Session: ‘Role of Geospatial Information in attaining Sustainable Development Goals’
Geospatial World Forum: Disruptive Business Models
26 May 2016

The Value of Open Data Sharing: Open Data, the SDGs and the economics of data infrastructure

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

Benefits of sharing geospatial data…

- What are the benefits of data sharing?
  - CODATA Report for GEO
  - Contribution towards realisation of SDGs
- What are the economic benefits of data sharing?
- How can we sustain the data infrastructure and what business models are appropriate for data repositories?

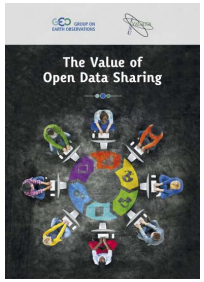
CODATA

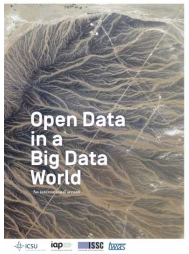
ICSU
International Council for Science

Principles, Policies and Practice

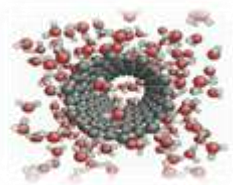
DC1
Data Citation Principles

ICSU
International Council for Science





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

Capacity Building

UNESCO

ICTP
The Abdus Salam International Centre for Theoretical Physics

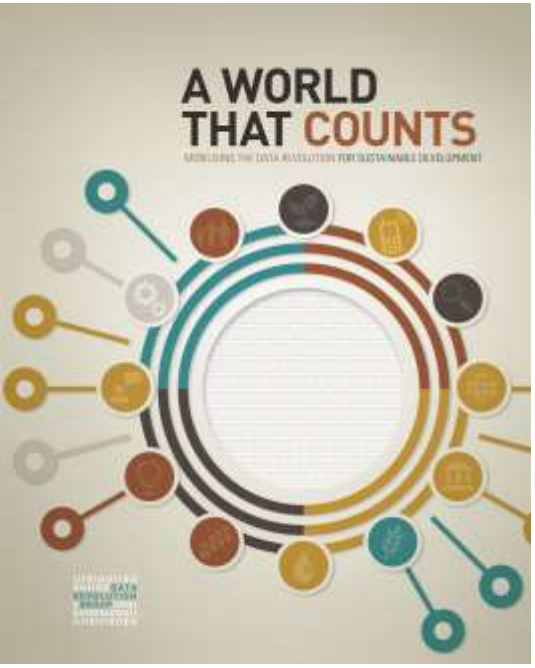
Data Revolution: Science as an Open Enterprise

Royal Society Report: Science as an Open Enterprise

CODATA President Geoffrey Boulton, FRS

- 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 and opportunities for science systems, technical and human.
  - Fundamental methodological issues for reproducibility and transparency.
- **Data for research should be intelligently open: accessible, assessible, intelligible, useable.**
- **Publications and data should be Open and available concurrently.**
- **Report with very significant impact: G8, H2020**

Data Revolution:
A World that Counts!



THIS IS THE REVOLUTION

Indonesia is one of the most social-media dense countries in the world today. Indonesians tweet about a range of topics, including the cost of living. A project by UN Global Pulse, the Indonesian Ministry of National Development Planning and the World Food Programme found public tweets mentioning food prices closely approximate official figures, leading to the development of a technology that extracts daily food prices from public tweets to generate a near real-time food price index. This data mining approach could be adapted to other food items and locations, not just leveraging Twitter but other crowd-sourced and social data sources.

Source: UN Global Pulse (http://www.unglobalpulse.org/nowcasting-food-prices)

- **Creating a world that counts:** Mobilising the Data Revolution for Sustainable Development.
- To meet the new sustainability goals *'there is an urgent need **to mobilise the data revolution for all people and the whole planet in order to monitor progress**, hold governments accountable and foster sustainable development.'*
- *Without immediate action**, gaps between developed and developing countries, between information-rich and information-poor people, and between the private and public sectors will widen**, and risks of harm and abuses of human rights will grow.*
  - Data quality and integrity
  - Data disaggregation (no-one should be invisible)
  - Data timeliness
  - Data transparency and openness
  - Data usability and curation
  - Data protection and privacy
  - Data governance and independence
  - Data resources and capacity
  - Data rights

The Role of Science in the SDGs

- Report from International Council for Science (ICSU) and International Social Science Council (ISSC) on the SDGs and indicators: http://www.icsu.org/publications/reports-and-reviews/review-of-targets-for-the-sustainable-development-goals-the-science-perspective-2015
- 'From a technical perspective there is a need to strengthen Earth Observation and ground-based monitoring and statistical capacities to address critical and persistent data gaps.'
- 'From an institutional perspective the need is to make sure that monitoring, review and reporting mechanisms are integrated into policymaking processes at all levels and that information is effectively used for improving decisions.'
- **Science and geospatial data can play an important role both in the monitoring of SDGs and in the changes which will help towards achieving them.**

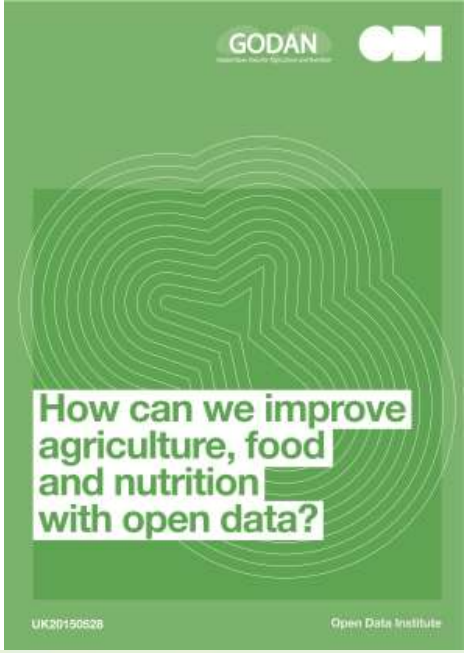
Data Revolution: how can we improve … with open data?

- GODAN-ODI Report: improving agriculture, food and nutrition with open data.
- *'Although the amount of data openly available is constantly increasing,* ***there are still challenges related to data management, licensing, interoperability and exploitation. There is a need to evolve policies, practices and ethics around closed, shared, and open data.****'*
- **Enabling more efficient and effective decision making** > lowers cost of accessing information and underpins tools that farmers themselves can use.
- **Fostering innovation to benefit everyone** > an opportunity that must not be missed for creating new businesses and jobs in 'new data-powered innovation ecosystems'.
- **Driving organisational and sector change through transparency** > open data is essential to understanding complex systems, interventions, targets, change.
- **Availability is not enough** > essential that the data be interoperable and machine-readable.
- Problem oriented and solution-based data strategies.
- Develop infrastructure and human capacity.

Boosting crop yields with a best practice knowledge bank: Plantwise

Plant pests and disease are currently responsible for about 40% of global crop production losses.[19] **Plantwise** helps smallholder farmers in developing countries deal with plant health issues. It aims to increase food security and improve rural livelihoods by reducing crop losses from pests and diseases. It does so by combining global and local open access data from sources such as CABI's databases, research publications and governmental data. It makes the data available and easy to search for via an online platform. Reports of disease from plant clinic operations on the ground are also used to supplement the knowledge bank and notify local partners of pest issues.

In two years the Plantwise knowledge bank has become a vital tool to support plant clinic operations in 33 countries. Over 600,000 farmers from 198 countries have visited the knowledge bank including over 9,000 factsheets to access critical agricultural data on crop pest prevalence and best practices to help manage and prevent potential crop loss from pests and diseases.[20]



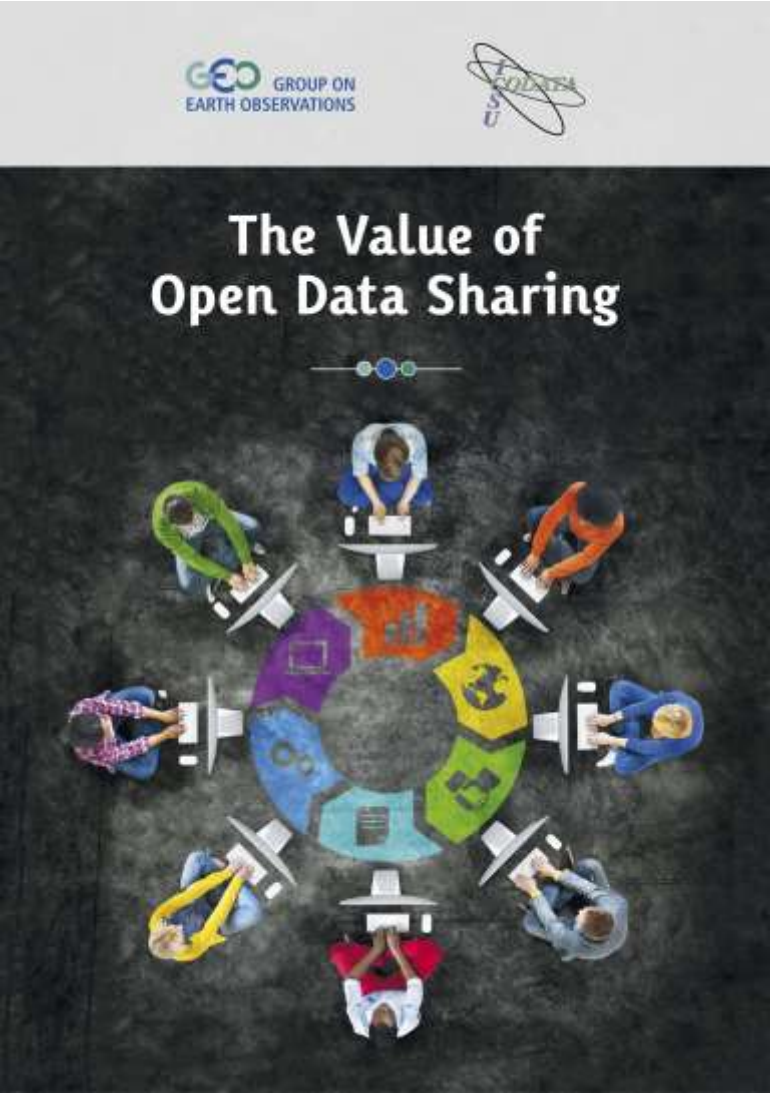
Improving crop varieties with open data on breeding trials: AgTrials

Cultivar testing is an important means of improving crop varieties. A wide range of trials are taking place on sites all over the world, addressing issues such as drought tolerance, heat stress, and soil management. However, almost all of the data generated has been inaccessible to other researchers – filed away on laboratory hard drives, or sometimes lost completely due to bad data management.

By compiling data from agronomic and plant breeding trials and making it open, the Global Agricultural Trial Repository (**AgTrials**)[38] hosted by a CGIAR Research Programme on Climate Change, Agriculture and Food Security (CCAFS) offers a rich knowledge base to inform ongoing, collaborative research, while eliminating unnecessary and costly duplication of efforts.

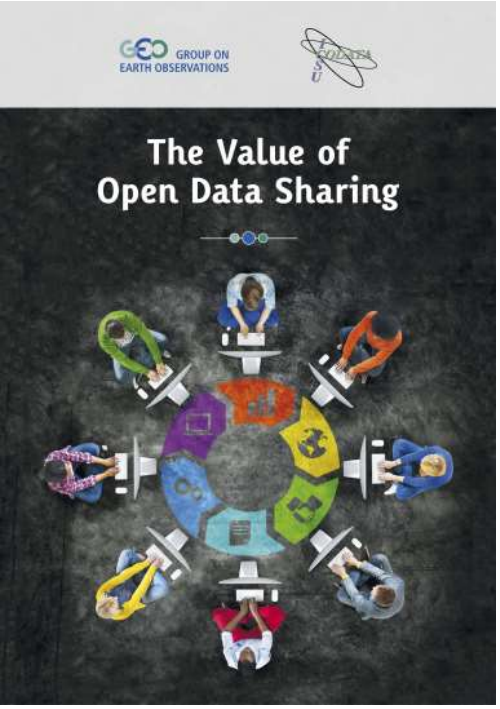
Scientists used 250 open AgTrials datasets to build crop models specific to the West Africa region. The models are used to project the local impacts of climate change, and define breeding programmes for adaptation.[39]

Societal Benefits of Data Sharing



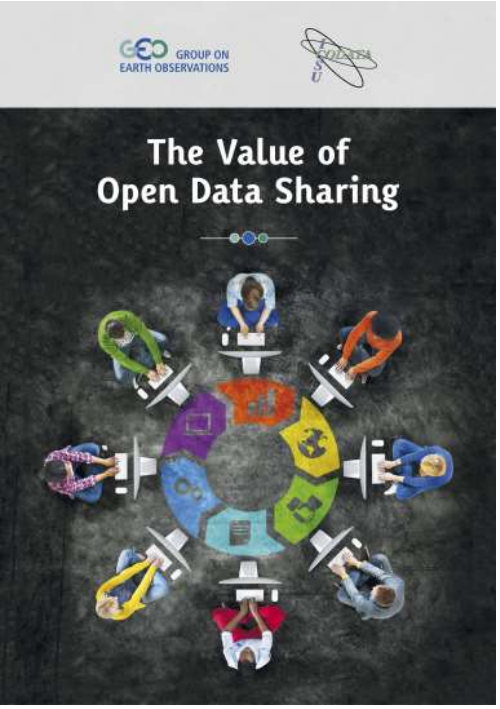
- 'The societal benefits of Earth observations cannot be achieved without data sharing'.
- Provides a concise, accessible, high level synthesis of key arguments and evidence of the benefits and value of open data sharing.
- Particular, but not exclusive, reference to Earth Observation data.
- Benefits in the areas of:
  - Economic Benefits
  - Social Welfare Benefits
  - Research and Innovation Opportunities
  - Education
  - Governance
- Available at http://dx.doi.org/10.5281/zenodo.33830

The Value of Open Data Sharing



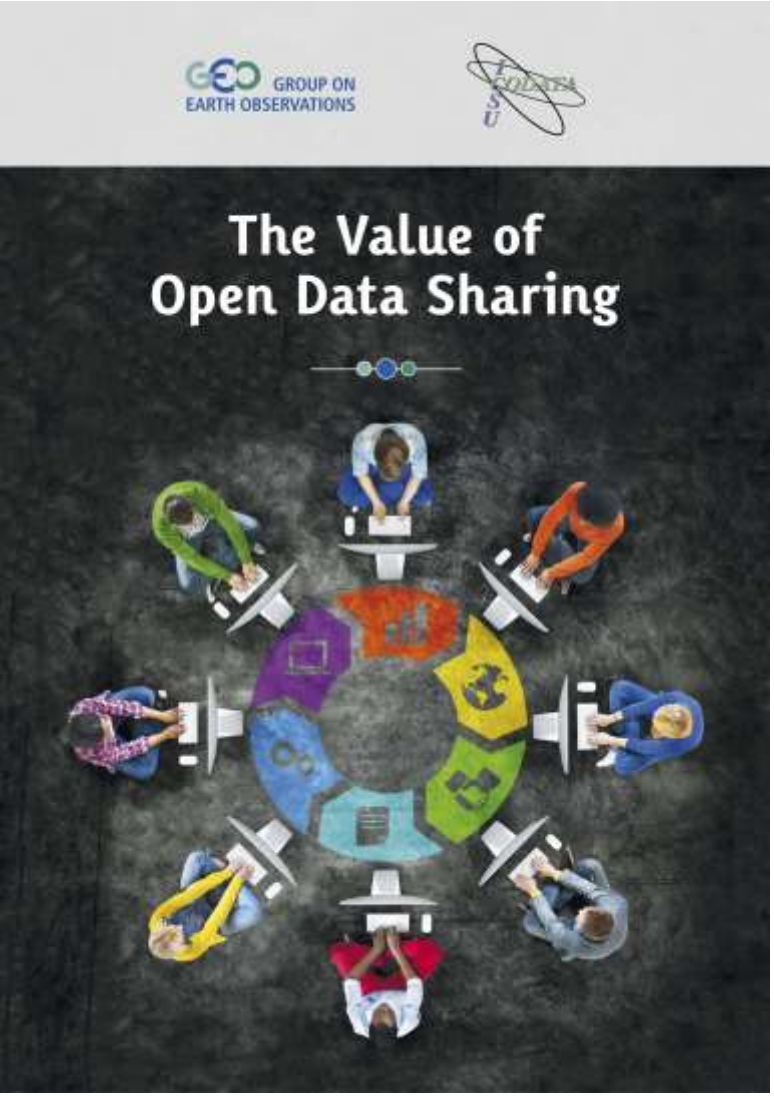
- **Social Welfare**
  - Political and ethical arguments in favour of public access to publicly funded data.
  - Social expectation of availability of this resource to public and to enterprises.
- **Research and Innovation Opportunities**
  - EO data have particularly strong research and education applications.
  - Very strong global interest in the use of EO data for research (SBAs)
  - Benefits accrue across disciplinary and national boundaries such that it not appropriate to restrict access.
  - Data mining and 'transdisciplinary' data integration for major research questions.
  - Verification of results, reproducibility, depends on open data.
  - Citizen science.
  - Downstream applications and commercial innovation.

The Value of Open Data Sharing



- **Education**
  - Considerable market demand for data skills, including GIS.
  - Considerable use of EO data in teaching.
- **Governance**
  - Open data for better research for better decision-making.
  - Essential component for capacity mobilisation and improved environmental, agricultural management in lower and middle income countries.
  - Equitable access to data ‘patrimony’ for all countries, including those with lower financial resources.

Economic Benefits of Data Sharing



- ‘Many studies and reports have documented the positive value of openness for EO data, specifically, and for various other types of data and information, more generally.’
- Weiss 2002: quantified considerable economic benefits of making meteorological data open ($400-700M in gross receipts; businesses and employment).
- Houghton 2011: apart from economic benefits, gross saving for Australian Bureau of Statistics of AU$3.5M by eliminating charging and management structure.
- Houghton 2014: Estimate unrealised benefits of research data of AU$1.4-4.9BN set against estimated AU$130-200M cost of data infrastructure.
- **Open data and open data infrastructure has a significant economic benefit.**

Figure 3: The value and impacts of the three UK data centres

	Investment & Use Value (Direct)	Contingent Value (Stated)	Efficiency Impact (Estimated)	Return on Investment (Estimated)
ESDS	Based on Depositor & User Counts	Based on User Count	Based on User Count	Based on Data Spend
	Investment Value £23m per annum	**Willingness to Pay** £25m per annum	**Survey Community** £68m-£112m per annum	**Increased RoI** [*From Additional Use*] £58m - £233m 2.5 to 10-fold RoI
	Use Value £24m per annum	**Willingness to Accept** £81m per annum (£111m per annum)	**Wider Community** £100m plus per annum	
ADS	Based on Deposit & Download Counts	Based on Download Count	Based on User Count	Based on Data Spend
	Investment Value £1.2m per annum	**Willingness to Pay** £1.1m per annum (Constrained by capacity to pay)	**User Community** [*ADS data use*] £13m per annum	**Increased RoI** [*From Additional Use*] £2.4m - £9.7m 2.1 to 8.3-fold RoI
	Use Value £1.4m per annum	**Willingness to Accept** £7.4m per annum	**User Community** [*All activity time*] £58m per annum	
BADC	**Investment Value** £2.8m per annum	**Willingness to Pay** £5.2m per annum	**User Community** [*BADC data use*] £10m per annum	**Increased RoI** [*From Additional Use*] (*Non Recreate*) £11m - £34m 4 to 12-fold RoI
	Use Value £2.3m per annum	**Willingness to Accept** £16m per annum	**User Community** [*All activity time*] £58m per annum	

The Challenge: Sustainable Business Models for Data Repositories

- Research funder policies – quite rightly – mandate data stewardship.
  - OECD Principles and Guidelines, 2007
  - G8 Science Ministers Statement, 2013
  - Major funders in US, UK, EC Horizon 2020 data policy etc.
- Increasing need for data repositories and data stewardship.
  - Increasing volume presents a challenge.
  - Requirements for stewardship present a greater challenge.
- **Sustaining digital data infrastructure is a major issue for science policy!**
- Genuine concern that current funding models will prove inelastic and not meet the growing requirements – concern on the part of repositories and funders.
- Witnessing Innovation
  - Changes in funding / business models (ADS, DANS, ICPSR)
  - Innovative business models (Dryad, FigShare)

The Challenge: Sustainable Business Models for Data Repositories

- Policy agreement that the cost of data stewardship is an essential, integral part of the cost of doing research.
- Strong value proposition for data infrastructure and data sharing.
  - CODATA White Paper for GEO: *The Value of Open Data Sharing*: http://dx.doi.org/10.5281/zenodo.33830
- Relatively little work has been done on the economics and business models of data infrastructure.
  - Blue Ribbon Task Group Report on Sustainable Digital Preservation: http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf
  - Sustaining Domain Repositories for Digital Data: A White Paper (ICPSR): http://datacommunity.icpsr.umich.edu/sites/default/files/WhitePaper_ICPSR_SDRDD_121113.pdf
- Pressing need for work on who pays and how: analysis of income streams, of innovative funding models, of willingness to pay and responsibilities, of business models in general.
- **OECD Global Science Forum is the ideal setting for this work.**

OECD Global Science Forum Project: Sustainable Business Models for Data Repositories

- Questions to address:
  1. How are data repositories currently funded?
  2. What innovative income streams are available?
  3. What means of restraining costs are available?
  4. How do income streams match willingness/ability to pay of various stakeholders?
  5. How do income streams/willingness to pay fit together into a **sustainable** business model?
- Builds on previous work of RDA-WDS Interest Group: http://dx.doi.org/10.5281/zenodo.46693
- Broader landscape study of current funding models, May-August 2016.
- Focus group on innovative income streams and on cost restraint, workshop October 2016.
- Micro and macro economic analysis of business models, Oct 2016-Jan 2017.
- Test business models with stakeholder groups, workshop Feb-March 2017.
- Policy recommendations based on concrete business model options, April-May 2017.

The Case for Open Data in a Big Data World

- **Science International Accord on Open Data in a Big Data World:** http://www.icsu.org/science-international/accord
- Presents a powerful case that the profound transformations mean that data should be:
  - Open by default
  - Intelligently open
- Supported by four major international science organisations.
- Lays out a framework of principles for how the vision of Open Data in a Big Data World can be achieved.
- **Please endorse the Accord on Open Data in a Big Data World: http://bit.ly/endorse_open_data_accord**
- Parallel initiative for an African open science capacity initiative.

CODATA Big Data / Open Data Capacity Initiative

- Important to mobilise capacity in Africa to take advantage of the data revolution and to mitigate unequal realisation of its benefits.
- Open data and big data represents a major opportunity, but if LMICs are to take advantage of the scientific opportunities available, capacity development is urgent!
- Holistic 'science systems' approach: policies, procedures, incentives, data infrastructure, scholarly communications, skills and training.
- **Keystone is to establish an Open Data forum or platform with coordinating role.**
- Support from Department of Science and Technology in South Africa.
- Pilot initiative will be resourced by DST and implemented by staff from South African Academy of Sciences, under direction from CODATA.

Centre of Excellence and Open Data at JKUAT

(Jomo Kenyatta University of Agriculture and Technology)

- International Workshop on Open Data for Science and Sustainability in Developing Countries, August 2014, UNEP-UNESCO and JKUAT, Nairobi, Kenya.
- Kenyan 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 JKUAT
- **'JKUAT has now established an ICT Centre of Excellence and Open Data (iCEOD) that was part of the Nairobi-CODATA conference recommendation'**
- CODATA working with the iCEOD on data policy and training activities.

Current Best Practice for Research Data Management Policies

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

Mimi Hudson and Laura Molloy

May 2014

Building the Initiative

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

Benefits of Open Data and the SDGs

- Strong evidence of the benefits of open data sharing.
- Importance of open data both for monitoring of SDGs and for interventions required to help us achieve them!
- Economic benefits of open data sharing and the ROI for data infrastructure.
- Need for sustainability models for that infrastructure.
- Need for initiatives that build the science base around open data.
- **CODATA Report for GEO on Benefits of Data Sharing:** http://dx.doi.org/10.5281/zenodo.33830
- **Endorse the Accord on Open Data in a Big Data World:** http://bit.ly/endorse_open_data_accord
- **Come to International Data Week:** http://www.internationaldataweek.org/

International Data Week 2016

- Jointly organised by CODATA, RDA and WDS: **12-16 September, Denver, Colorado, USA**
- Combines 1) two-day research conference, SciDataCon 2016, 2) an international data forum focusing on policy discussion, intersections with open public data and data science, data driven innovation, 3) RDA Plenary 8.

INTERNATIONAL DATA WEEK 2016

WWW.INTERNATIONALDATAWEEK.ORG

Organized by:

RESEARCH DATA ALLIANCE

ICSU WORLD DATA SYSTEM

SciDataCon 2016

http://www.scidatacon.org/2016/

- **Part of International Data Week:** Organised by CODATA and WDS: **12-13 September, Denver, Colorado, USA**
- Scientific conference on 'advancing the frontiers of data in research'.
- Deadline for session proposals, 18 March 2016: 48 of 90+ proposals selected
- Deadline for paper submissions, 30 May 2016: http://www.scidatacon.org/2016/submit/
- Major event to look at data issues, needs, challenges and contributions of all research areas.

Thank you for your attention!

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