Forschungsdatenmanagement:
Chancen und Herausforderungen aus Sicht der Hochschulleitungen
Hochschulrektorenkonferenz
Bonn, Germany
16 December 2016

Managing Research Data in Universities and the Challenge of Open and FAIR Data

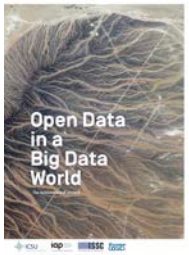
Dr Simon Hodson
Executive Director, CODATA
www.codata.org

CODATA

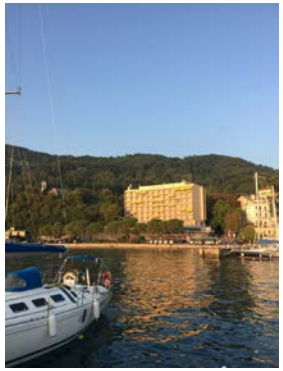
Principles, Policies and Practice

DC1 Data Citation Principles

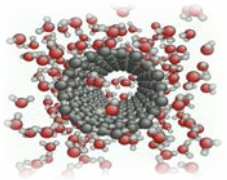
ICSU International Council for Science



Capacity Building



Frontiers of Data Science



]u[ubiquity press open scholarship

Data Science Journal

IDW 2016, 11-17 Sept, Denver, CO.

INTERNATIONAL DATA WEEK 2016
WWW.INTERNATIONALDATAWEEK.ORG

Why Open Science / FAIR Data?

- Good scientific practice depends on communicating the evidence.
- Boulton: to fail to communicate the data that supports scientific assertions is malpractice.
- Open research data are essential for reproducibility, self-correction.
- Research data produced by publicly funded research are a public asset.
- 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.
- Open data practices have transformed certain areas of research.
- Research data should be open by default; but there are good reasons for restricting some data.
- **Intelligently Open Data to FAIR Data.**
- **As open as possible, as closed as necessary.**
- Research data is an asset to research institutions and universities.
- Essential role of RDM in universities in the data ecosystem.

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, 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
- Science International Accord on Open Data in a Big Data World, Dec 2015: http://bit.ly/opendata-bigdata

International Perspectives on RDM

Drivers

- Good research practice.
- Open Science, Open Data and return on investment.
- Data assets, resources to address major international research questions.

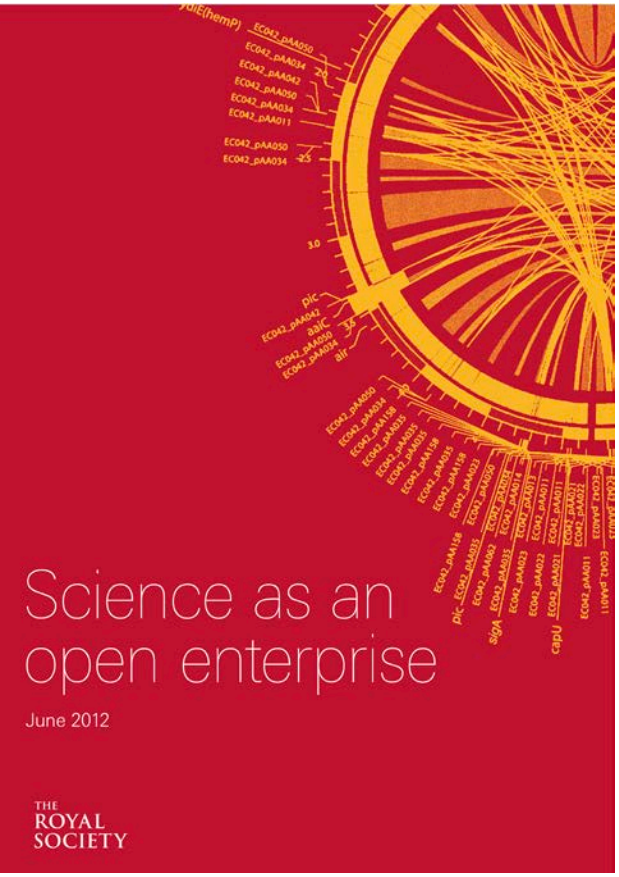
Policy Frameworks

- OECD > Funders Data Policies > Increasing number of institutional data policies > Journal data policies.
- Intelligently Open Data and FAIR data.
- Endorsed by international coordinating organisations (e.g. ICSU, Belmont Forum, GEO) and international research programmes.

Institutional Responses

- Good practice: Interpretation and implementation of data **policies**.
- Data as an asset: development of data **ecosystems and infrastructures**.
- Data Science Skills Agenda: development of researcher **skills**.
- Recognition and reward: promotion of **good practice** (e.g. data citation).

Science as an Open/Public Enterprise



- The digital age has brought a data revolution that presents science with major challenges and opportunities.
- Incredible advances in certain disciplines due to the ability to create and analyse data at scale.
- Open data / immediate release of data is a significant accelerating factor (particularly in genomics, astronomy, earth observation etc).
- Concern over (lack of) reproducibility and distance between article and data (scholarly communications relatively slow to adapt to the digital age).
- Concern to ensure that society / public is engaged with science / research: opportunities of citizen science,
- **Data for research should be intelligently open: accessible, assessible, intelligible, useable.**
- **FAIR: Findable, Accessible, Interoperable, Reusable.**
- **Publications and data should be Open and available concurrently: argues that not to make data concurrently open is scientific malpractice.**

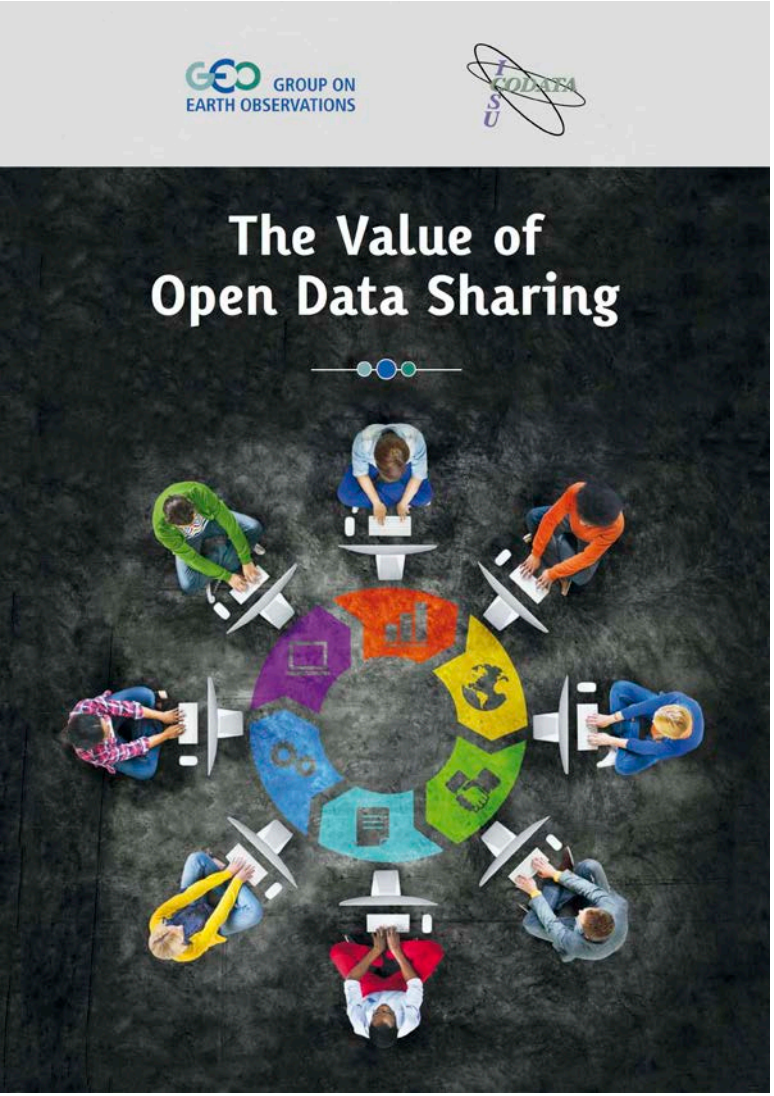
Benefits of Open Data: some examples from GEO

- **Barbara Ryan, Director of Secretariat GEO, TED-X Talk Barcelona**
- In 2008 US Government was convinced to make Landsat Data openly available, for free.
- Under charging, the highest number of downloads was 53 scenes per day.
- Now over 5700 scenes per day are downloaded.
- Spanish deforestation research: under the charging regime data access alone would have cost €260M
- **CODATA produced a White Paper on the Value of Data Sharing for the GEO-XII Plenary: http://dx.doi.org/10.5281/zenodo.33830**

https://www.youtube.com/watch?v=9umWTFgFIVs

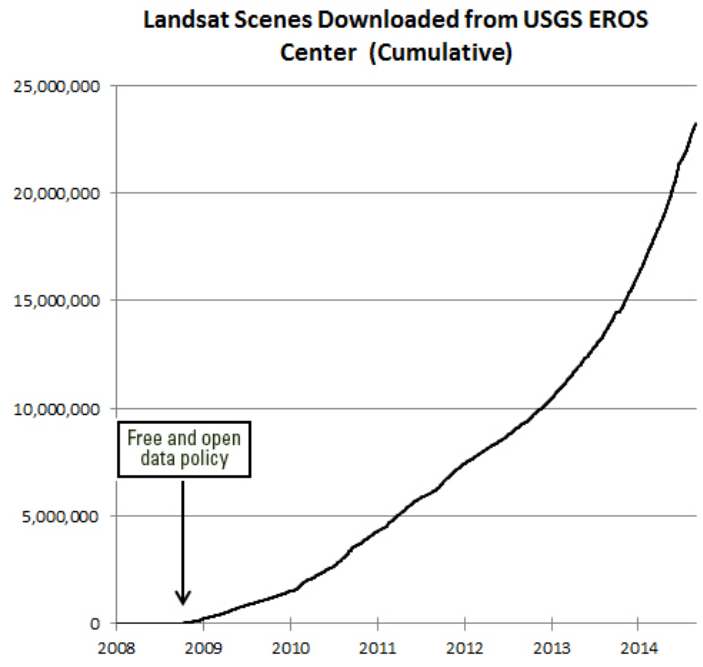
GROUP ON EARTH OBSERVATIONS

The Value of Open Data Sharing



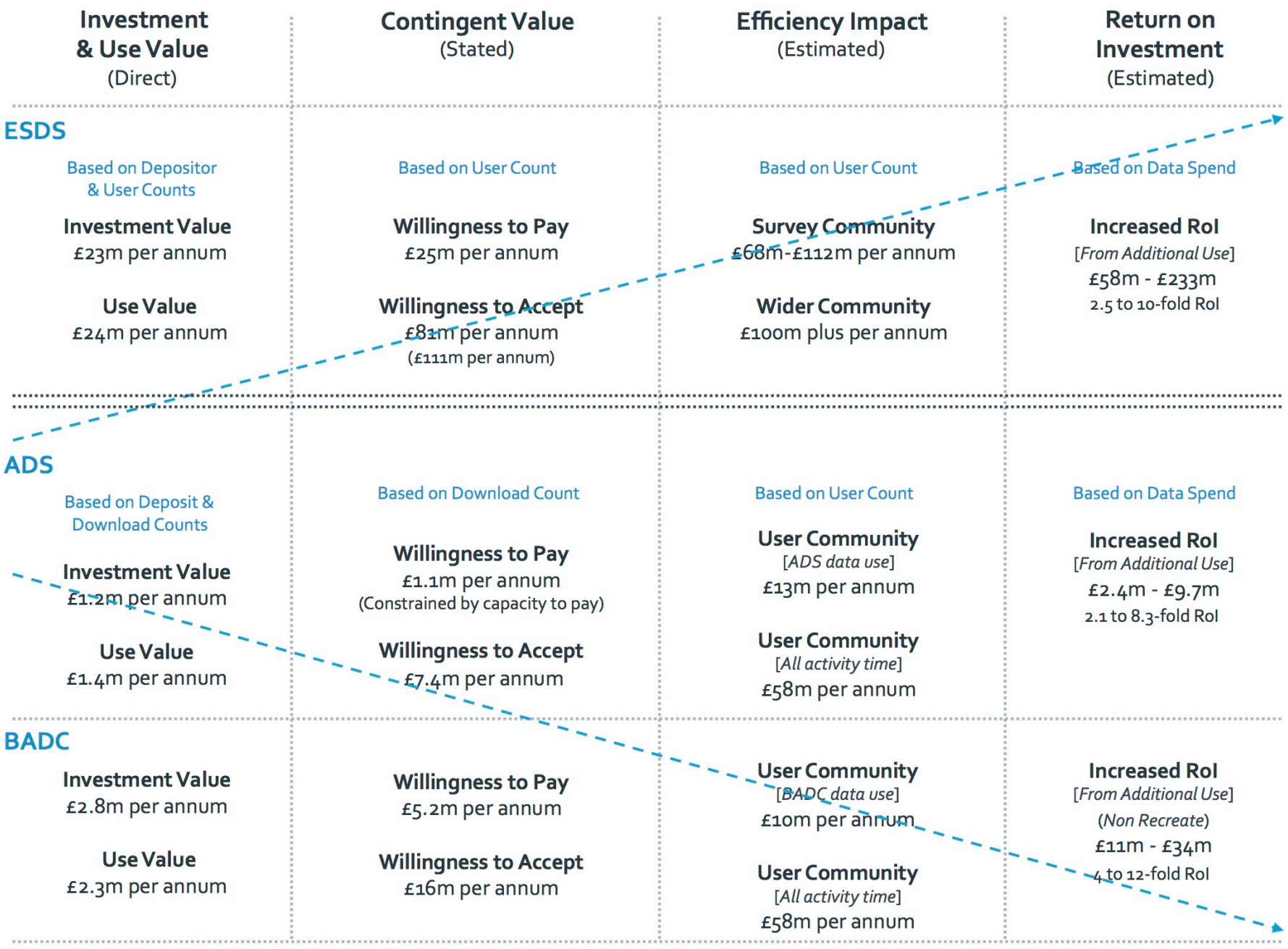
- Report by CODATA for GEO, the Group on Earth Observation.
- Provides a concise, accessible, high level synthesis of key arguments and evidence of the benefits and value of open data sharing.
- Particular, but not exclusive, reference to Earth Observation data.
- Benefits in the areas of:
  - Economic Benefits
  - Social Welfare Benefits
  - Research and Innovation Opportunities
  - Education
  - Governance
- Available at http://dx.doi.org/10.5281/zenodo.33830
- **GEO DSWG is building on this work with further examples: would be valuable to work with this community.**

Economic Benefits of Data Sharing: LandSat



- **2006 Study** estimated the loss in case of a data gap as equivalent to US$935 M.
- **2011 Study** estimated benefits of landsat-sourced information for agriculture as US$858 M just for the state of Iowa.
- **2015 Study** estimated worldwide economic benefit of US$2.19 BN.
- Estimated benefit in US of US$1.8 BN.
- Valuing Geospatial Information: Using the Contingent Valuation Method to Estimate the Economic Benefits of Landsat Satellite Imagery: http://dx.doi.org/10.14358/PERS.81.8.647 **(Paywall… Irony…)**
- **Open data and open data infrastructure has a significant economic benefit.**

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



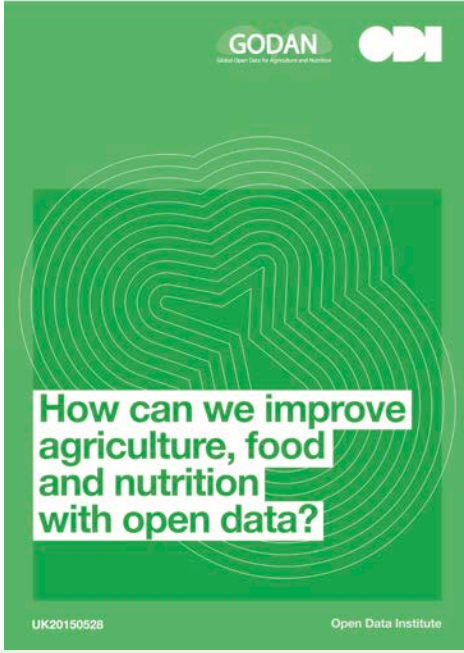
Data Revolution: how can we improve … with open data?

- 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]

The Challenge: Business Models for Sustainable 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, TAIR; DANS, ICPSR)
  - Innovative business models (Dryad, FigShare)

OECD Global Science Forum Project: Business Models for Sustainable 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 survey of current funding models, May-Oct 2016.
- Focus group on innovative income streams and on cost restraint, workshop Nov 2016.
- Micro and macro economic analysis of business models, Nov 2016-March 2017.
- Test business models with stakeholder groups, workshop March 2017.
- Policy recommendations based on concrete business model options, April-June 2017.
- **Explicitly includes how data repositories and services in RPOs / Universities might be sustainably resourced.**

OECD BETTER POLICIES FOR BETTER LIVES

ICSU WORLD DATA SYSTEM

RDA RESEARCH DATA ALLIANCE

Universities and Research Data: Opportunities and Challenges

Open and FAIR Research Data Presents Major Opportunities for Universities

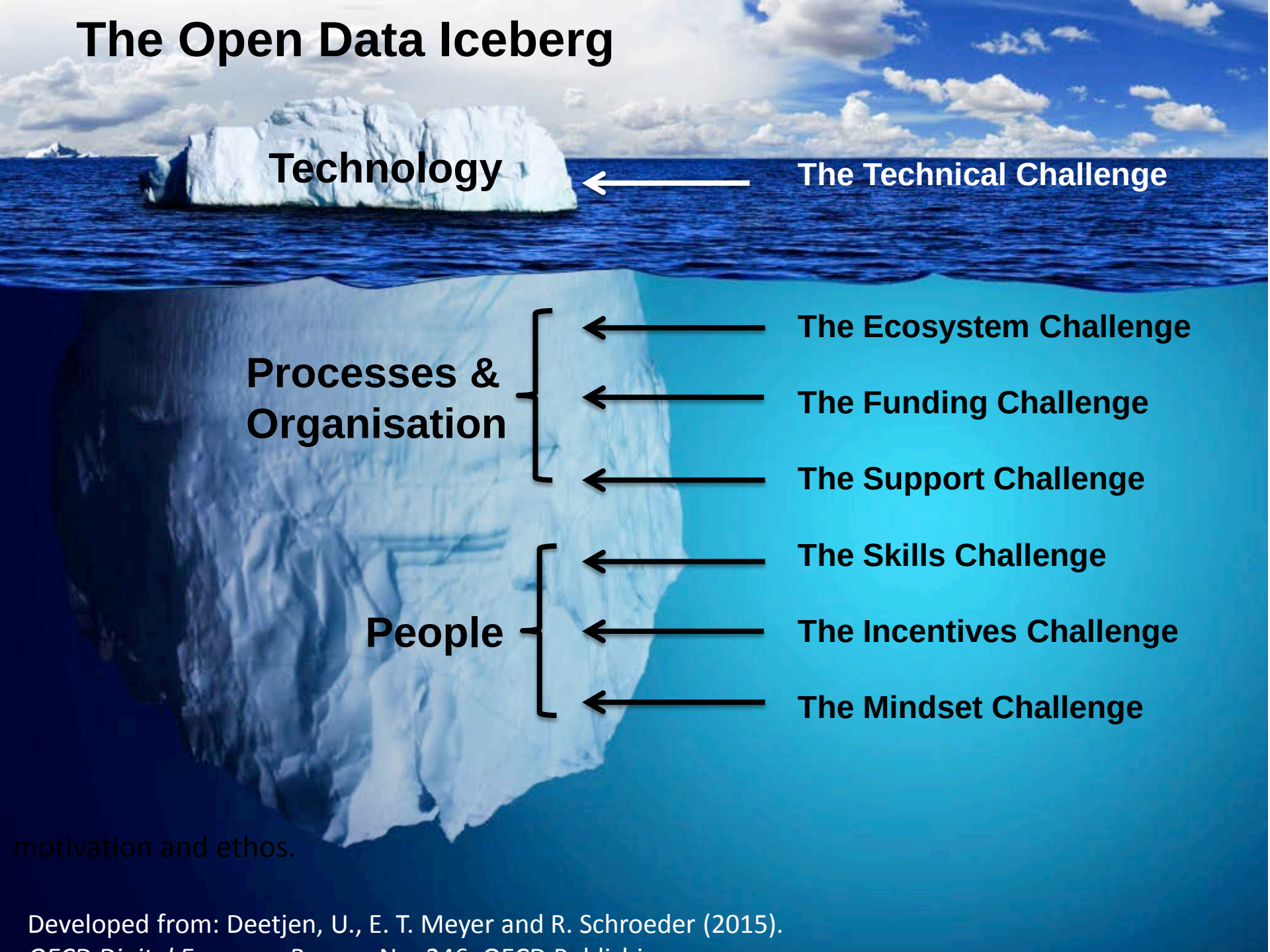
- Research intensive universities will be data intensive universities.
- Supporting researchers' use of data is a key strategic mission and enabler.
- A university's reputation is increasingly built on all research outputs and wider societal and economic impact: data is core to this.
- **Development of significant data collections of research intensive universities.**
- The way in which the contribution to research of both the individual researcher and the institution will increasingly be measured on the basis of data outputs as well as research articles.

Challenges for RDM in Universities

- Finding niche in data ecosystem.
- Supporting the long term stewardship of research data, as part of a flexible and collaborative system.
- Sustainability and finance.
- Supporting data through the lifecycle.
- Policy development: unpicking Open and FAIR data
- Culture and incentives: what's in it for us?
- Training and support.

Resources:

- CODATA http://www.codata.org/ ; https://doi.org/10.5281/zenodo.165830
- Research Data Alliance and World Data System.
- National Programmes in UK, Australia => ANDS http://www.ands.org.au/
- LERU RDM Roadmap: http://www.leru.org/files/publications/AP14_LERU_Roadmap_for_Research_data_final.pdf
- RfII "Recommendations regarding structures, processes, and financing for research data management in Germany': http://www.rfii.de/?wpdmdl=2075

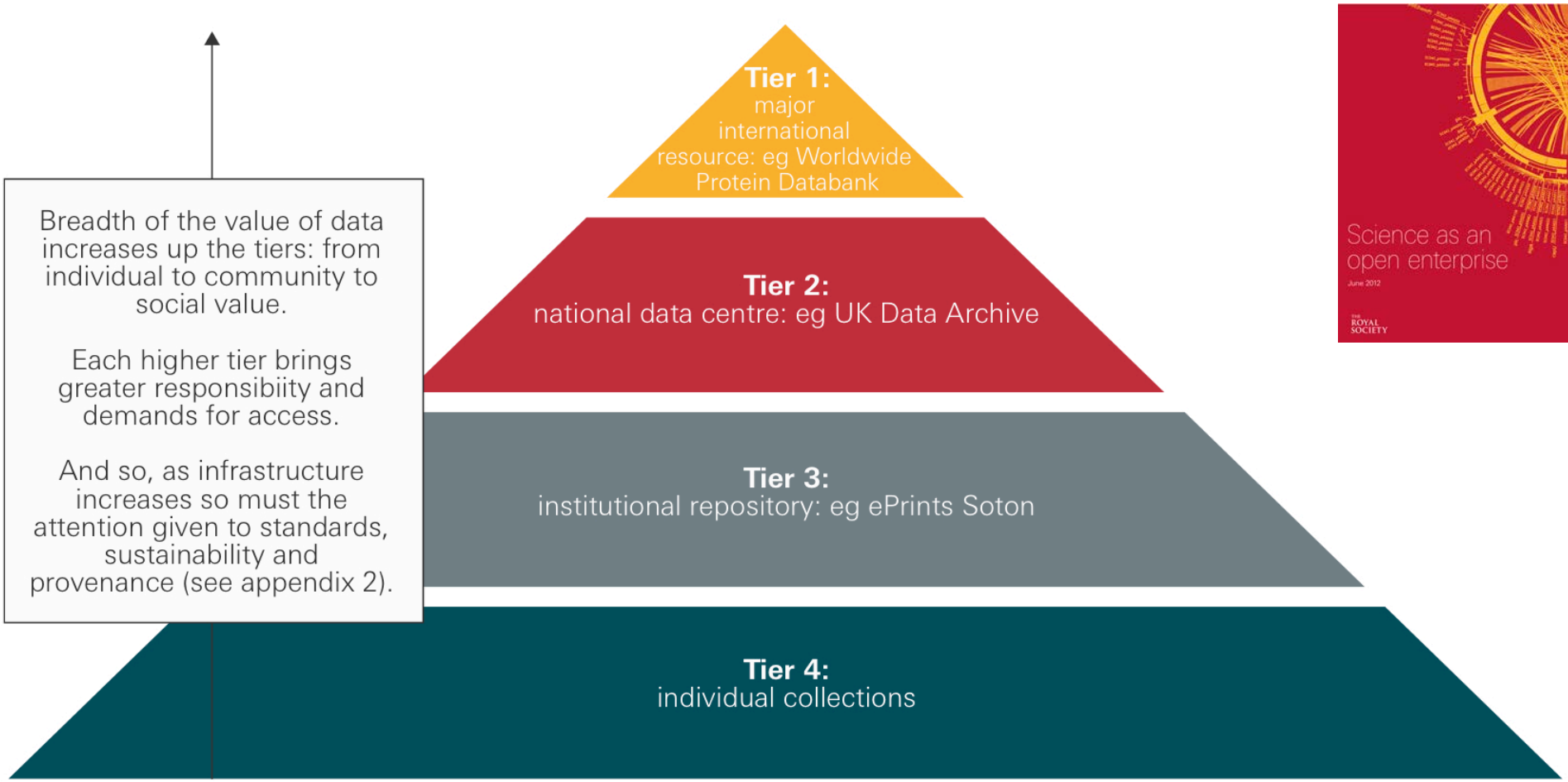


Developed from: Deetjen, U., E. T. Meyer and R. Schroeder (2015).

Research Data Infrastructures / Roadmaps

- **Research priorities and gap analysis.**
- **Ecosystem:** what is the provision of RDIs for particular disciplines through national and international initiatives?
  - International disciplinary and EC developments.
  - National infrastructures (RfII recommends Nationale Forschungsdateninfrastruktur) 'collaborative, flexible etc'.
  - Institutional infrastructures and responsibilities: key participants in NFDI, have interests and responsibilities in RDM.
- **Role of Research Institutions:** Is lifecycle support and long tail being supported in institutions.
- RDIs are not just hardware, but 'part of a research ecosystem', so must address: governance; training, personnel and career structures' sustainable funding; access and outreach to national, public and commercial partners.
- **Roadmap for Data Infrastructure**
  - Co-design to meet national needs and research priorities.
  - Opportunities for shared infrastructures
  - Examples of good governance and sustainable funding models.
  - **Sustainable Business Models for RDIs**

Increasing the Value of Data Resources

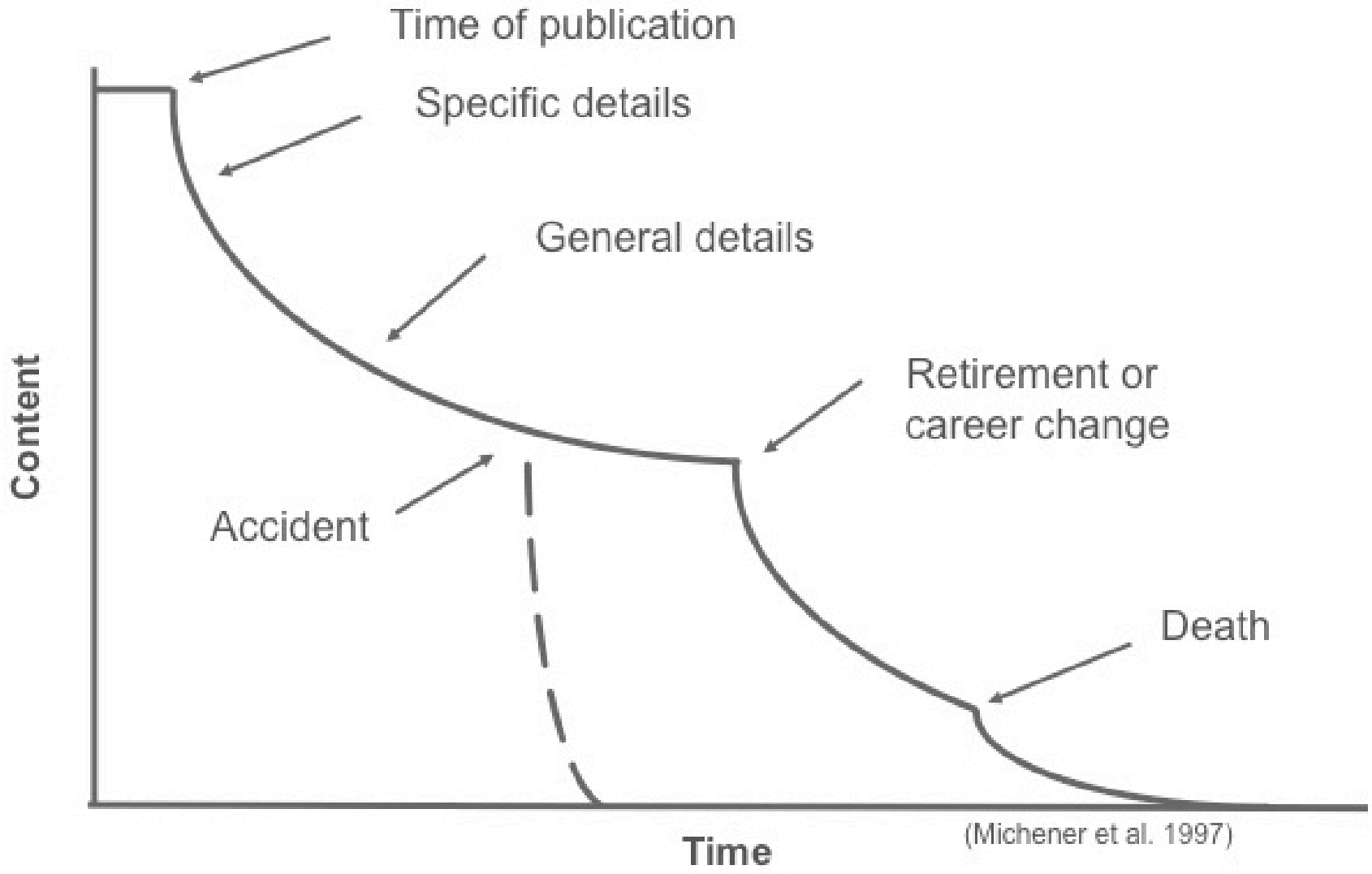


The Data Pyramid: taken from Royal Society Report, *Science as an Open Enterprise:* *http://royalsociety.org/policy/projects/science-public-enterprise/report/*

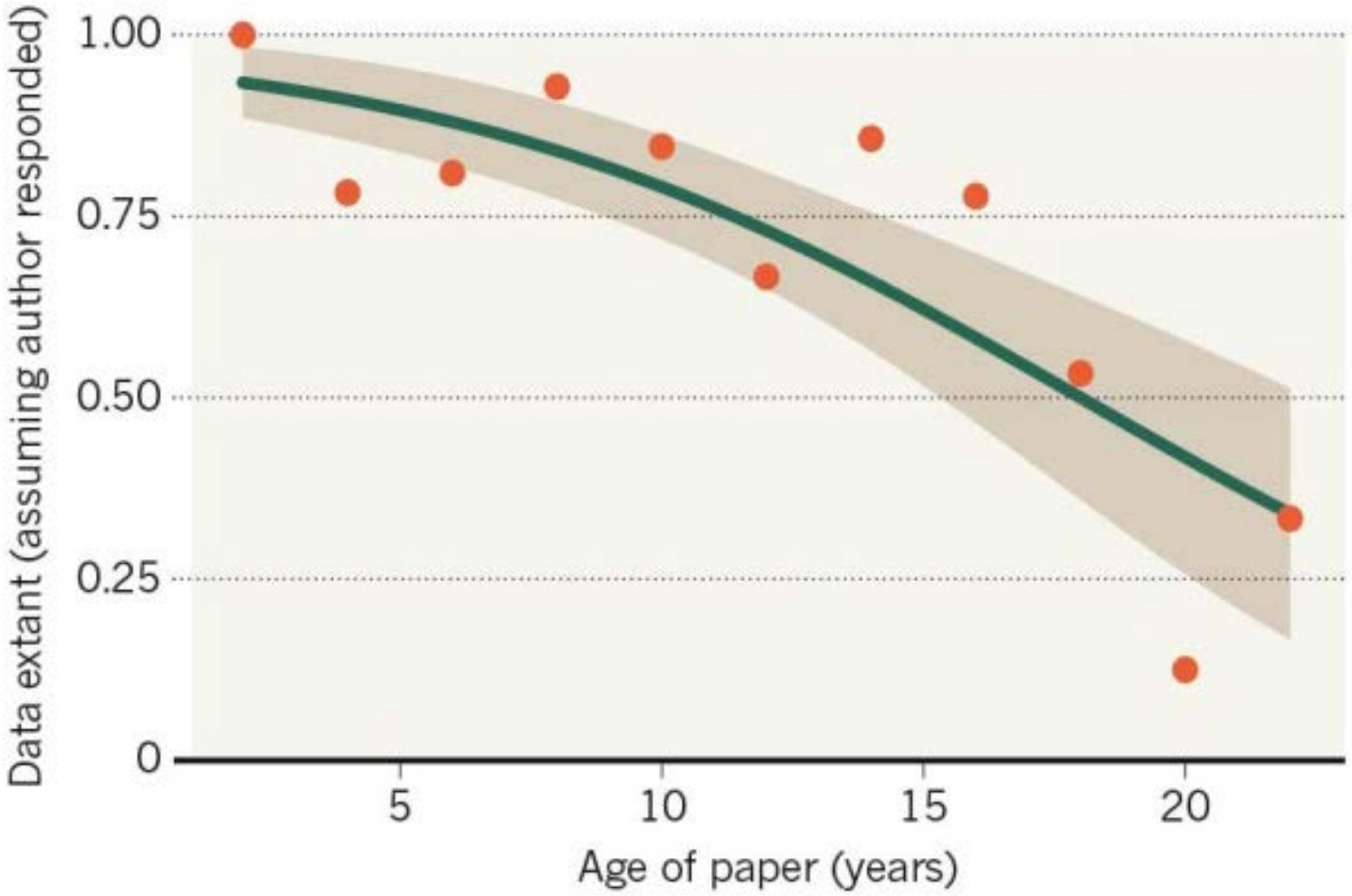
Where should research data go?

Homogenous data collections essential for research	• Earth observation data; • Genetic data; • Social science survey data…	National and international data archives
Significant data outputs of publicly funded research	• Significant data outputs from funded projects; • Raw and analysed experimental data…	National or institutional data archives; data papers
Data underpinning research publications	• Raw and analysed data for reproducibility (evidence); • Data behind the graph…	Dedicated data archives (e.g. Dryad)

Data Entropy

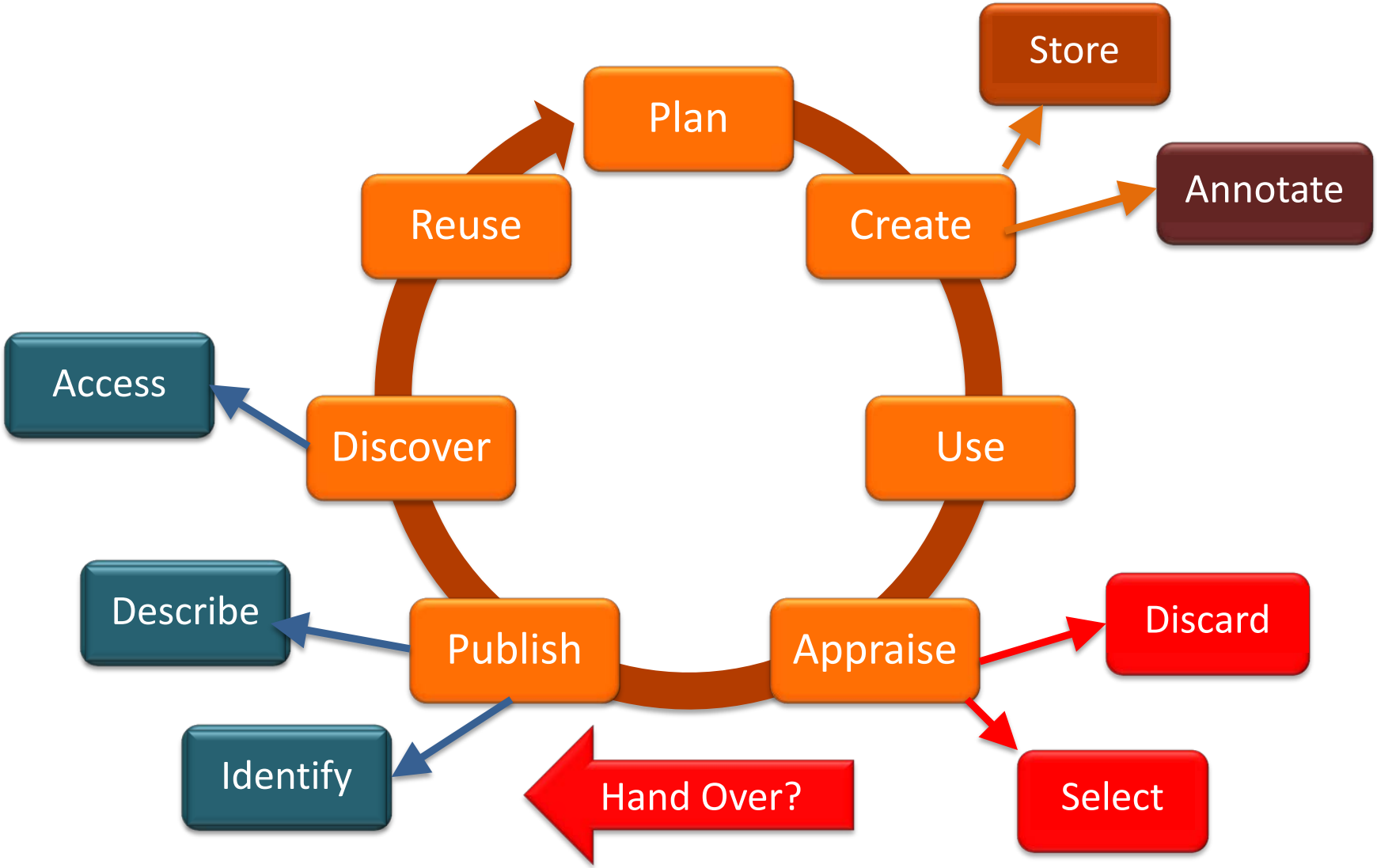


80% of ecology data irretrievable after 20 years (516 studies)

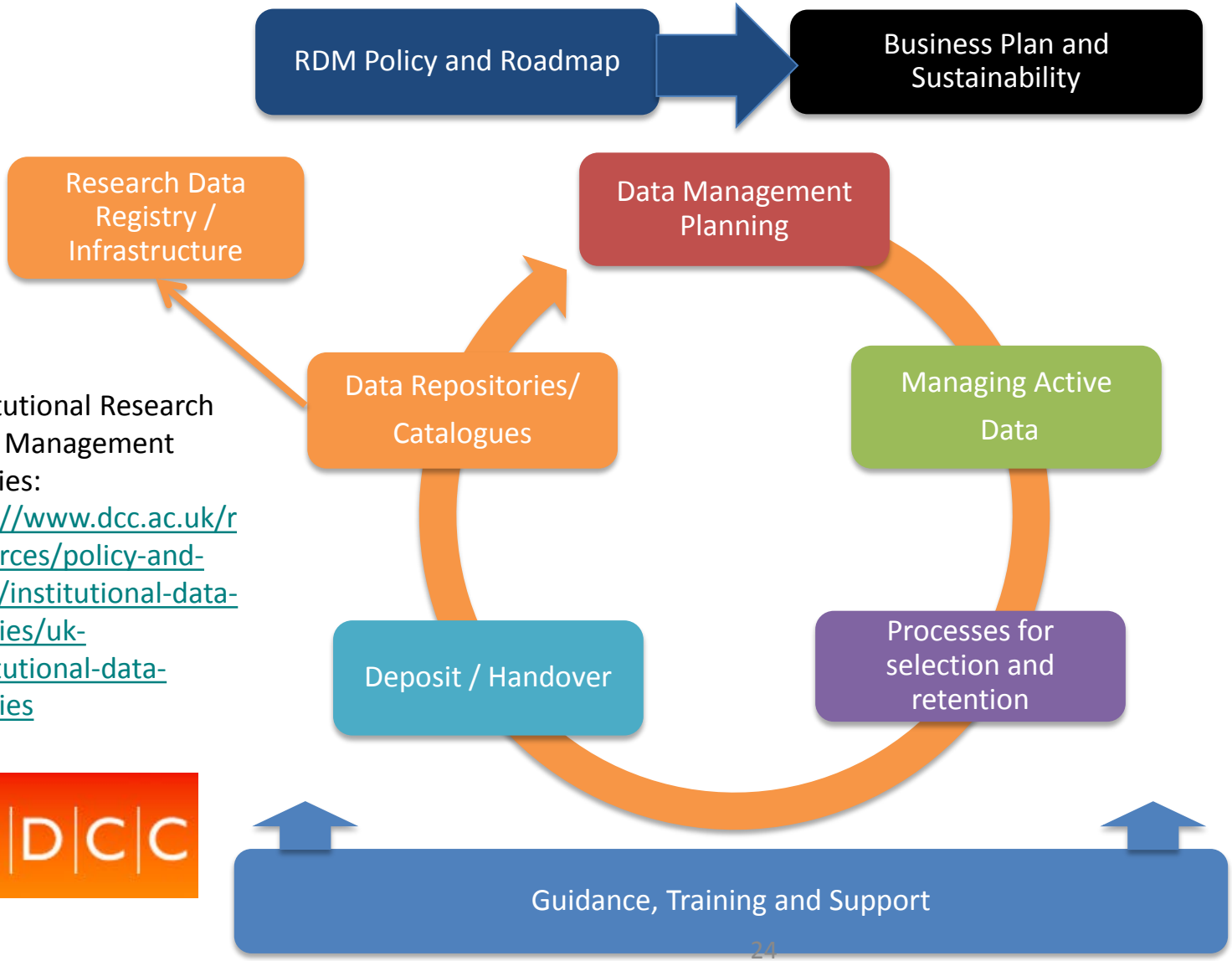


Vines TH *et al*. (2013) Current Biology DOI: 10.1016/j.cub.2013.11.014

Supporting the Research Data Lifecycle



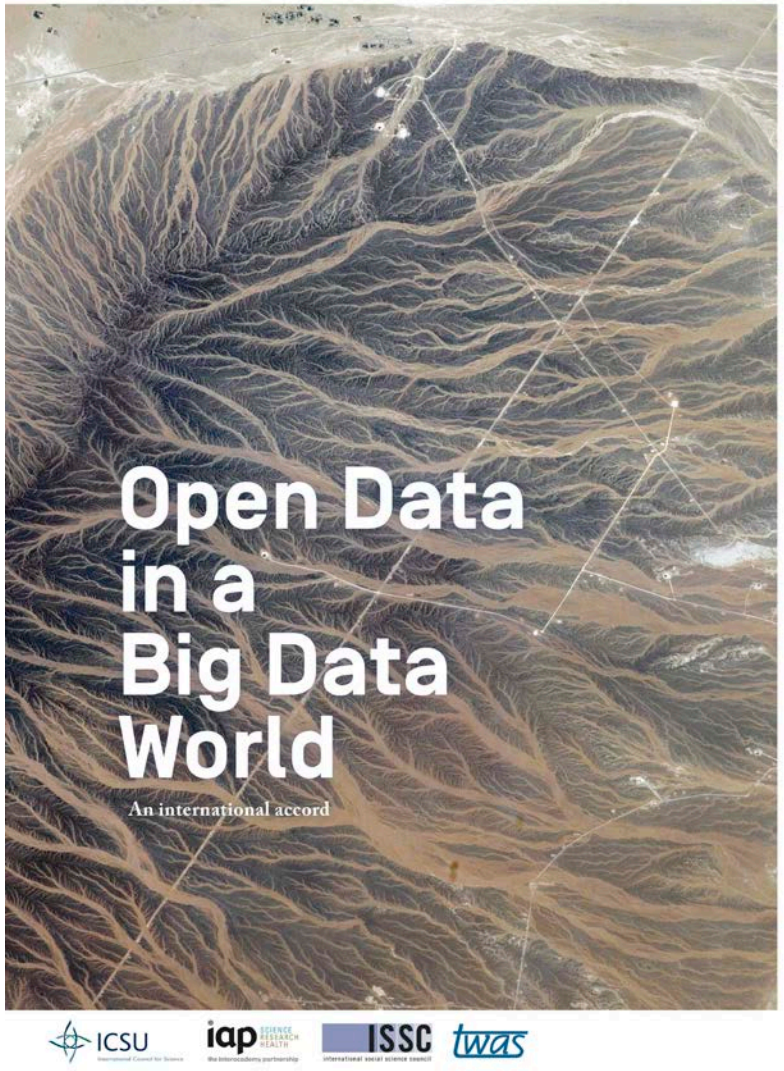
Components of RDM support services



Institutional Research Data Management Policies: http://www.dcc.ac.uk/resources/policy-and-legal/institutional-data-policies/uk-institutional-data-policies

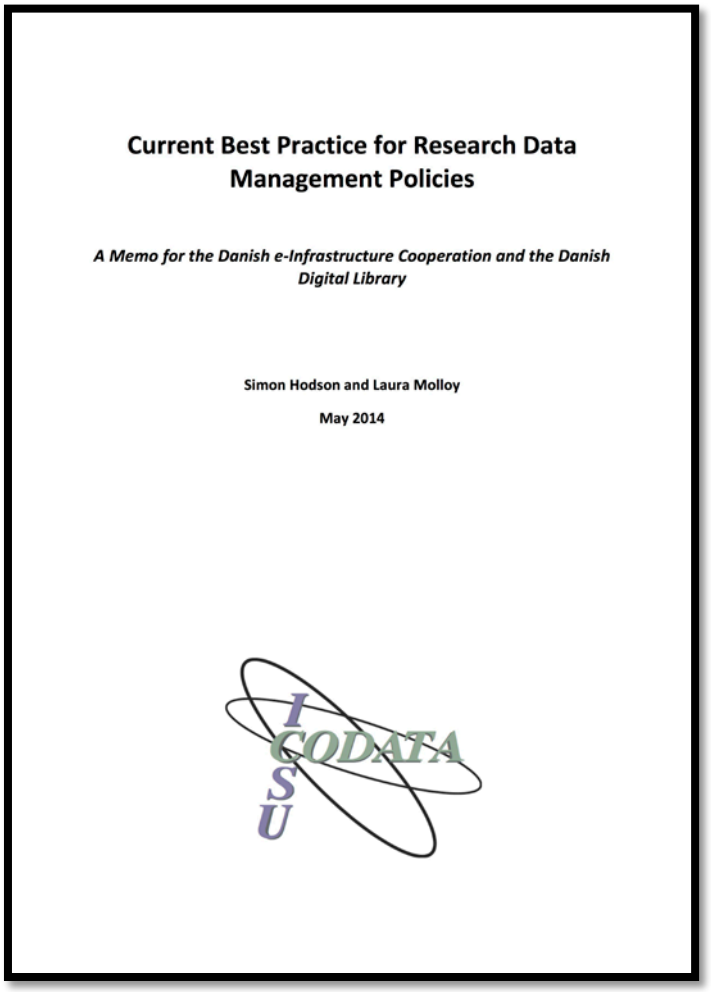
Principles and Enabling Practices

- **Science International Accord on Open Data in a Big Data World: http://www.science-international.org/**
- Supported by four major international science organisations.
- 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.
  - National Open Research Strategy.
  - Open data policies and guidance at national and institutional level.
  - Collaborative infrastructures for certain research disciplines.
  - Development of institutional infrastructure for research collaboration and data stewardship/RDM.
  - Mechanisms (infrastructure and policy) to ensure concurrent publication of data as research output.
  - Promotion of data skills (researchers and data stewards).
  - Open practices and assessment of research contribution.
  - Explore the boundaries of open (particularly privacy, IPR).

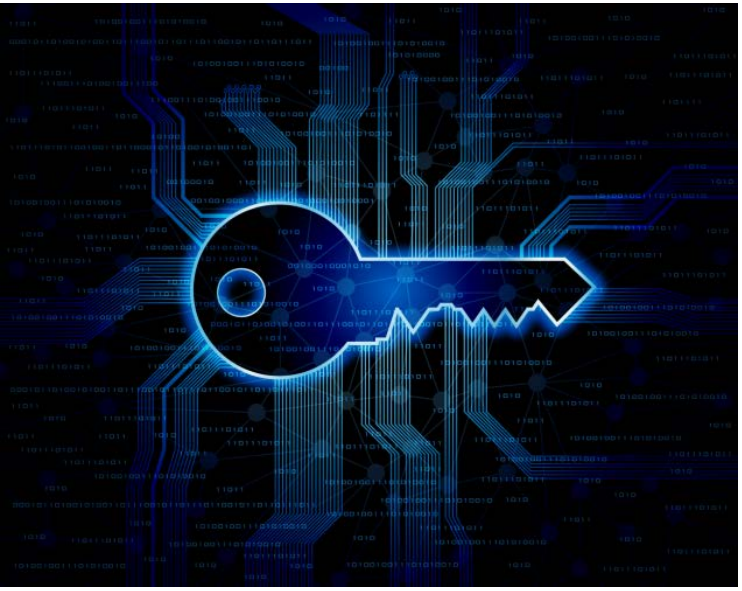


Resources: Current Best Practice for Research Data Management Policies

- **Expert report commissioned by CODATA member.**
- Provides comprehensive summary of best practice in funder data policies.
- Identifies key elements to be addressed:
  1. Summary of policy drivers
  2. Intelligent openness
  3. **Limits of openness**
  4. **Definition of research data**
  5. **Define data in scope**
  6. **Criteria for selection**
  7. Summary of responsibilities
  8. Infrastructure and costs
  9. DMP requirements
  10. Enabling discovery and reuse
  11. **Recognition and reward**
  12. Reporting requirements, compliance monitoring
- Zenodo: http://dx.doi.org/10.5281/zenodo.27872



Boundaries of Open



- For data created with public funds or where there is a strong demonstrable public interest, **Open should be the default.**
- **Proportionate** exceptions for:
  - Legitimate **commercial** interests (sectoral variation)
  - **Privacy** ('safe data' vs Open data – the anonymisation problem)
  - **Public interest** (e.g. endangered species, archaeological sites)
  - **Safety, security** and dual use (impacts contentious)
- All these boundaries are fuzzy and need to be understood better!
- Should not allow these issues to lead to blanket exemptions: e.g. H2020.
- **There is a need to evolve policies, practices and ethics around closed, shared, and open data.**

Barriers to Data Availability / Publication

Researchers concerns:

- Concern that data may be misused or misunderstood.
- Concern that will lose scientific edge if sharing before fully exploited.
- Desire to retain control of a professional asset.
- **Concern that will not be credited.**
- **Lack of career rewards for data publication.**
- See ODE report, using Parse.Insight findings: http://www.alliancepermanentaccess.org/wp-content/uploads/downloads/2011/11/ODE-ReportOnIntegrationOfDataAndPublications-1_1.pdf
- Culture in particular research disciplines; availability of infrastructure.
- **Fundamentally, researchers are reluctant to expend effort sharing data because they do not feel that data is adequately exposed or credited.**

Nature special issue on data sharing: http://www.nature.com/news/specials/datasharing/index.html

Integrating Data with Scholarly Communications and Rewards

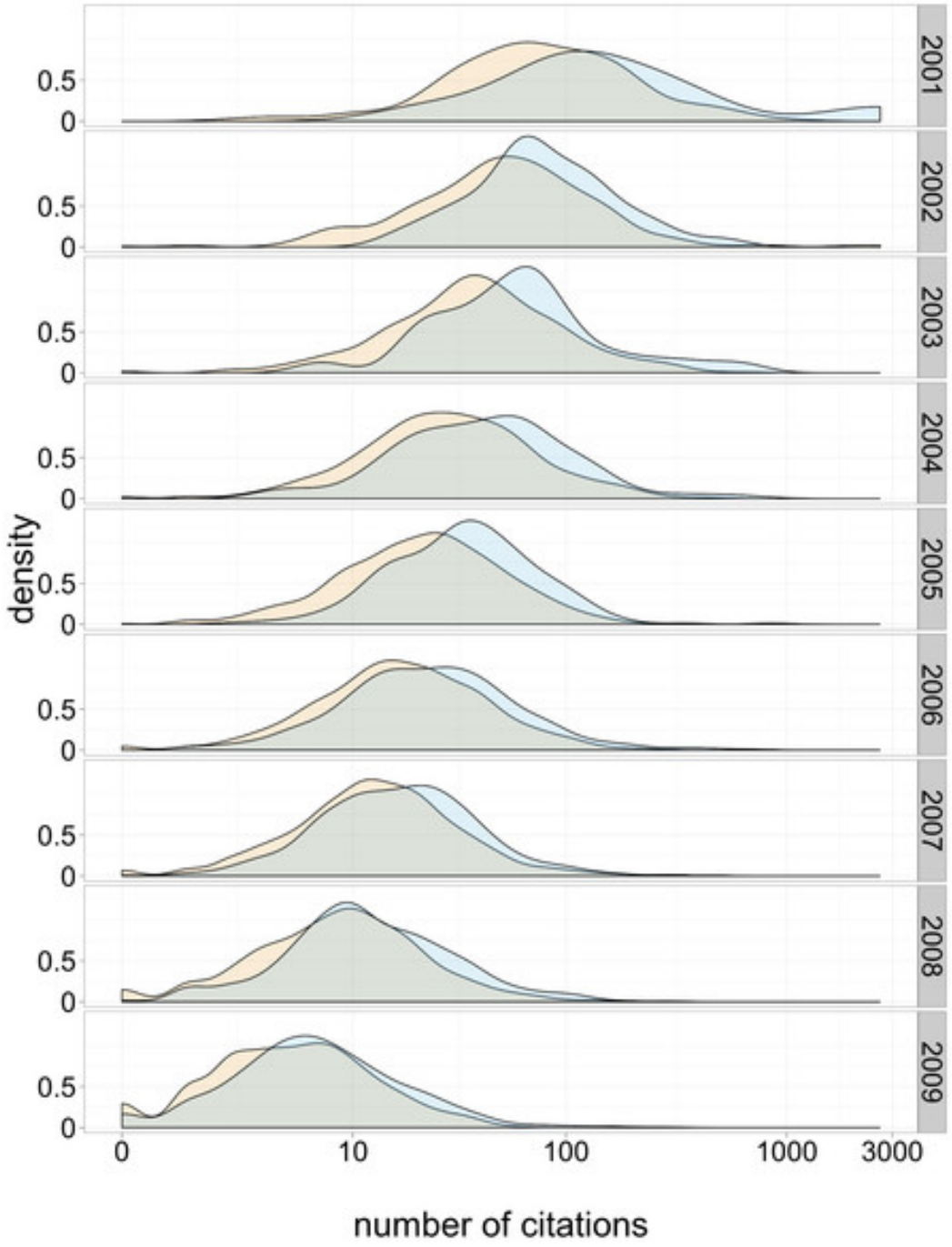
- The way in which the contribution to research of both the individual researcher and the institution will increasingly be measured on the basis of data outputs as well as research articles.
- Researcher's reputation will increasingly rely on data – data is one of the assets which build reputation and enable collaborations.
- When we make our knowledge and research outputs open for discussion, increasingly that means data too.
- **Data is integrated with the process of scholarly communications.**

Developments: Journal Data Policies

- **Dryad Joint Data Archiving Policy, Feb 2010: http://datadryad.org/jdap**
- This journal **requires, as a condition for publication**, that data supporting the results in the paper should be archived in an appropriate public archive, such as GenBank, TreeBASE, Dryad, or the Knowledge Network for Biocomplexity.

- **PLOS Data Availability Policy, revised Feb 2014:** http://www.plosone.org/static/policies.action#sharing
- PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exceptions.

- **Jisc work to develop registry of journal data policies; BioSharing https://biosharing.org/**
- **Likely new initiative through RDA to encourage development and adoption of journal data policies.**
- **CODATA working with ICSU to encourage ISUs to address data policy from disciplinary perspective.**



Citation advantage of having archived Gene Expression Omnibus data

Examined **10,555** studies that created gene expression microarray data, comparing those that made data available and those that didn't.

Studies that made data available in a public repository received 9% more citations than similar studies for which the data was not made available.
Increased citation of 30% for those published 2004-5.

Piwowar and Vision (2013), PeerJ DOI:10.7717/peerj.175

Incentives: 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:

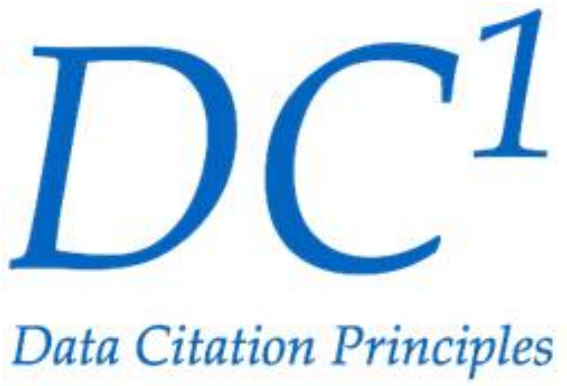
http://bit.ly/data_citation_principles

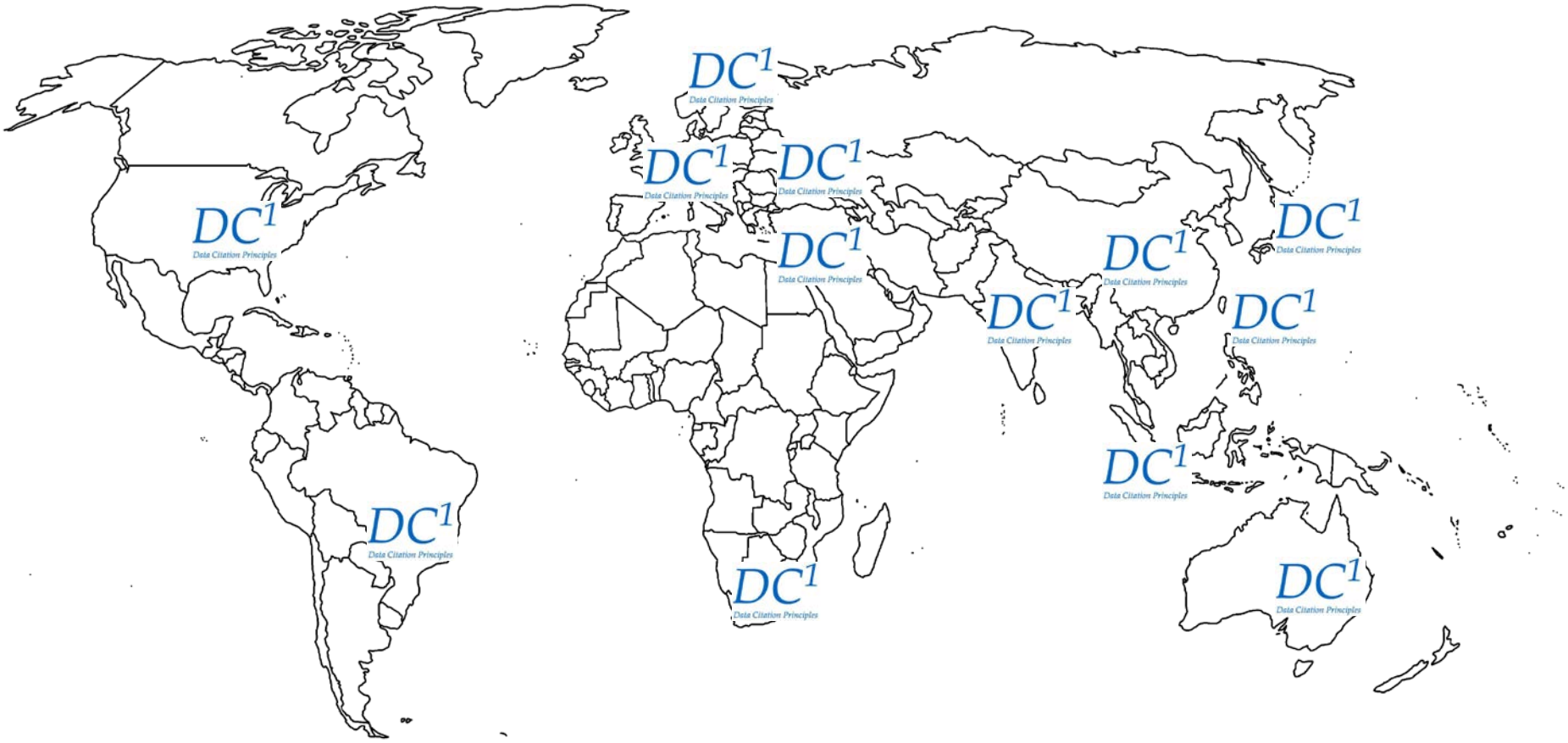
DC1

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!

2015: China, Australia, Japan, India and South Africa.
2016: USA, Israel, Russia *+ Finland (Nov) and Taiwan (Dec).*
2017: France, Korea, Indonesia, Brazil…

Synthesis Report of first 8 workshops to be published in soon!

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 Research Data Science 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

CODATA-RDA School of Research Data Science

- **First School of Research Data Science, 1-12 August 2016, ICTP, Trieste**
- Funding for students and tutors provided by ICTP, TWAS, CODATA, ACU, RDA Europe, GEO and GODAN.
- Attended by 70 students from all around the world.

The Abdus Salam International Centre for Theoretical Physics

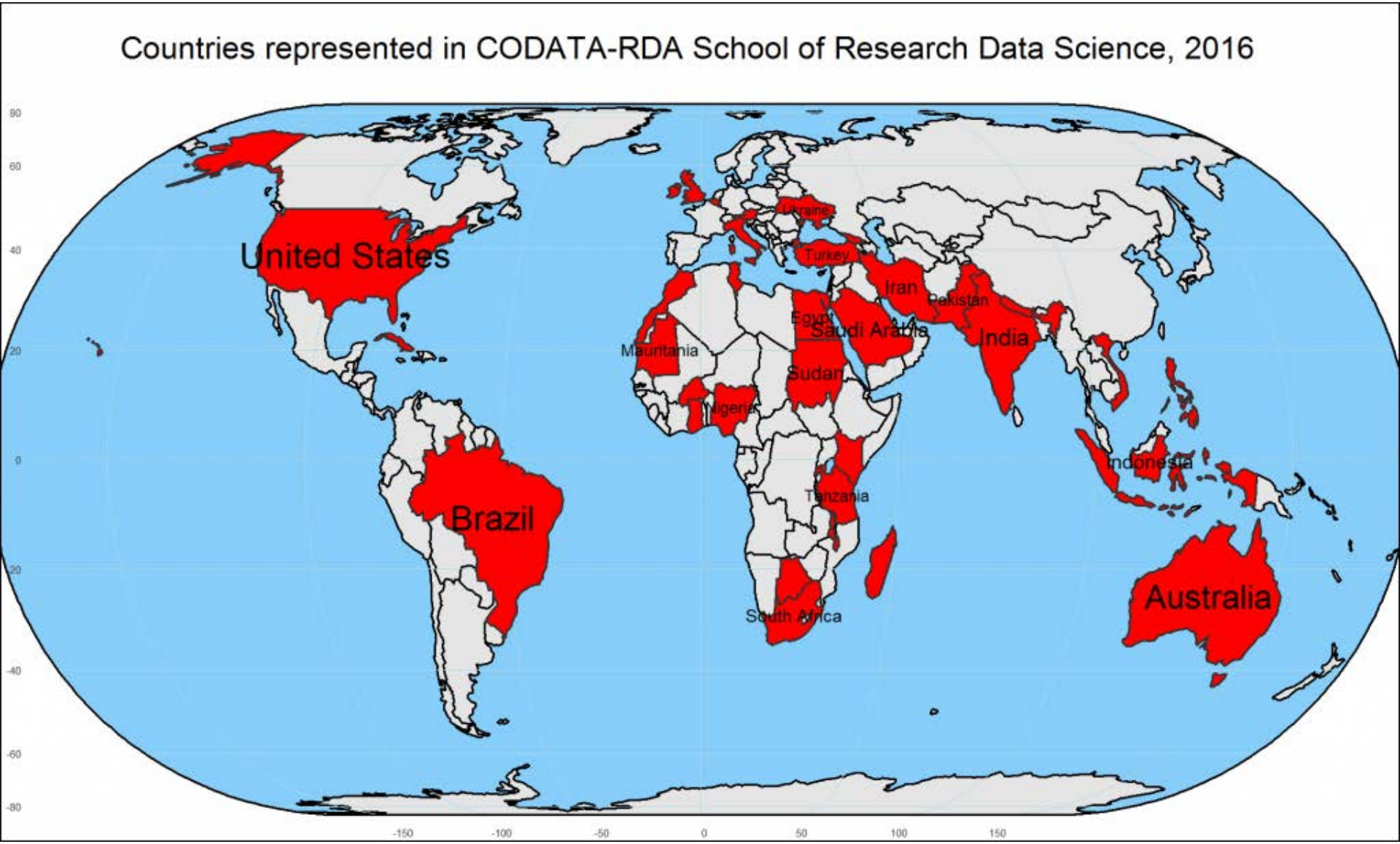
RESEARCH DATA ALLIANCE EUROPE

The Association of Commonwealth Universities

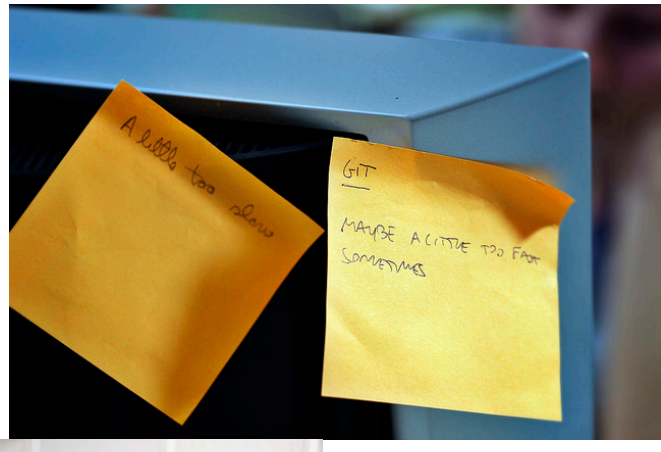
GODAN Global Open Data for Agriculture & Nutrition

#DataTrieste

#DataTrieste



#DataFoo…



- **Programme for #datatrieste** http://bit.ly/School_of_Research_Data_Science-Programme
- **School will repeat at Trieste in July 2017 and in 2018**
- Possibly with addition of one week more advanced courses for particular disciplines.
- **Will run foundational two week course at ICTP INESP in Sao Paolo, Brazil, December 2017.**
- Schools can be run with a greater or lesser degree of support and coordination from the international convenors.
- **Keen to encourage a network of schools, but also local schools with lower central input.**

- Discussions with possible partners in South Africa and India.
- Keen to explore opportunities with CODATA National and Union Members.

Opportunities, Challenges and Support!

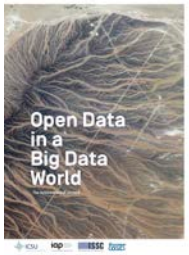
- Turn towards data is a major development in 21st Century Science.
- Universities' missions oblige them to engage with this development.
- Significant opportunities for those universities that develop data expertise and specialised data collections.
- Challenges: some technical challenges, but more challenges of funding, coordination, culture, policy and practice are more significant.
- Need for flexible, collaborative infrastructures; **and for institutions to take key role in these things.**
- Major progress is being made internationally with RDM in institutions.
- **RDM in Universities, International Conference, planned for c.3-7 July 2017, Göttingen.**
- Engage with international developments through CODATA and RDA.

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

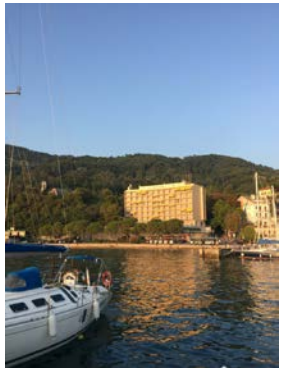
Principles, Policies and Practice

DC1 Data Citation Principles

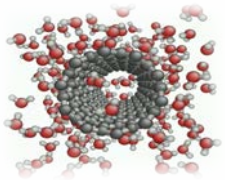
ICSU International Council for Science



Capacity Building



Frontiers of Data Science



]u[ubiquity press open scholarship

Data Science Journal

IDW 2016, 11-17 Sept, Denver, CO.

INTERNATIONAL DATA WEEK 2016
WWW.INTERNATIONALDATAWEEK.ORG

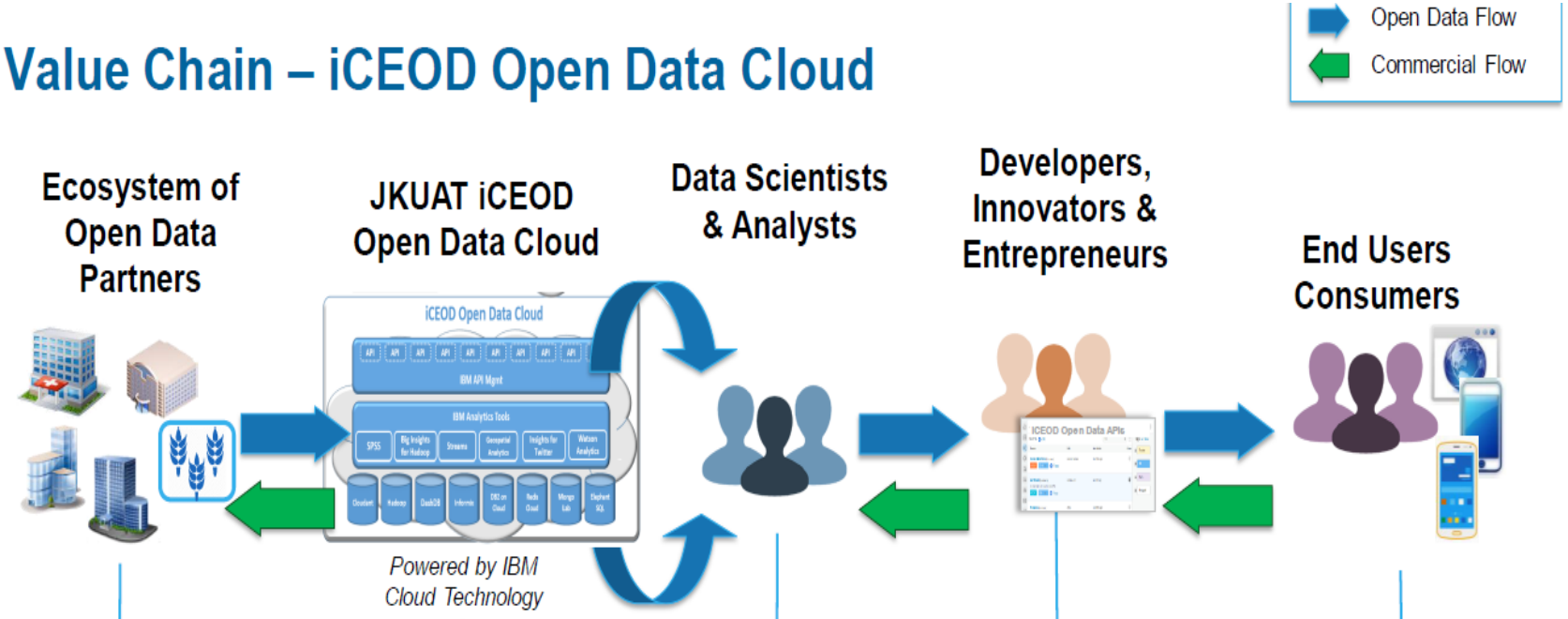
INTERNATIONAL COUNCIL FOR SCIENCE

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

iCEOD Value Chain: Data and Society Agriculture and Nutrition

Value Chain – iCEOD Open Data Cloud



Open Data from iCEOD and their partner ecosystem including
- Hospitals
- Research Partners
- Agriculture Partners
- Governments

JKUAT Staff and PhD. students expose and monetize APIs across IBM enabled platforms

iCEOD Staff, JKUAT faculty, PhD students and other Research Partners enrich and extend the Open Data by applying analytics to unlock valuable insights

Developers, ISVs, Entrepreneurs access Open Data through a catalogue of free and monetized APIs using these to create innovative apps and solutions

End Users take advantage of these innovative Apps to address their needs and problems, commercialised paid and free Apps with in-app purchases

An Open Research Data Strategy for Poland?

- Collaboration on a national workshop to develop a national open research data / open science strategy for Poland.
- CODATA leads met earlier this year with representatives from Ministry of Science and Education and with Open Science Centre to plan a workshop in 2017.
- Draws strongly on the approach of the accord.
- **Stakeholders and Responsibilities:** governments/funders, universities and research institutions, institutional libraries, national academies and learned societies, national and international research and data infrastructures, publishers and journal editorial boards.
- **Working Groups on Enabling Practices:** boundaries of open, normative values (sharing, timeliness), non-restrictive reuse and TDM, incentives, interoperability, sustainability of data infrastructure, data literacy.

OPEN SCIENCE PLATFORM

Open Research Data

- More qualitative indicators on data policy framework, data infrastructure and skills/education?
- **Policy Development:**
  - To what extent have government funders, charitable funders, research institutions, journals and publishers developed and implemented data policies?
- **Data Infrastructures:**
  - Availability and scope of data infrastructures?
  - Assessment of what data infrastructure exists, for what subject areas?
- **Development of data related skills:**
  - Skills and competence frameworks, incorporating data related skills (e.g. EDISON Project http://www.edison-project.eu/ and FOSTER https://www.fosteropenscience.eu/)
  - To what extent are data skills being addressed in curricula (recommendations, policy, skills frameworks etc).

Open Research Data

- **What proportion of publicly-funded research data is 'openly available'?**
- **What proportion of publicly-funded research data is 'well-managed'?**
- Current information from funders, research output assessment, RPO CRIS infrastructure is short of giving useful indicators.
- Research assessment exercises (do these pay enough attention to data)?
- How effective are institutional CRIS in capturing information about datasets? CERIF metadata standard was updated to include datasets in 2013.
- ResearchFish platform used by RCUK is capturing information about datasets.
- ***A policy and priority issue…***

- **Opportunities…**
- Production of **data management plans**: e.g. Horizon 2020 pilot, IDRC pilot, DMPonline tools > various funders require data management plans. Provides some form of indicator.
- **National aggregator services**: e.g. Research Data Australia currently has over 115K datasets… https://researchdata.ands.org.au/
- Journal data availability policies and **data citation**. TR Data Citation Index.

What is Open Research?

- Strong case that Open Research is not just Open Access (to scientific literature) + Open (Research) Data.
- One definition Includes:
  - Open Methodology (including open practices such as open lab-book etc).
  - Open Data
  - Open Source (and open availability of software, code and algorithms used).
  - Open Peer Review
  - Open Access
  - Open Educational Resources
- Other forms of communication and output (e.g. open access beyond traditional, scholarly peer-reviewed article…).
- **Impact and engagement with society? Science communication, citizen science, data-aware/open science aware society? This was a strong motivation of the Royal Society Report, for example.**

- ***Some of these may be contentious (open peer review, some aspects of open methodology…)***