CODATA:
Past - Present - Future

Geoffrey Boulton

General Assembly
Denver, USA
11 September 2016

CODATA decade by decade

60s-70s - basics
- Fundamental constants
- Key values for thermodynamics
- Computer use
- Compendium of numerical data projects

70s – higher level issues
- Chemical kinetics
- Numerical data by disciplines
- Data dissemination
- Handllng of experimental data
- Thermodynamic data systemization

80s – addressing disciplinary specifics
- Data in the biosciences & geosciences
- Directory of protein & nucleic acid sequencing
- Access to biological data banks
- Biodiversity international standards
- Materials database standards
- Chemical thermodynamic tables

90s – managing data
- Databases for Experimentation
- Electronic Publishing
- Data Access Commission
- Global Plant Checklist Network
- Data/Information and Visualization
- Mathematical Methodologies for Data
- Handling and Knowledge Extraction
- Data Quality and Database Compatibility

- **2000s – addressing the digital revolution**
- Data Citation: use of DOIs for data;
- International Polar Year Data
- International Symposium on Open Access and the Public Domain
- Digital Data and Information in Science;
- OECD Data Policy Recommendations.
- GEO Data Sharing Principles

2010s - open data, open science
- Policy and practice
- Frontiers of scientific use of data
- Capacity building

The digital revolution

storage – analysis – communication

Global information storage capacity
In optimally compressed bytes

2007
19 Exabytes
2000
1993
1986
Analogue Storage
Digital Storage
280 Exabytes
2014 - 4000 Exabytes
2016 - 20,000 Exabytes
Explosion of the Digital revolution

The technological bases for open science
if we choose to use them!

Based on:
http://www.martinhilbert.net/WorldOnfoCapacity.html

1 Exabyte=10^{18} bytes

Exploiting the Data Revolution:

The CODATA strategy

April 2014

The ICSU Committee on Data for Science and Technology

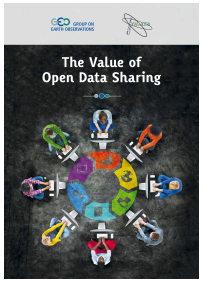
Executive Committee 2014-2016

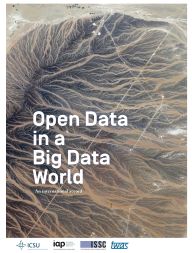
CODATA – Strategic Priorities

Principles, Policies and Practice

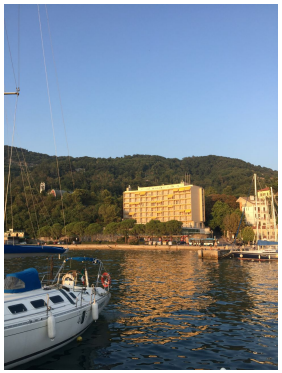
DC1 Data Citation Principles

ICSU International Council for Science

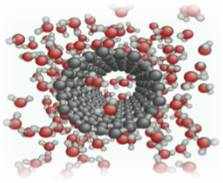


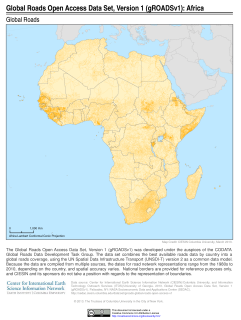


Capacity Building



Frontiers of Data Science





ubiquity press open scholarship

Data Science Journal

SciDataCon 2016, 11-13 Sept, Denver, CO.

INTERNATIONAL DATA WEEK 2016
WWW.INTERNATIONALDATAWEEK.ORG

Organized by:

Executive Committee Anaysis of TG Proposals

Principles, Policies and Practice

- Citizen Science and the Validation, Curation, and Management of Crowdsourced Data
- Globalising Open Science & Access to Research Software
- Practice and Impact of Digital Data Citation
- Data at Risk*

Capacity Building

- Agriculture Data, Knowledge for Learning and Innovation
- Building Foundational Training in Research Data Science
- Open Data Strategy in Africa
- Preservation of and Access to Scientific and Technical Data in Developing Countries *

Frontiers of Data Science

- Coordinating Data Standards amongst Scientific Unions
- Materials Data in IoT and Big-Data Era
- Advancing Informatics for Microbiology
- Earth and Space Science Data Interoperability*
- Linked Open Data for Global Disaster Risk Research
- Science and Management of Physical Objects in the Digital Era

.

Strongly endorse

Endorse

* Endorse subject to conditions

Do not endorse

CODATA – Emergent issues

Principles, Policies and Practice

- Standards for data discovery, use & interchange (TG proposal – possible Commission)
- Citizen science & crowd-sourced data (TG proposal)
- Citing and managing digital objects
- A data system ecology for decision makers

Frontiers of Data Science

- Machine learning
- Internet of things
- Standards for reproducibility
- Linked semantic data

Capacity Building

Open Data Platforms

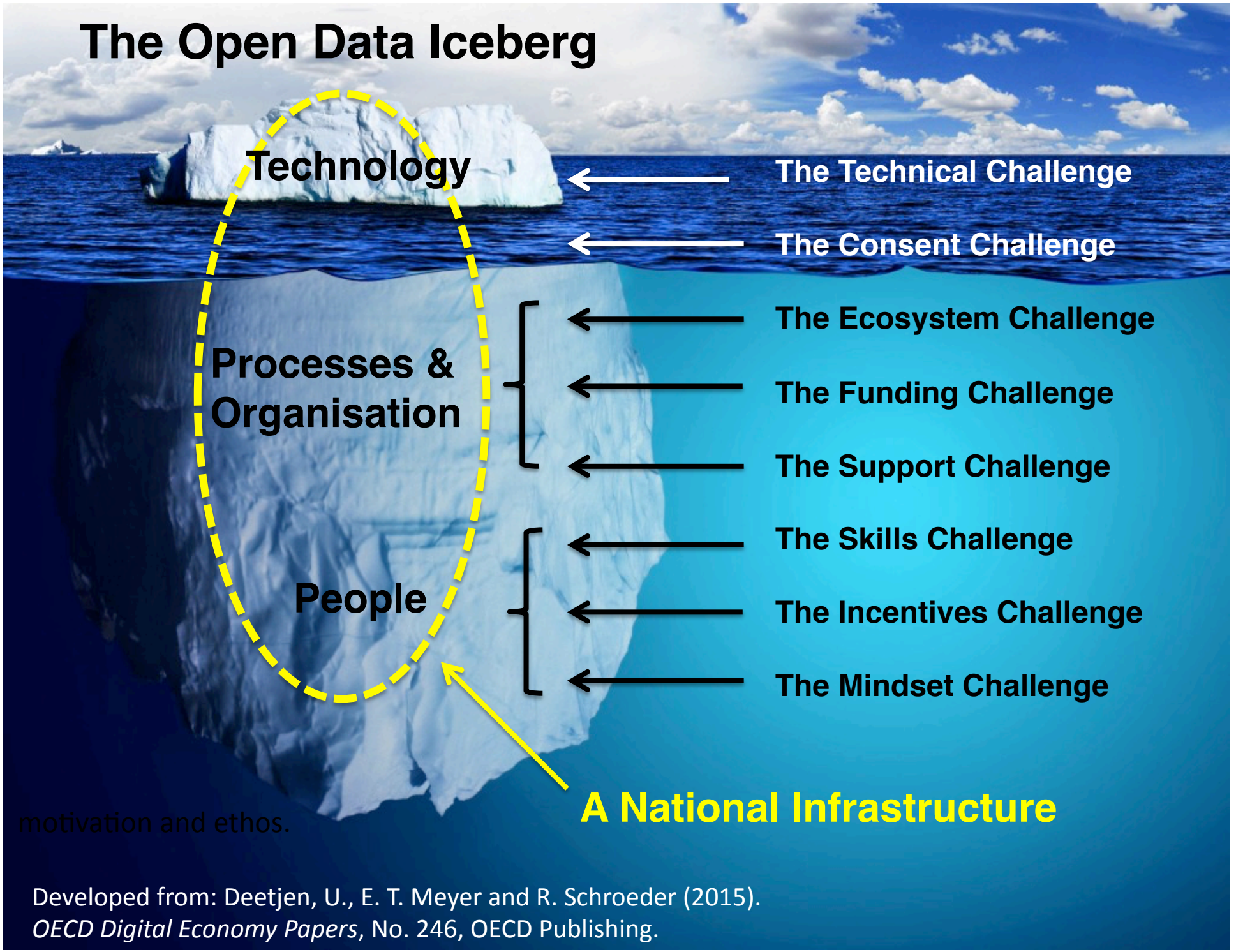
- Africa
- Latin America & Caribbean

(Note TG proposals)

Governance

Focus on:

- Relations with National Committees
- Relations with Scientific Unions
- Individual & Corporate Membership
- Serving the whole research enterprise



Developed from: Deetjen, U., E. T. Meyer and R. Schroeder (2015). *OECD Digital Economy Papers*, No. 246, OECD Publishing.

An Open Data Platform for Africa

Principles
– the rules of the game (the Accord)

Policies
– governmental priorities

Processes
– stakeholders applying the rules to support open data

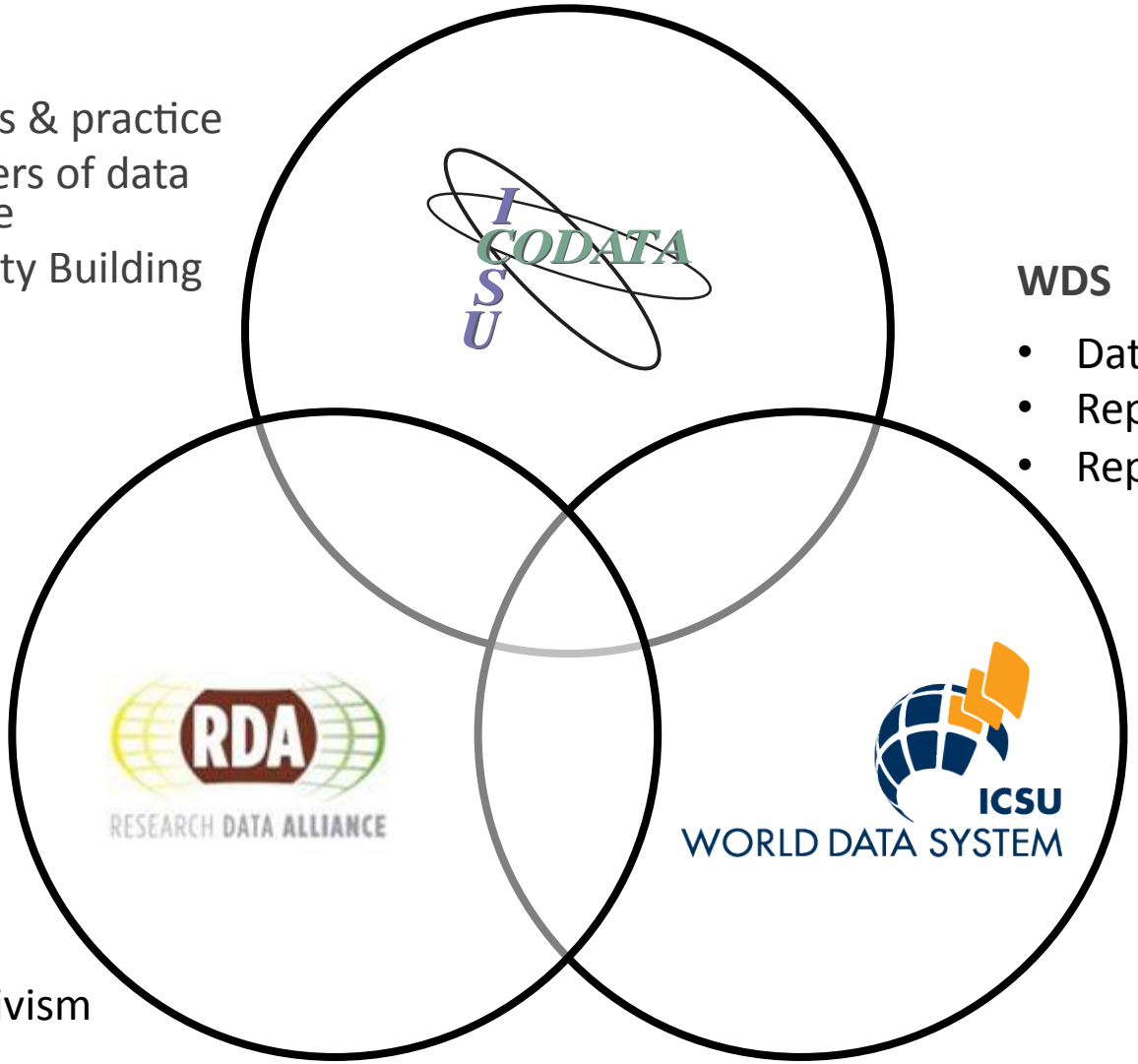
Planning – Coordination

- Hardware procurement
- Software tools
- Repository/database ecology
- Capacity development
- Support structures,
- Management
- Incentives
- Skill development

International Research Data Collaboration

CODATA

- Policies & practice
- Frontiers of data science
- Capacity Building



WDS

- Data stewardship
- Repository Standards
- Repository Access

RDA

- Data sharing
- Interoperability
- Community activism

Summary of Executive Committee Recommendations in an Ordered Table

1	Coordinating Data Standards amongst Scientific Unions	Closely aligned with CODATA strategy. **STRONGLY ENDORSE**
2	Building Foundational Training in Research Data Science	Closely aligned with CODATA strategy. **STRONGLY ENDORSE**
3	Practice and Impact of Digital Data Citation	Closely aligned with CODATA strategy. **STRONGLY ENDORSE**
4	Linked Open Data for Global Disaster Risk Research (LODGD)	Closely aligned with CODATA strategy. **STRONGLY ENDORSE**
5	Agriculture Data, Knowledge for Learning and Innovation	Valuable new proposal, aligned with CODATA strategy. **ENDORSE**
6	Citizen Science and the Validation, Curation, and Management of Crowdsourced Data (with WDS)	Valuable new proposal. **ENDORSE**
7	Preservation of and Access to Scientific and Technical Data in Developing Countries (PASTD)	Can be more closely aligned to CODATA strategy and achieve greater impact **ENDORSE SUBJECT TO CONDITIONS**
8	Earth and Space Science Data Interoperability (ESSDI)	Can be more closely aligned to CODATA strategy and achieve greater impact **ENDORSE SUBJECT TO CONDITIONS**
9	Data at Risk (DAR)	Can be more closely aligned to CODATA strategy and achieve greater impact **ENDORSE SUBJECT TO CONDITIONS**
10	Globalising Open Science & Access to Research Software (with GYA)	Valuable collaboration, but workplan not strong enough **DO NOT ENDORSE**
11	Open Data Strategy in Africa	Recommend incorporate with other activities. **DO NOT ENDORSE**
12	Science and Management of Physical Objects in the Digital Era	TG has not delivered enough in the last term **DO NOT ENDORSE**
13	Advancing Informatics for Microbiology (TG-AIM)	TG has not delivered enough in the last term **DO NOT ENDORSE**
14	Materials Data in IoT and Big-Data Era	Proposal incomplete and workplan not strong enough **DO NOT ENDORSE**
15	Protecting Historical Data	Recommend incorporate with other activities. **DO NOT ENDORSE**