IG Data for Development
RDA 10th Plenary Meeting
19 September 2017

Open Science and FAIR Data: Essential Components for Sustainable Development in a Big Data World

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

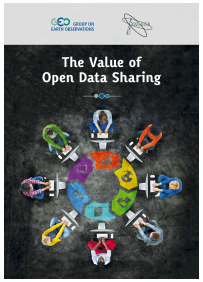
CODATA Prospectus:

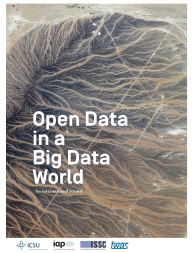
https://doi.org/10.5281/zenodo.165830

Principles, Policies and Practice

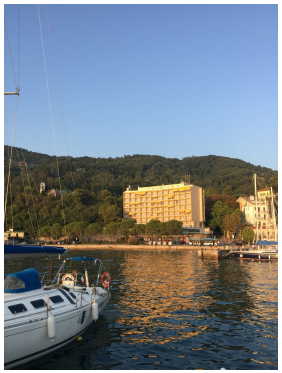
DC1 Data Citation Principles

INTERNATIONAL COUNCIL FOR SCIENCE

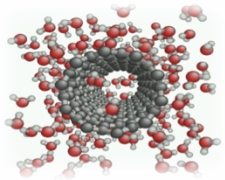


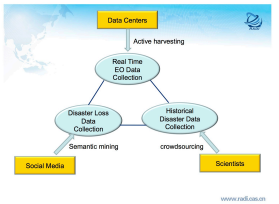


Capacity Building



Frontiers of Data Science





ubiquity press open scholarship

Data Science Journal

INTERNATIONAL DATA WEEK 2016
WWW.INTERNATIONALDATAWEEK.ORG

CODATA 2017, Saint Petersburg 8-13 Oct 2017

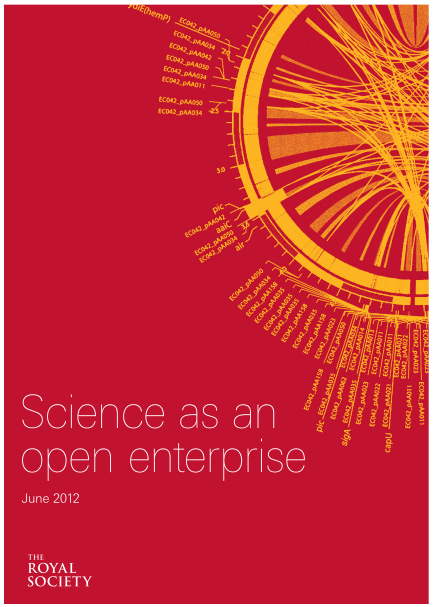
http://codata2017.gcras.ru/

Why Open Science / FAIR Research Data?

- **Open data foster innovation and accelerate scientific discovery through reuse of data within and outside the academic system.**
  - Research data produced by publicly funded research are a public asset, a public good.
  - There is a strong interface between many areas of research and application by professionals, practitioners.
  - **A World that Counts:** call for a world that knows itself better through Open Data, and uses more diverse data sources to monitor and achieve the SDGs.
  - **Open Data and Agriculture:** large number of case studies of interface between research and practitioners.
  - CODATA Report for GEO, 'The Value of Open Data Sharing', summarising benefits including the economic, research, educational, societal.

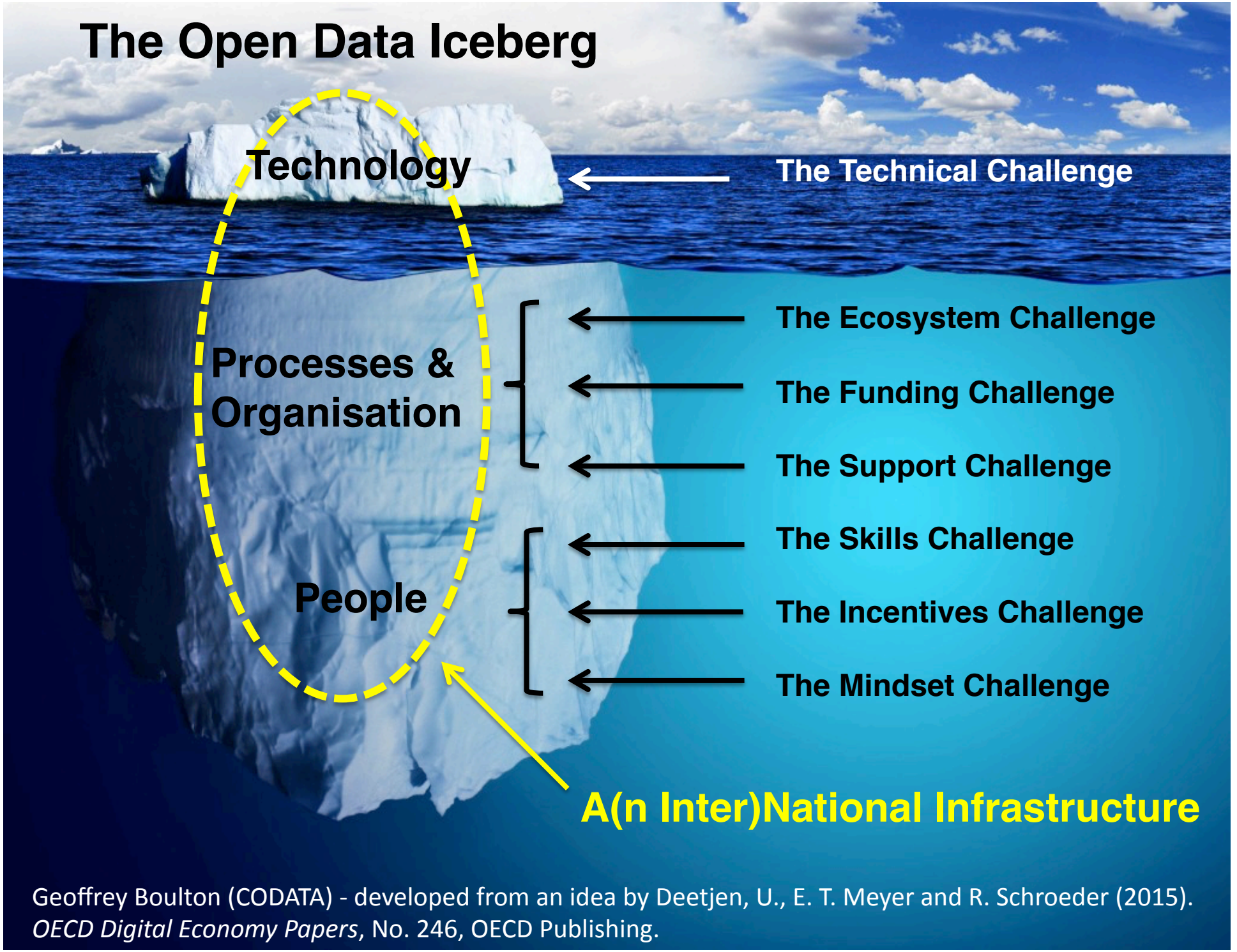
Africa, Data and Open Science

- 21st century is the century of data.
- Data skills and infrastructure will be essential for economic advancement and for sustainable development.
- We need to create **a 'world that counts'** that gathers data and uses data to understand itself.
- Open data is essential to increase impact of research and translation for practitioners.
- African governments, research and education systems and universities have an interest in developing data skills and infrastructure.
- African universities have an essential role to play as educators of a data savvy generation and as the stewards of the data created by African research.
- The data from many research projects conducted in Africa is not looked after in African institutions.
- African institutions need to present their research outputs, including data, as a shop window and a record of their activities, achievements, impact.



The Case for Open Data in a Big Data World

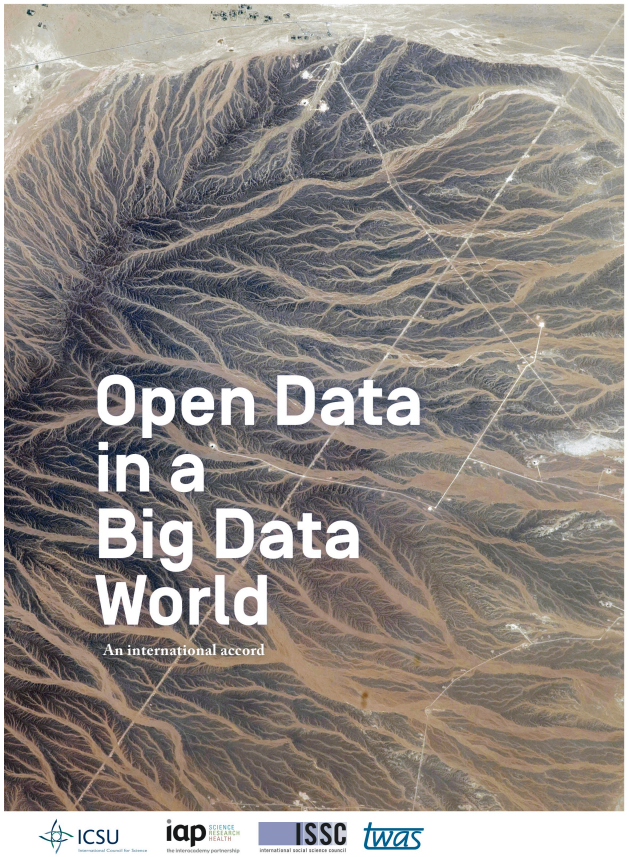
- **Science International Accord on Open Data in a Big Data World: http://www.science-international.org/**
- Supported by four major international science organisations.
- Presents a powerful case that the profound transformations mean that data should be:
  - Open by default
  - Intelligently open, 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**



Geoffrey Boulton (CODATA) - developed from an idea by Deetjen, U., E. T. Meyer and R. Schroeder (2015). *OECD Digital Economy Papers*, No. 246, OECD Publishing.

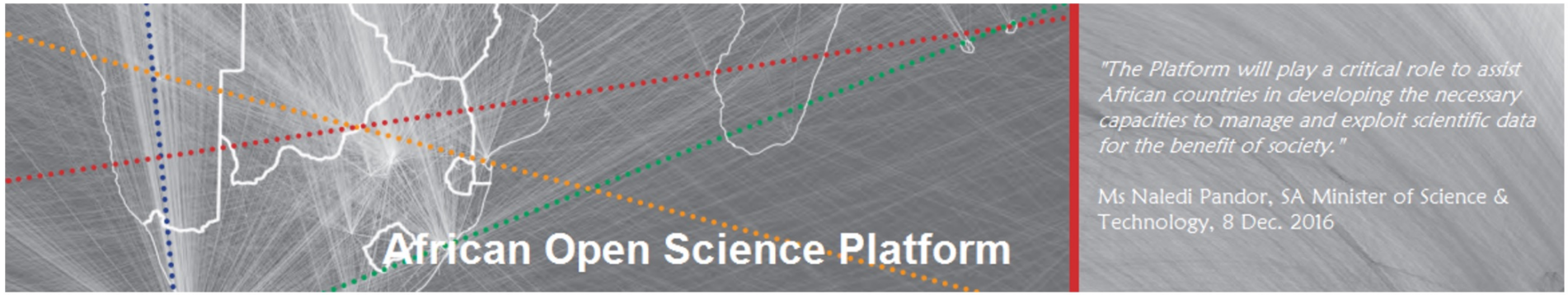
Framework for National and Institutional Data Strategies

- National / Institutional Open and FAIR Data Strategy.
- Open data **policies and guidance** at national and institutional level.
- Clarify the **boundaries of open** (particularly privacy, IPR).
- Mechanisms (infrastructure and policy) to ensure **concurrent publication of data as research output**.
- Data 'publication' and citations of data included in **assessment of research contribution**.
- Promotion of **data skills** (researchers and data stewards).
- Development **of institutional infrastructure** for research collaboration and data stewardship/RDM.
- **Collaborative infrastructures** for certain research disciplines, nationally, regionally to pool expertise and lower costs.



African Open Science Platform

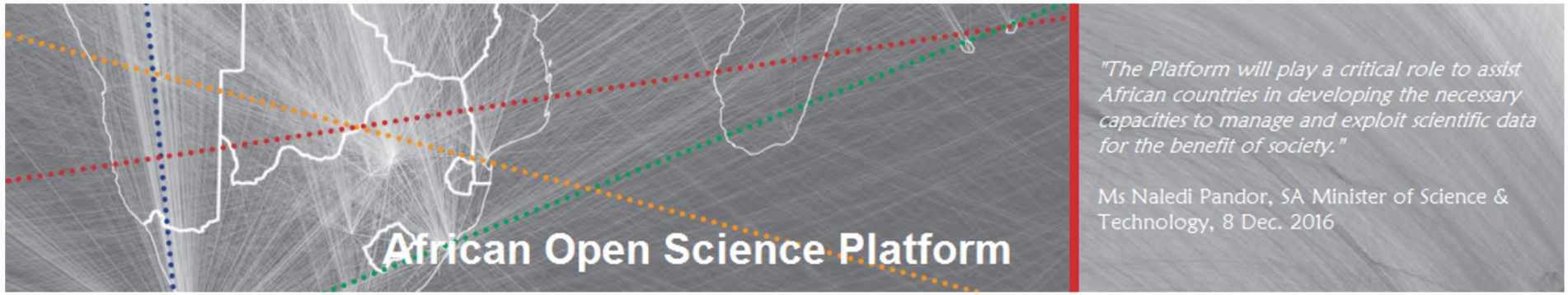
- Approach outlined in the Accord on Open Data in a Big Data World: Open Science is fundamental to 21st Century discovery and needs to be supported.
- Development of coordinating 'Platform' approach to support development of Open Science in Africa.
- Pilot project a direct result of the Accord. **Funding from DST / NRF for African-wide initiative.**
- Three year pilot starting in Dec 2016 with scoping workshop.
- ASSAf, South African Academy of Sciences, playing the key implementation role.



African Open Science Platform

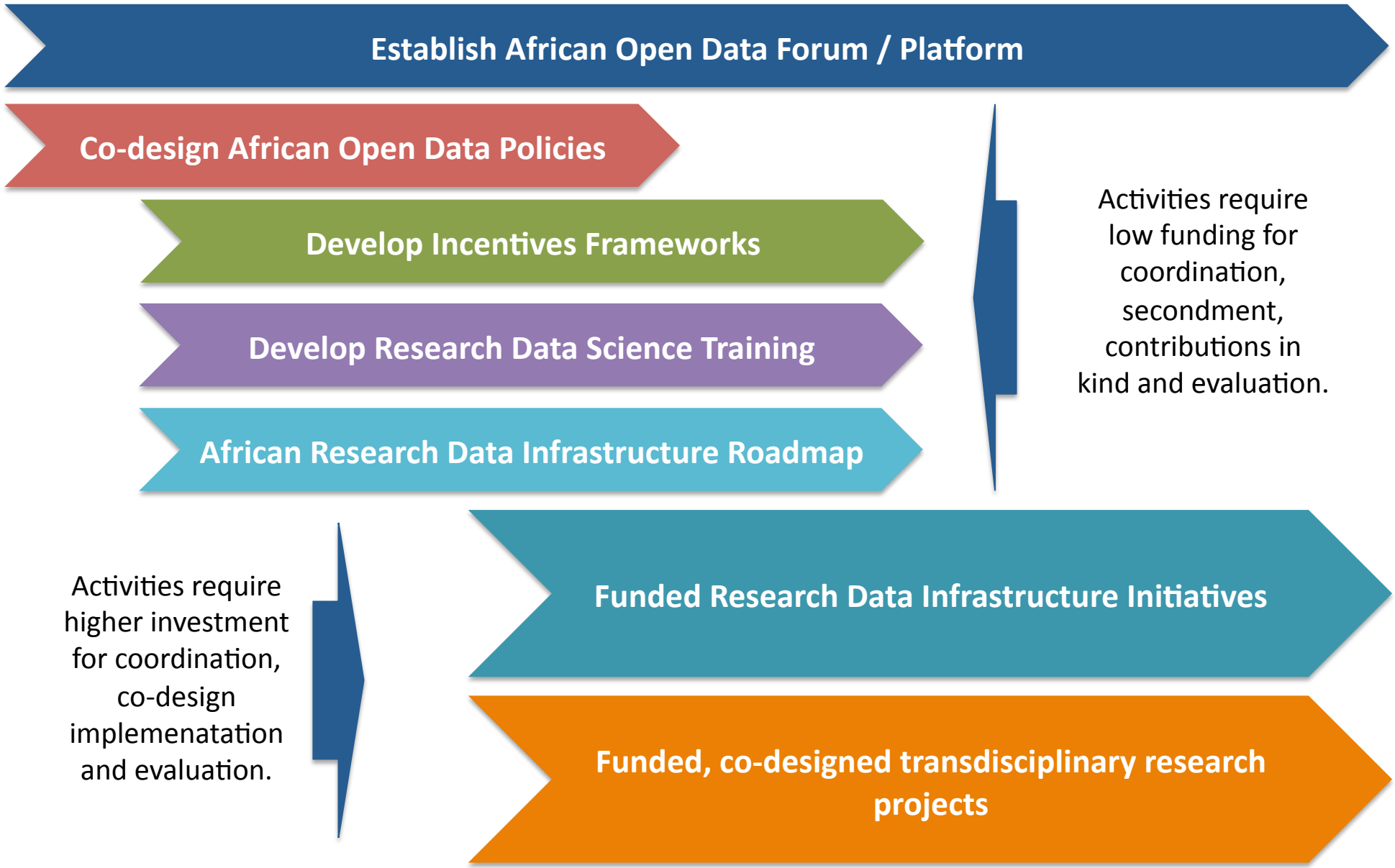
African Open Science Platform

- High level advisory council, chaired by Khotso Mokhele
  - Ismail Serageldin (founding director of Library of Alexandria)
  - Ogunlade Davidson (former minister former Minister of Electricity and Water Resources, Sierra Leone)
  - Heide Hackmann (Executive Director, International Council for Science)
- 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 and others.
- Meetings of National Data Fora in Madagascar (September) and Botswana (October).

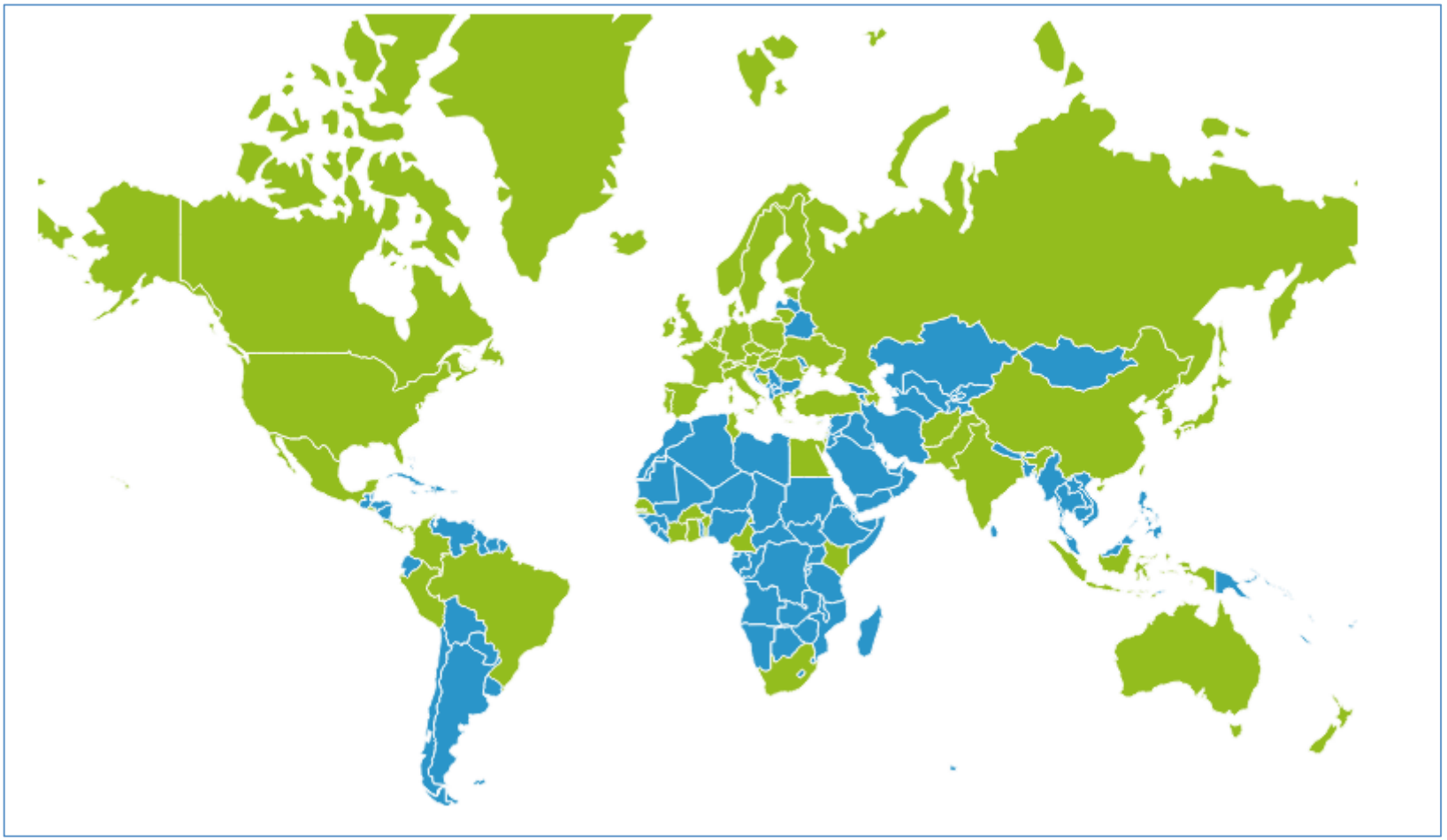


African Open Science Platform

African Open Science Platform
Pilot Project Workpackages



Where should research data go?



Country coverage in Re3Data.org (registry of data repositories.

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

DCC

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**

#DataTrieste

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, organised 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**

The Abdus Salam International Centre for Theoretical Physics

twas

CODATA 2017: Global Challenges and Data Driven Science

http://codata2017.gcras.ru/

http://conference.codata.org/2017/

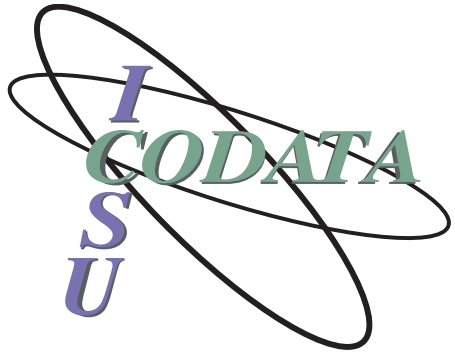
Major conference themes:

1. Achievements in Data Driven Science, in all research disciplines
2. Earth Observations Data and the Earth's System
3. Data and Disaster Risk Research
4. Data Driven and Sustainable Cities
5. Big Data in Scientific and Commercial Sectors
6. Data Analysis, Event Recognition and Applications
7. National and International Data Services
8. Research Data Services in Universities
9. Coordination of Data Standards and Interoperability
10. FAIR Data and the Limits of Open Data
11. Metrology, Reference Data and Monitoring Data

- **Saint-Petersburg, 8-12 October**
- http://codata2017.gcras.ru/
- Special collectcin in Data Science Journal

- **International Data Week 2018, in Africa**

INTERNATIONAL COUNCIL FOR SCIENCE



Thank you for your attention!

Slide credits: in this presentation, I have reused, adapted and credited some slides from presentations by Geoffrey Boulton, Danny Kingsley, Michael Lautenschlager, Li Guoqing, Marshall Ma and the Dryad Data Repository

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

Why Open Science / FAIR Data?

- **Good scientific practice depends on communicating the evidence.**
  - Open research data are essential for reproducibility, self-correction.
  - Academic publishing has not kept up with age of digital data.
  - Danger of an replication / evidence / credibility gap.
  - Boulton: to fail to communicate the data that supports scientific assertions is malpractice
- **Open data practices have transformed certain areas of research.**
  - Genomics and related biomedical sciences; crystallography; astronomy; areas of earth systems science; various disciplines using remote sensing data…
  - **FAIR data helps use of data at scale, by machines, harnessing technological potential.**
  - Research data often have considerable potential for reuse, reinterpretation, use in different studies.
- **Open data foster innovation and accelerate scientific discovery through reuse of data within and outside the academic system.**
  - Research data produced by publicly funded research are a public asset.