents (2018-22)
  - Ordinary members (8) (2018-20)
- International Task Group Topics
  - US initiatives for International Input
- Operation of the US National Committee within BRDI
  - Continue as informal subcommittee
  - Players
    - CODATA International Officers
    - US TG Chairs
    - Co-opted (e.g. Past Officers)
    - BRDI Members -- Additional involvement

Thank you!

Bonnie C. Carroll
Secretary General

Many Slides courtesy of
Simon Hodson
Executive Director CODATA
www.codata.org
http://lists.codata.org/mailman/listinfo/codata-international_lists.codata.org
Email: simon@codata.org
Twitter: @simonhodson99
Tel (Office): +33 1 45 25 04 96 | Tel (Cell): +33 6 86 30 42 59

CODATA Recommended Values of the Fundamental Physical Constants, 2014: http://dx.doi.org/10.5281/zenodo.22826

2014 CODATA RECOMMENDED VALUES OF THE FUNDAMENTAL CONSTANTS OF PHYSICS AND CHEMISTRY NIST SP 959 (Aug 2015)

See: P. J. Mohr, D. B. Newell, and B. N. Taylor, arxiv.org/pdf/1507.07956v1.pdf (2015). A more extensive listing of constants is available in the reference given above and on the NIST Physical Measurement Laboratory Web site: physics.nist.gov/constants.

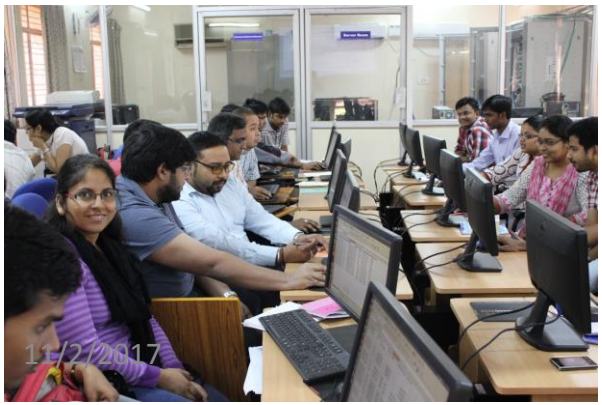
Quantity	Symbol	Numerical value	Unit
speed of light in vacuum	c, c_0	299 792 458 (exact)	$\mathrm{m\ s^{-1}}$
magnetic constant	μ_0	$4\pi \times 10^{-7}$ (exact)	$\mathrm{N\ A^{-2}}$
electric constant $1/\mu_0 c^2$	ϵ_0	$8.854\,187\,817... \times 10^{-12}$	$\mathrm{F\ m^{-1}}$
Newtonian constant of gravitation	G	$6.674\,08(31) \times 10^{-11}$	$\mathrm{m^3\ kg^{-1}\ s^{-2}}$
Planck constant	h	$6.626\,070\,040(81) \times 10^{-34}$	J s
$h/2\pi$	$\hbar$	$1.054\,571\,800(13) \times 10^{-34}$	J s
elementary charge	e	$1.602\,176\,6208(98) \times 10^{-19}$	C
fine-structure constant $e^2/4\pi\epsilon_0\hbar c$	α	$7.297\,352\,5664(17) \times 10^{-3}$	
inverse fine-structure constant	α^{-1}	137.035 999 139(31)	
Rydberg constant $\alpha^2 m_e c/2h$	R_∞	10 973 731.568 508(65)	$\mathrm{m^{-1}}$
Bohr radius $\alpha/4\pi R_\infty$	a_0	$0.529\,177\,210\,67(12) \times 10^{-10}$	m
Bohr magneton $e\hbar/2m_e$	μ_B	$927.400\,9994(57) \times 10^{-26}$	$\mathrm{J\ T^{-1}}$

DataTrieste Film on Vimeo

CODATA-RDA School of Research Data Science 2017, ICTP, Trieste: #DataTrieste

CODATA and Data Science Capacity Building: Training

CODATA ISI Workshop on Big Data for Science, Indian Statistical Institute, Bangalore, 9-20 March 2015
http://drtc1.isibang.ac.in/bdworkshop/



CODATA LIPI Workshop on Big Data, Tools and Data Management, LIPI, Jakarta, 2-4 Sept 2015
http://www.codata.org/events/workshops/workshops-2015/training-workshop-on-big-data-jakarta

CODATA and Data Science Capacity Building: Training

CODATA Training in Big Data for Science
Beijing, 4-20 June 2014
http://bit.ly/CODATA-China_Training_2104-Call

CODATA PASTD Training Workshop on Open Data, Kenya, Jomo Kenyatta University of Science and Technology, 3-5 August 2014
http://bit.ly/codata-training-jkuat

CODATA-RDA School of Research Data Science

- Contemporary research – particularly when addressing the most significant, transdisciplinary research challenges – increasingly depends on a range of skills relating to data. **These skills include the principles and practice of Open Science and research data management and curation, the development of a range of data platforms and infrastructures, the techniques of large scale analysis, statistics, visualisation and modelling techniques, software development and data annotation.** The ensemble of these skills, relating to data in research, can usefully be called 'Research Data Science'.

software carpentry

DATA CARPENTRY
MAKING DATA SCIENCE MORE EFFICIENT

Leverage

- **From September 2015 to September 2016, the annual income from membership fees of c.€205K leveraged further investment in activities to a total of over €1.9M: a leverage ratio of over 9.6:1**. This estimate includes external contributions to events, Task Groups and similar activities, sponsorship obtained as well as host and participant investment in events. As a specific example, in August 2016, the CODATA-RDA School of Research Data Science was held at ICTP in Trieste. CODATA's own investment in the event totals c.€10K in travel and student support. The event as a whole leveraged an additional c.€270,000 in support, comprising international and local travel and accommodation for experts and students as well as sponsorship and local expenses. **It should be noted that this estimate considerably undervalues the CODATA's leveraging power as it does not include any estimate for contributions in kind (e.g. co-chairs time).**

- **Next two years, concerted outreach to expand membership and to engage more with National Committees.**

Policy Push for Open Research Data

- The three Bs (Budapest, Berlin and Bethesda) and Open Access, 2002-3
- OECD Principles and Guidelines on Access to Research Data, 2004, 2007
- UK Funder Data Policies, from 2001, but accelerates from 2009
- NSF Data Management Plan Requirements, 2010
- Royal Society Report 'Science as an Open Enterprise', 2012
- OSTP Memo 'Increasing Access to the Results of Federally Funded Scientific Research', Feb 2013
- G8 Science Ministers Statement, June 2013
- G8 Open Data Charter and Technical Appendix, June 2013
- EC H2020 Open Data Policy Pilot, 2014; Adoption of FAIR Data Principles, 2017.
- Science International Accord on Open Data in a Big Data World, Dec 2015: http://bit.ly/opendata-bigdata

The Case for Open Data in a Big Data World

- **Science International Accord on Open Data in a Big Data World: http://www.science-international.org/**
- Supported by four major international science organisations.
- Presents a powerful case that the profound transformations mean that data should be:
  - Open by default
  - Intelligently open, FAIR data
- **Lays out a framework of principles, responsibilities and enabling practices for how the vision of Open Data in a Big Data World can be achieved.**
- Campaign for endorsements: over 150 organisations so far.
- **Please consider endorsing the Accord: http://www.science-international.org/#endorse**