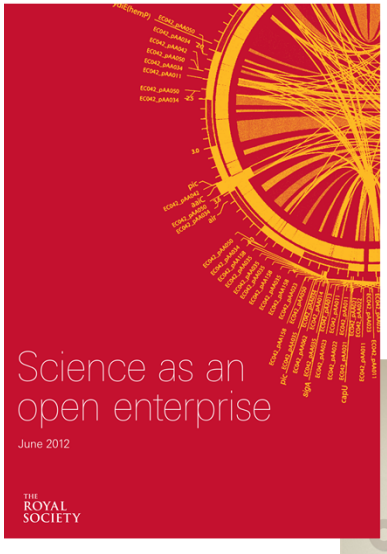
CODATA US Meeting
Washington DC
19 November 2015

# Responses to the data revolution: CODATA's work on data policy, data science and capacity building

Dr Simon Hodson
Executive Director, CODATA
www.codata.org

# Data Revolution:
# Science as an Open Enterprise






- The digital age has brought a data revolution that presents science with major challenges and opportunities.
- Opportunities because we can gather unprecedented volumes and types of data and analyse them far more quickly.
- **Exploiting these opportunities is the major challenge of international science.**
  - Challenges for data infrastructure, networks and analysis.
  - Fundamental methodological issues for reproducibility and transparency.
  - Challenges and opportunities for science systems, technical and human.
- **Mobilising Big Data requires Open Data!**
- **Data for research should be intelligently open: accessible, assessible, intelligible, useable.**
- **Creating a world that counts:** Mobilising the Data Revolution for Sustainable Development.
- **GODAN-ODI Report:** improving agriculture, food and nutrition with open data.

# CODATA Strategy: Mobilising the Data Revolution

CODATA President Geoffrey Boulton, FRS Chair of *Science as an Open Enterprise* Report

Simon Hodson CODATA Executive Director


New CODATA Executive Committee elected at GA in New Delhi, Nov 2014

Mobilising capacity to take advantage of the data revolution is **the** major priority for international science.

CODATA strategy lays out three priorities and a plan that shows we can **deliver benefits** for members on these priorities.

Presents a vision for international promotion and coordination of data ecosystems.

Promote intelligently open data

- **data policies:** supporting implementation of data principles and practice

Adapt to the transformation in research

- **data science:** addressing the frontier issues of data science

Promote data skills, data scientists, data managers

- **research data capacity building** (particularly in LMICs)

# CODATA and Open Data Policies


Current Best Practice for Research Data Management Policies

*A Memo for the Danish e-Infrastructure Cooperation and the Danish Digital Library*

Simon Hodson and Laura Molloy

May 2014

Expert Report on data policies for Danish e-Infrastructure Group

Regional Workshops on Data Citation Principles and Practice, South Africa, October


ICT Centre of Excellence and Open Data
Jomo Kenyatta University of Agriculture and Technology

Home | About Us | Open Data | Research and Innovation | Technology Transfer | Contacts Us

ICT Centre of Excellence and Open Data- iCEOD > CODATA Kenya

**CODATA Kenya**

The main aim of CODATA is to improve the quality, reliability, management and accessibility of data of importance to all fields of science and technology. CODATA has three main priorities as per its strateg entitled "Exploiting the Data Revolution:the CODATA strategy".

1) Data Policy- Supporting implementation of data principles and practices
2) Data Science- Addressing the frontiers of data science and its adaptation to scientific research.
3) Data Education- Capacity building (particularly in low and middle income countries (LMICs)

At the CODATA International Workshop on Open Data for Science and Sustainability in Developing Countries held at JKUAT and at UNESCO in Nairobi, it was agreed that JKUAT, would join CODATA on behalf of Kenya. Later, the CODATA General Assembly held in October, 2014 in India, accepted and welcomed Kenya warmly into the CODATA family.

- **CODATA Data Policy Committee:** key means of delivery.
  - 'The Data Agenda for International Science'
- **Register of Good Practice and Data Policy Assessment Tool**
  - Expert Report on data policies Danish e-Infrastructure Group: http://dx.doi.org/10.5281/zenodo.27872
  - Means of assisting good practice and self-evaluation for national authorities, research institutions and data intensive programmes.
- **Regional Workshops on Data Citation Principles and Practices.**
- **Developing Data Strategies at regional, national and institutional levels.**
  - Collaborating with Polish Science Ministry on Data Policy development
  - Collaborating with CODATA Kenya, JKUAT, on Data Policy development and data strategy.

OPEN SCIENCE PLATFORM

Open Research Data: Implications for Science and Society

# International Workshop on Open Data for Science and Sustainability in Developing Countries

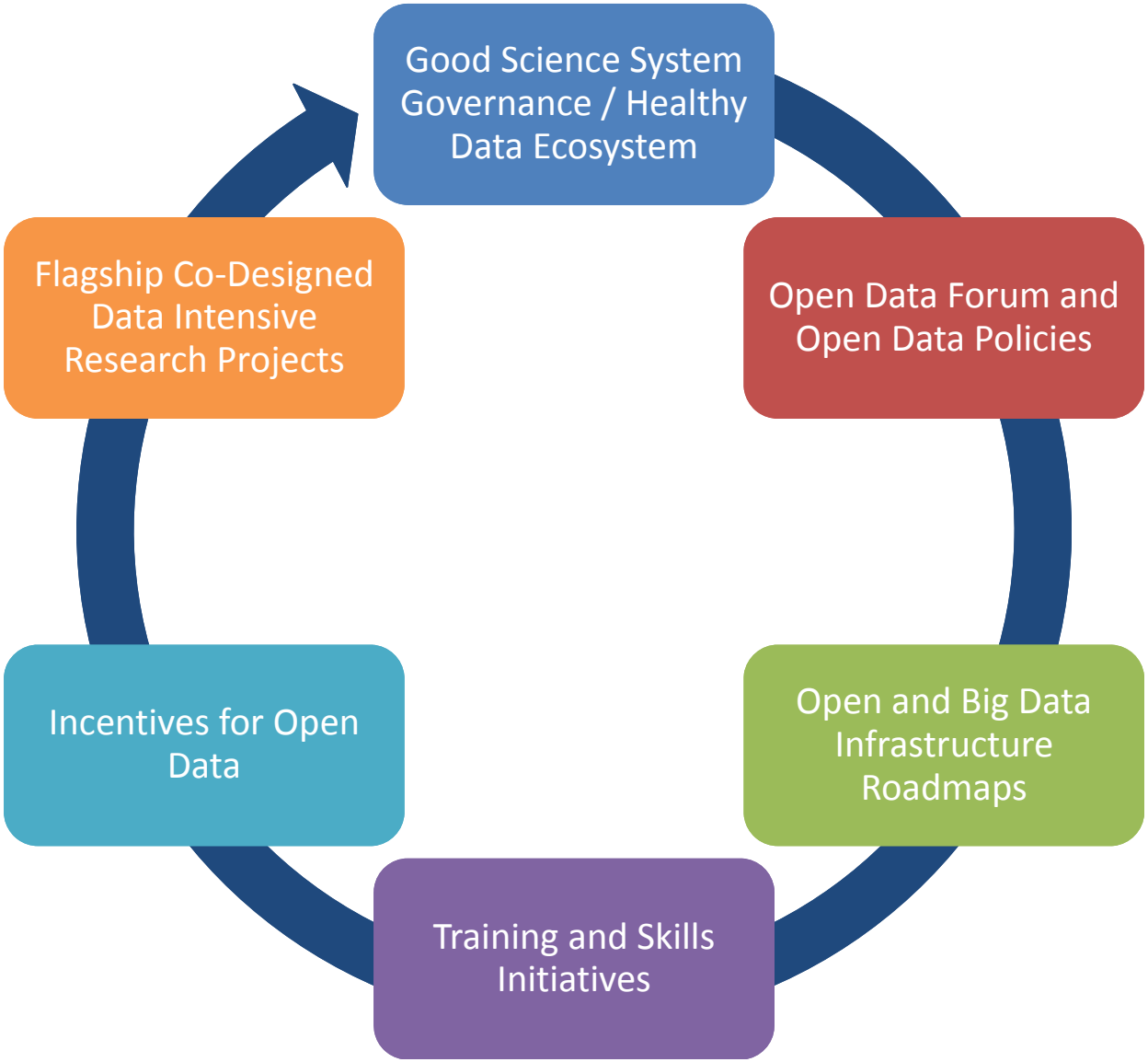
- Strong endorsement for the workshop from Kenyan Cabinet Secretary and from local universities and research institutes.
- Cabinet Secretary Dr. Fred Matiang'i: called on CODATA and other international organisations to 'become more visible in education and capacity-building, by developing science and educational programs and activities that focus on data and information' in developing countries.
- Announced data centre to be established at Jomo Kenyatta University of Agriculture and Technology.
- **'JKUAT has now established an ICT Centre of Excellence and Open Data (iCEOD) that was part of the Nairobi-CODATA conference recommendation'**
- Working with CODATA on data management policies and development of iCEOD: http://www.codata.org/membership/national-members/kenya

# ‘Science International’ and Open Science Capacity Building Initiative

- **Science International**: conceived as an annual summit for international science organisations: inc. ICSU, TWAS, IAP, ISSC.
- Provide a unified, international voice for science.
- First edition will be 7-9 Dec 2015, Pretoria, South Africa.
- Coincides with G77 Science Meeting.
- **CODATA leading on the document: Science International Accord on Big Data / Open Data**

- Will launch a broader international **Capacity Building Initiative in Big Data / Open Data**.
- Support from Department of Science and Technology in South Africa.
- Holistic ‘science systems’ approach: policies, procedures, incentives, data infrastructure, scholarly communications, skills and training.
- Research data science summer schools an essential component of this.

# Big Data / Open Data Capacity Mobilising Intiative



# Data Policies: Data Citation

If publications are the stars and planets of the scientific universe, data are the 'dark matter' – influential but largely unobserved in our mapping process

## Task Group on Data Citation Principles and Practices

Out of Cite, Out of Mind

http://bit.ly/out_of_cite

Joint Declaration of Data Citation Principles:

https://www.force11.org/datacitation

Background and Developments:

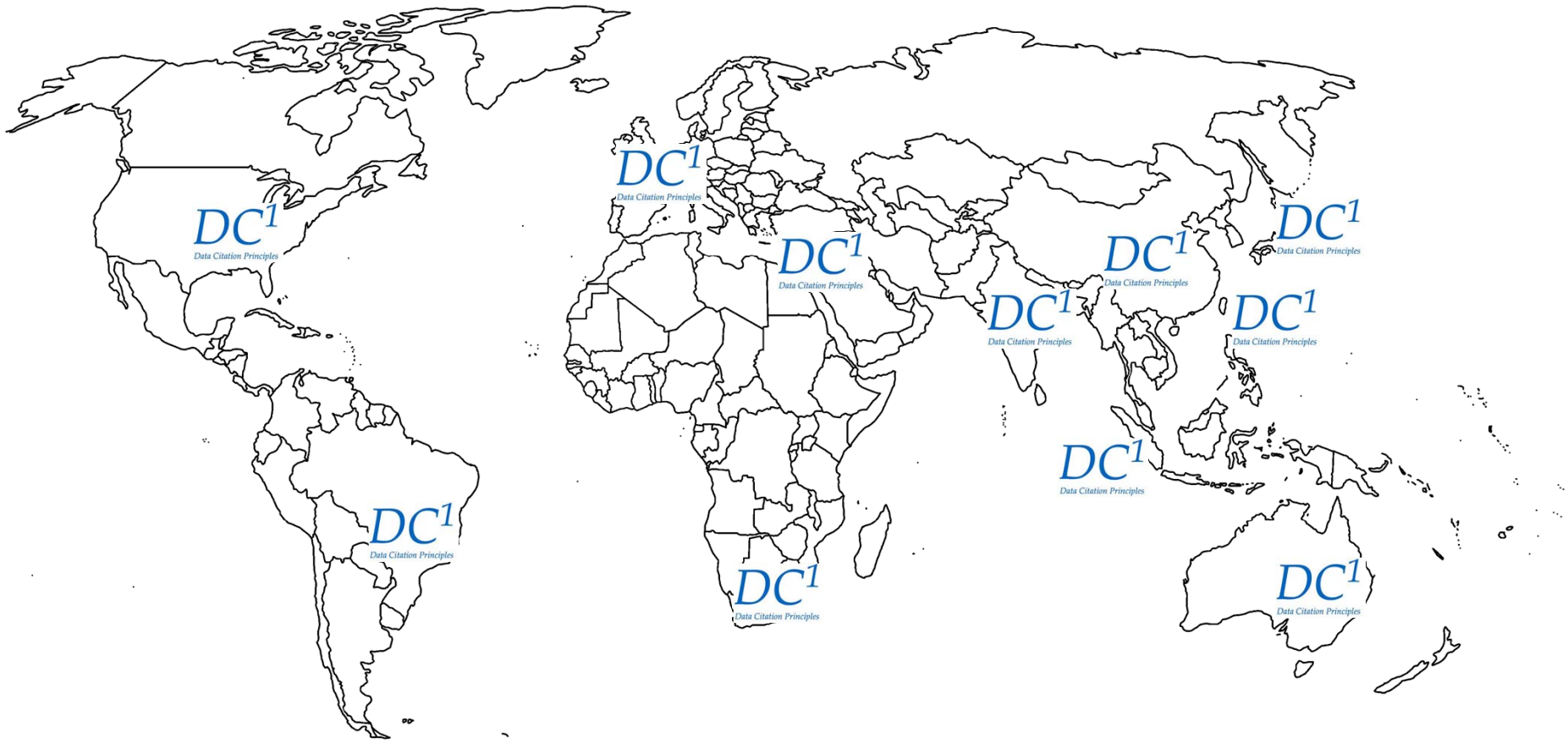
http://bit.ly/data_citation_principles

DC1

Data Citation Principles

# Data Citation: From Principles to Practice

- CODATA Task Group on Data Citation 'Data Citation: From Principles to Practice, A Focus on the Research Policy and Funding Community': http://www.codata.org/task-groups/data-citation-standards-and-practices
- Organising an international series of implementation and adoption workshops.
- **Promote the implementation of data citation principles in the research policy and funding communities throughout the world.**
- **Stakeholders** include: government, funders, research performing institutions, research administrators, research librarians, researchers, learned societies, publishers, data archives, journal editors …
  - What is the policy environment for data citation?
  - What are current attitudes to data citation?
  - What infrastructure currently exists to support data citation?
  - What specific plans for implementation were identified?




DC1
Data Citation Principles

**We are taking Data Citation workshops on a world tour!**

China… then Australia, Japan, India and South Africa. Plus: USA, Taipei, Korea, Indonesia, Brasil, EC, France, Isreal…

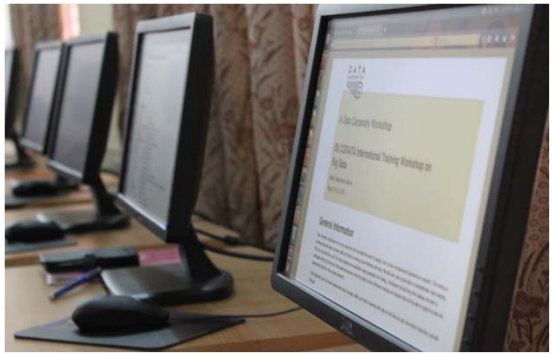
# CODATA and Data Science

## Capacity Building: Training

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

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

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



# Research Data Science Summer Schools

**CODATA-RDA Research Data Science Science Summer Schools will:**

- address a recognised need for Research Data Science skills across disciplines;
- follow an accredited curriculum;
- provide a pathway from a broad introductory course for all researchers (Vanilla) through more advanced and specialised courses (Flavours and Toppings);
- be reproducible: all materials will be online with Open licences;
- be scalable: emphasis will be placed on Training New Teachers (TNT) and building sustainable partnerships;
- **pay particular attention to the needs of young researchers in LMICs.**

# Learning Outcomes (Introductory / Vanilla)

**Learning Outcomes**

1. To understand the importance of data sharing and open science.
2. To understand the importance of annotation and metadata.
3. To have an introductory understanding of developing software for research.
4. To understand how to query and modify an SQL database and the principles behind relational databases.
5. To understand basic principles of statistics, how to deploy machine learning techniques and understand their limitations.
6. To be able to generate simple visualisations of data in an informative fashion.
7. To understand the basic taxonomy of cloud computing.
8. To be able to launch and deploy a Virtual Machine on a Cloud Computing platform.
9. To be able to use a batch submission process.

# Research Data Science Summer Schools

**First Vanilla School, 1-12 August, ICTP, Trieste**

- ICTP providing accommodation and meals for up to 120 students.
- Total 30K euros funding for student travel committed by ICTP, TWAS and CODATA.
- **Priority for students from LMICs.**
- Other sponsors and funders welcome!
- Explore regional schools with TWAS and ICSU regional offices.

software carpentry

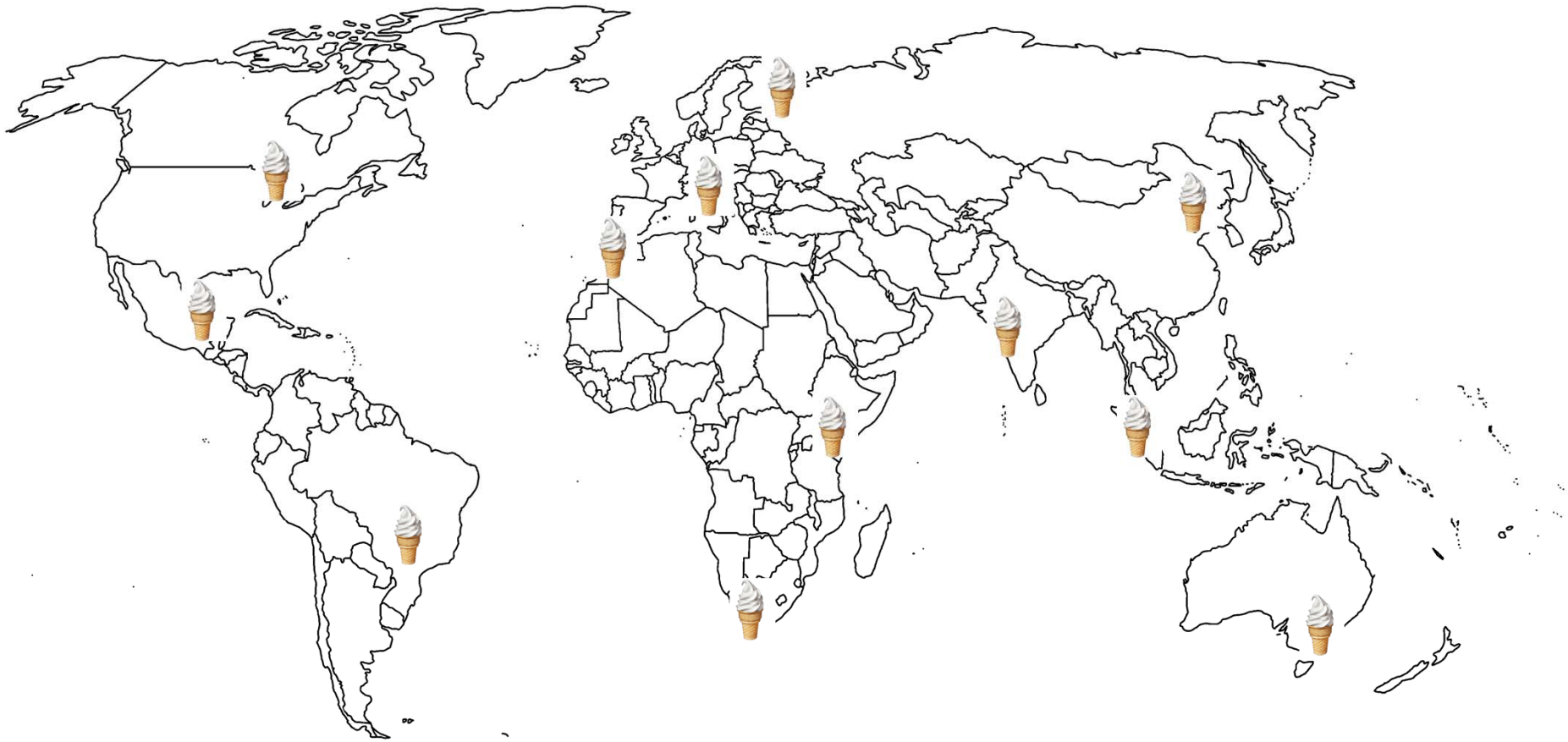
DATA CARPENTRY
MAKING DATA SCIENCE MORE EFFICIENT

DCC

GEO

ICTP The Abdus Salam International Centre for Theoretical Physics

twas

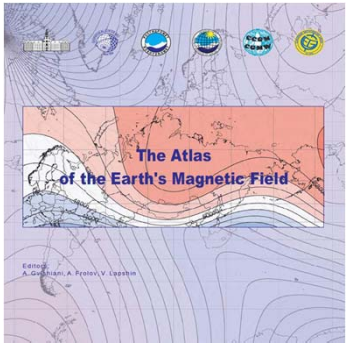


# We are designing Vanilla for a world tour!

Italy… then South Africa, Mexico, Brasil, USA, Kenya, India, Australia, China, Russia, Indonesia.

# Challenges in Data Science

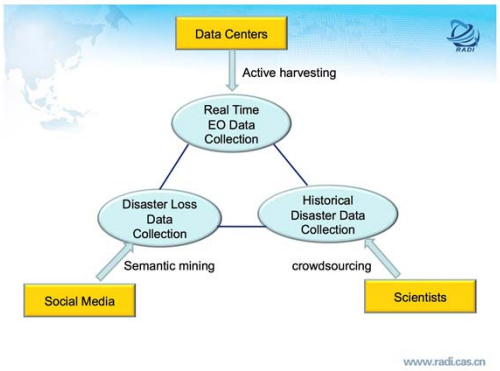
## Earth-Space Science Interoperability




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

Conferences in Sochi 2016 and St. Petersburg 2017

## LOD Global Disaster Data

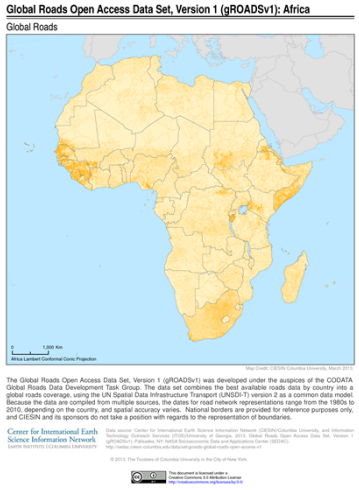



Preparing White Paper on Use of LOD for Disaster Data:
http://www.codata.org/task-groups/linked-open-data-for-global-disaster-risk-research

## Data at Risk

Memory of the World

Developing 'Principle Guidelines for Data at Risk':
http://bit.ly/DAR-guidelines



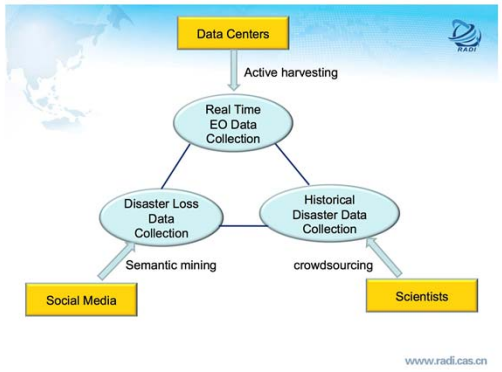

## Global Roads Data

Review of Global Roads Data Development Methodologies:
http://bit.ly/globalroads-methods

# Integrating Geospatial Data on the Web

**Coverages and Earth Observation in Linked Data (CEO-LD)**

- Funded project led by CODATA.
- UK-China collaboration to implement/validate OGC-W3C standards: http://www.w3.org/2015/ceo-ld/
- Builds on collaborations with RADI, Institute of Remote Sensing and Digital Earth of the Chinese Academy of Sciences and CODATA TG on LOD for Disaster Research.
- Project runs Sept 2015 – May 2016.
- Output will be a draft standard on coverages in Linked Data.



# Data Strategies for Interdisciplinary, Integrated Research into Smart Cities

Are the data available?

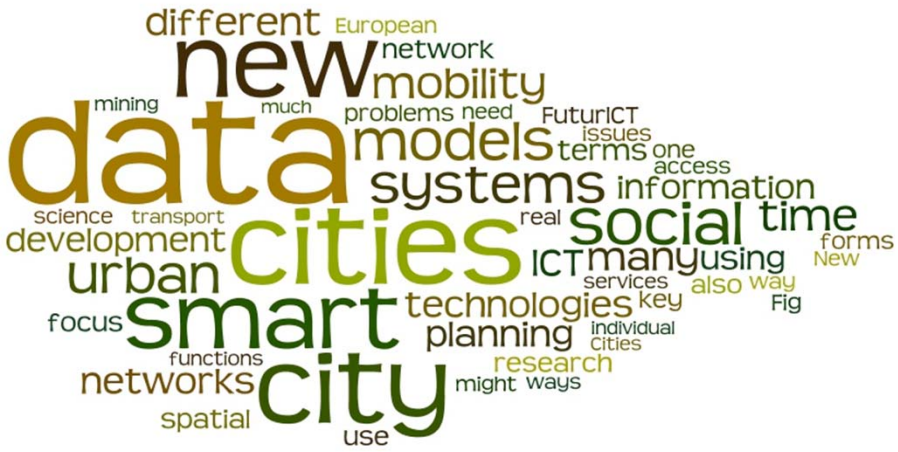
What policies are in place?

How may the data be integrated, analysed?

What are the ethical implications?

What skills are required?

- ISSC World Social Science Fellows Seminar on Big Data in an Urban Context, ICSU Programme on Urban Health and Wellbeing, Xiamen, China, 30 Nov-4 Dec: http://www.codata.org/news/50/62/Call-for-Applications-World-Social-Science-Fellows-Big-Data-in-Urban-Contexts
- **Follow with series of CODATA workshops on Big Data and Smart Cities**




# What do we mean by data science? What are the pressing issues?


- Strategic mission to address 'frontier issues in data science'.
- What do we mean by 'data science'?
- **Promote the study of data in research (data science) as a discipline.**
- Data science as the systematic study of data in research relates to all areas of scholarly inquiry.
- Contemporary research – particularly when addressing the most significant, transdisciplinary research challenges – increasingly depends on a range of issues relating to data. These issues 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 systematic study of these things, of data in research, can usefully be called 'Research Data Science.


- **Key Issues: 1) reproducibility in research; 2) statistical, epistemological and ethical issues in Big Data; 3) coordination of discipline vocabularies (with ISUs); 4) data integration and data strategies for international research programmes (e.g. Urban Health and Wellbeing)**

SciDataCon 2014
International Conference on
Data Sharing and Integration for Global Sustainability

ICSU
WORLD DATA SYSTEM

# Data Science Journal


- CODATA has relaunched the Data Science Journal with Ubiquity Press.
- New Editor-in-Chief, Sarah Callaghan, data scientist with British Atmospheric Data Centre and expert on many data issues.
- http://datascience.codata.org/
- *Dedicated to the advancement of data science and its application in policies, practices and management as Open Data to ensure that data are used in the most effective and efficient way in promoting knowledge and learning.*

- Ubiquity Press in a publisher with a strong interest in data and software publication and a strong OA ethos.
- Open Scholarship – Researcher Led Publishing.
- Transparent on APC, waivers in case of need.

]u[ ubiquity press
open scholarship

# International Data Week 2016

- Jointly organised by CODATA, RDA and WDS.
- Combines 1) two-day research conference, 2) a data forum focusing on policy discussion, 'hacktivism', intersections with open public data and data science, 3) RDA Plenary 8.
- **September or October 2016, USA…**


Organized by:

# CODATA Opportunities

- **CODATA: a platform for policy development, research into 'data science' and strategic capacity mobilisation.**
- Engage with the **Data Science Journal** as a forum for research into data issues: http://datascience.codata.org/
- Come to **International Data Week** 2016: http://www.internationaldataweek.org/
- Work with us on the **CODATA-RDA Research Data Science Summer Schools:** http://www.codata.org/working-groups/research-data-science-summer-schools; http://indico.ictp.it/event/7658/
- **Sign up to the Data Citation Principles**: https://www.force11.org/group/joint-declaration-data-citation-principles-final
- Propose a **CODATA Task Group** on a Frontier Data Issue.
- **Forthcoming reports:**
  - CODATA-RDA Legal Interoperability: http://bit.ly/legal-interop-IGs
  - RDA-WDS Income Streams: http://bit.ly/income-streams-draft-P6
  - The Value of Open Data Sharing: White Paper for GEO, the Group on Earth Observations

DC1 Data Citation Principles

INTERNATIONAL DATA WEEK 2016
WWW.INTERNATIONALDATAWEEK.ORG
Organized by:


ICSU

International Council for Science

# Thank you for your attention!

Slide credits: Christine Borgman, Geoffrey Boulton, Sarah Callaghan

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 (ICSU Committee on Data for Science and Technology), 5 rue Auguste Vacquerie, 75016 Paris, FRANCE