euroCRIS Strategic Membership Meeting
Slovak Centre of Scientific and Technical Information (SCSTI)
Bratislava, Slovakia
21 November 2017

FAIR Data: principles, implementation, implications

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

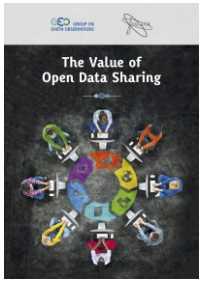
CODATA Prospectus:

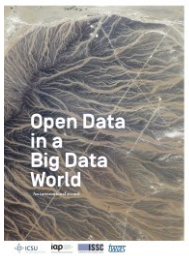
https://doi.org/10.5281/zenodo.165830

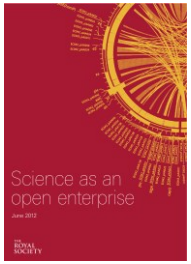
Principles, Policies and Practice

DC1 Data Citation Principles

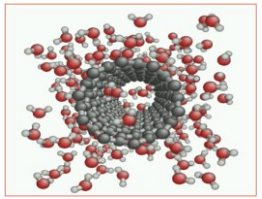
INTERNATIONAL COUNCIL FOR SCIENCE

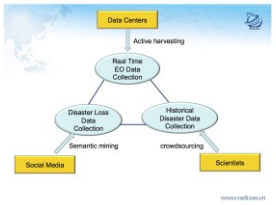






Frontiers of Data Science



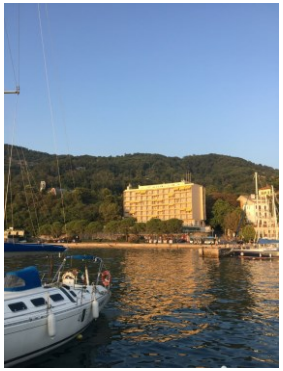


ubiquity press open scholarship

Data Science Journal

CODATA 2017, Saint Petersburg 8-13 Oct 2017

Capacity Building



中國科學院 CHINESE ACADEMY OF SCIENCES

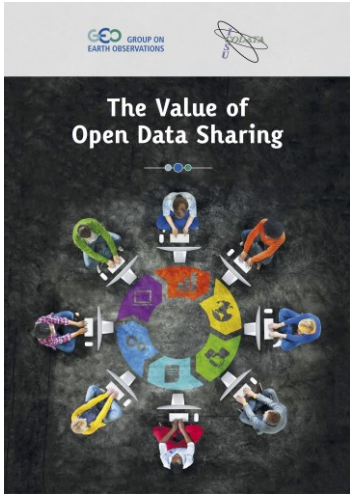
UNESCO

The Abdus Salam International Centre for Theoretical Physics

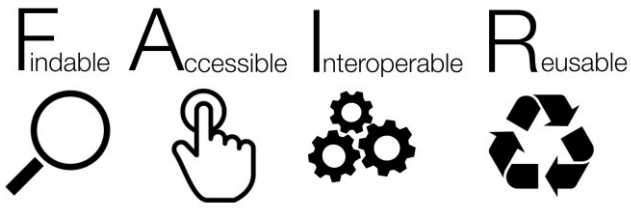
ICSU WORLD DATA SYSTEM

RDA RESEARCH DATA ALLIANCE

CODATA Data Policy Activities



- New Data Policy Committee, chaired by Paul Uhlir, international expert in Data Policies and member of CODATA Executive Committee.
- Current Best Practice for Research Data Management Policies http://dx.doi.org/10.5281/zenodo.27872
- The Value of Open Data Sharing, report for GEO http://dx.doi.org/10.5281/zenodo.33830
- Legal Interoperability, Principles and Implementation Guidelines https://doi.org/10.5281/zenodo.162241
- FAIR Data
  - Simon Hodson is chairing the European Commission's Expert Group on FAIR Data: http://bit.ly/FAIR_Data_Expert_Group
- OECD Global Science Forum and CODATA Project on Business Models for Sustainable Data Repositories: http://www.codata.org/working-groups/oecd-gsf-sustainable-business-models



OECD

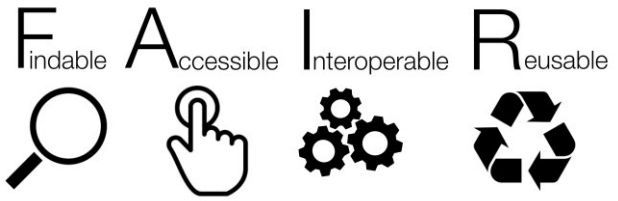
Why Open Science / FAIR Data?

- **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.

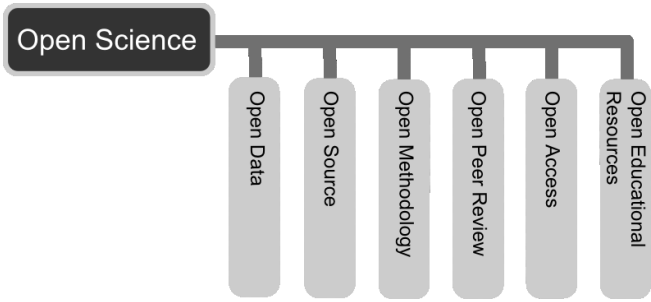
Policy Push for Open Research Data

- The three Bs (Budapest, Berlin and Bethesda) and Open Access, 2002-3
- OECD Principles and Guidelines on Access to Research Data, 2004, 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; Adoption of FAIR Data Principles, 2017.
- Science International Accord on Open Data in a Big Data World, Dec 2015: http://bit.ly/opendata-bigdata

Open Science

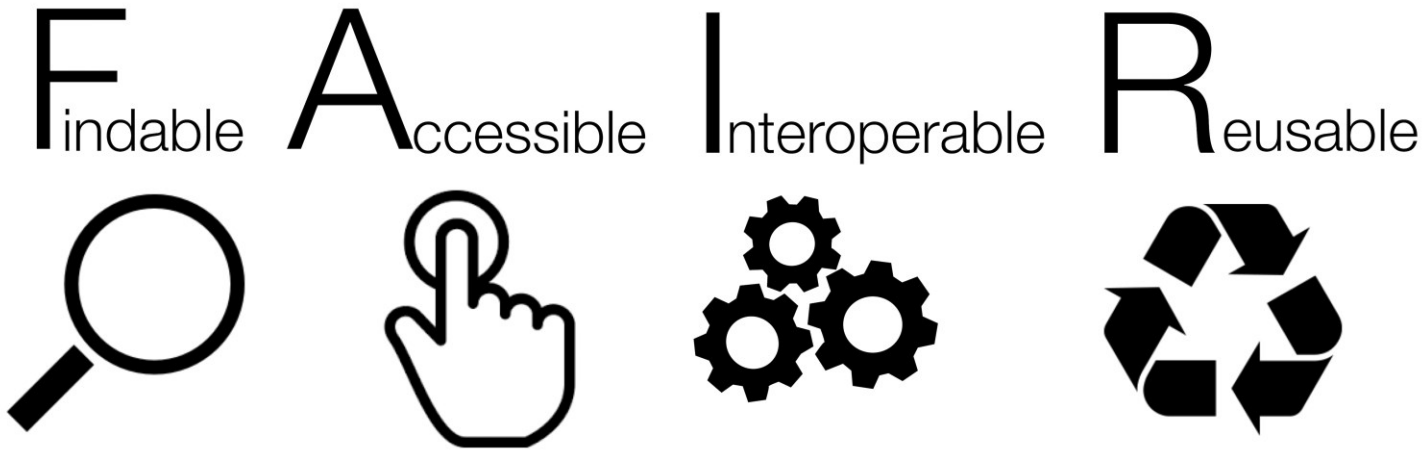


- What is Open Science:
  - Open access to research literature.
  - **Data that is as Open as possible, as closed as necessary.**
  - **FAIR Data (Findable, Accessible, Interoperable, Reusable).**
  - A shop window and repository of all research outputs.
  - A culture and methodology of open discussion and enquiry (including methodology, lab notebooks, pre-prints)
  - **Engagement with society and the economy in research activities (citizen science, co-design / transdisciplinary research, interface between research, development and innovation).**
- Research institutions have an opportunity to build their reputation around research specialisation: and this requires data specialisation and FAIR data collections.
- The Open Science ethos and co-design helps build collaboration between research institutions, government agencies and industry.



Emerging Policy Consensus? FAIR Data

- **FAIR Data** (see original guiding principles at https://www.force11.org/node/6062
  - **Findable:** have sufficiently rich metadata and a unique and persistent identifier.
  - **Accessible:** retrievable by humans and machines through a standard protocol; open and free by default; authentication and authorization where necessary.
  - **Interoperable:** metadata use a 'formal, accessible, shared, and broadly applicable language for knowledge representation'.
  - **Reusable:** metadata provide rich and accurate information; clear usage license; detailed provenance.



Attributes that give value to research data

- Builds on previous definitions…
- OECD Statement of Principles and Guidelines for Access Research Data: include a number of principles including accessibility, interoperability, quality, legal transparency, sustainability…
- Royal Society, 2012, *Science as an Open Enterprise*, Intelligent Openness: accessible, intelligible, assessable, usable.
- G8 Science Ministers' Statement, 2013, 'Open scientific research data should be easily discoverable, accessible, assessable, intelligible, useable, and wherever possible interoperable to specific quality standards.'

- FAIR Data now at the heart of H2020 policy, European Open Science Cloud etc.
  - **Under the revised version of the 2017 work programme, the Open Research Data pilot has been extended to cover all the thematic areas of Horizon 2020.**
- Current EC Guidance at http://bit.ly/EC_H2020_RDM_Guidance and http://bit.ly/EC_H2020_OpenData_Infographic

European Commission Expert Group on FAIR Data

- Chair of European Commission Expert Group on FAIR Data 'Making FAIR Data a Reality': http://bit.ly/FAIRdata-EG
- One of European Experts in ZA-EU Open Science Dialogue.
- Member of EOSC Advisory Board.

1. To develop **recommendations** on what needs to be done to turn each component of the FAIR data principles into reality
2. To propose **indicators** to measure progress on each of the FAIR components
3. To provide input to the proposed European Open Science Cloud (EOSC) action plan on how to make data FAIR
4. To contribute to the **evaluation of the Horizon 2020 Data Management Plan (DMP) template** and development of associated sector / discipline-specific guidance
5. To provide input on the issue of **costing and financing data management activities**

FAIR Guiding Principles (1)

- **To be Findable:**
  - F1. (meta)data are assigned a globally unique and persistent **identifier**
  - F2. data are described with rich **metadata** (defined by R1 below)
  - F3. metadata clearly and explicitly include the **identifier** of the data it describes
  - F4. (meta)data are registered or **indexed in a searchable resource**
- **To be Accessible:**
  - A1. (meta)data are retrievable by their **identifier** using a standardized communications protocol
  - A1.1 the **protocol** is open, free, and universally implementable
  - A1.2 the protocol allows for **an authentication and authorization** procedure, where necessary
  - A2. **metadata** are accessible, even when the data are no longer available

(Mons, B., et al., The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, http://dx.doi.org/10.1038/sdata.2016.18)

FAIR Guiding Principles (2)

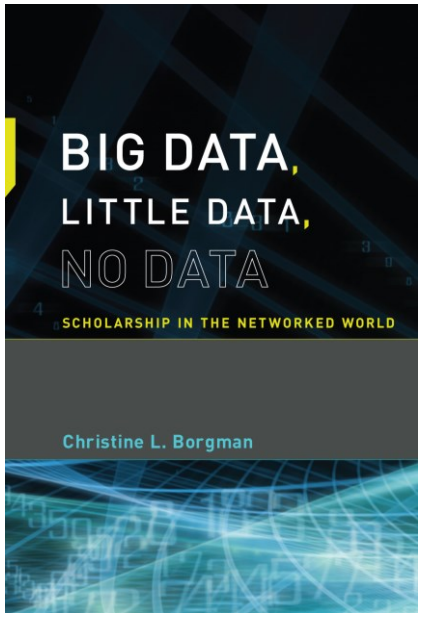
- **To be Interoperable:**
  - I1. (meta)data **use a formal, accessible, shared, and broadly applicable language for knowledge representation**.
  - I2. (meta)data use **vocabularies** that follow FAIR principles
  - I3. (meta)data include qualified references to other (meta)data
- **To be Reusable:**
  - R1. meta(data) are richly described with a plurality of accurate and relevant attributes
  - R1.1. (meta)data are released with a clear and accessible **data usage license**
  - R1.2. (meta)data are associated with detailed **provenance**
  - R1.3. (meta)data meet domain-relevant **community standards**

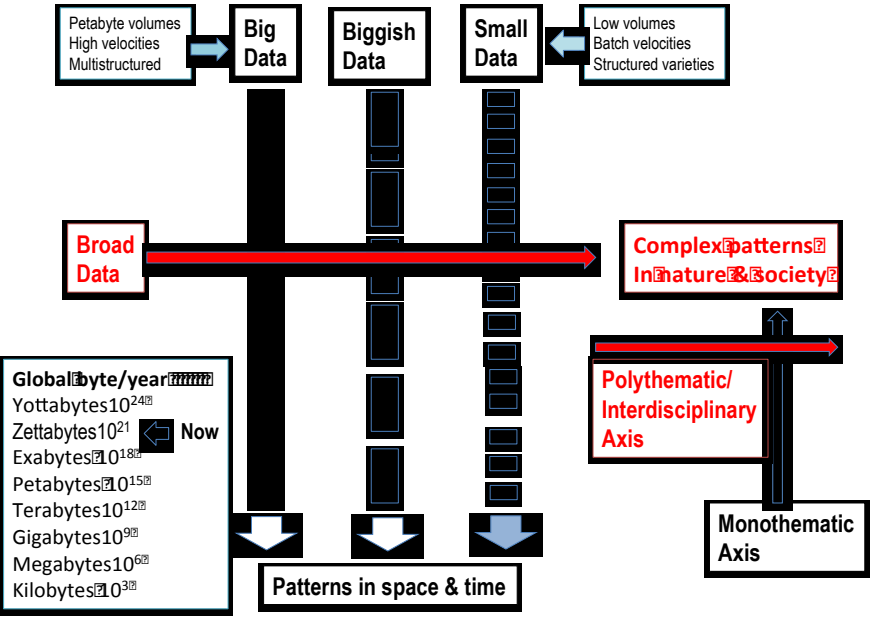
(Mons, B., et al., The FAIR Guiding Principles for scientific data management and stewardship, Scientific Data, http://dx.doi.org/10.1038/sdata.2016.18)

Open Science and FAIR Data: Implications and Directions

- Clarify the **boundaries of Open** for research data.
- Clarify and propagate **criteria and guidelines** for appraisal and selection.
- Refine and improve understanding of FAIR data, particularly I and R…
- Work with and across disciplines on **standards and vocabularies**, to help address I and R…
- Recognise that data stewardship must be regarded as part of the total cost of doing research.
    - Invest in **sustainable data infrastructure** (including repositories, stewardship, standards), and develop appropriate **business models** for sustainability.
- **Incorporate research data in the process of scholarly communication** and ensure that researchers, research groups and institutions receive adequate reward and recognition for their efforts.
- Address the **skills** requirements for data scientists, data stewards, data liaisons and researchers themselves.

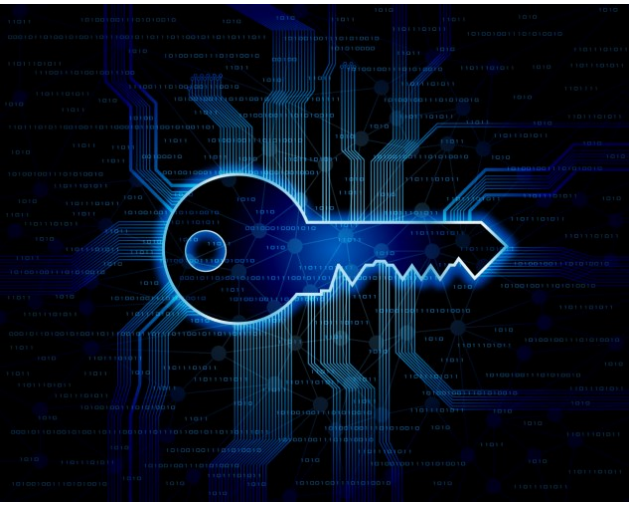
What are data? When are data?





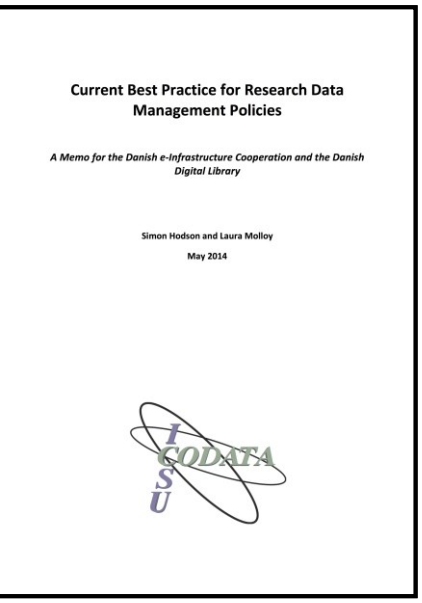
- Data is not the new oil. They are far more valuable because they are renewable.
- **Big Data** poses many challenges and opportunities: how do we manage it and develop the data science to extract information?
- **Small data** poses many challenges, because often they require detailed human-crafted metadata.
- Often still our challenge is a **lack of data**, no data.
- Data can be very varied: poses challenges of storage, management, analysis.
- Many disciplines have raw data, processed data, science ready data or data products.
- Important to understand which of these can and / or should be preserved and made available.

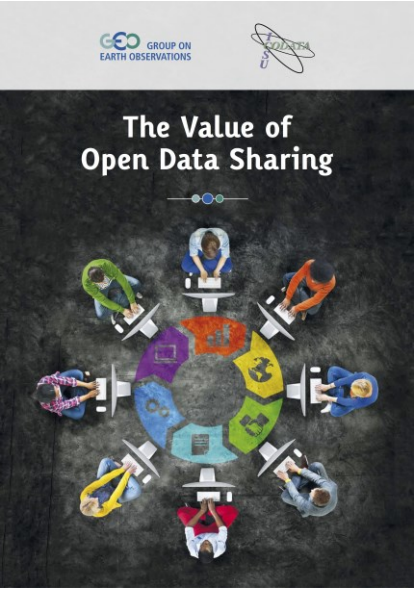
Boundaries of Open



- For data created with public funds or where there is a strong demonstrable public interest, **Open should be the default.**
- **As Open as Possible as Closed as Necessary.**
- **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!
- **A great deal of data is not affected by these issues and can and should be open.**
- **There is a need to evolve policies, practices and ethics around closed, shared, and open data.**

Research data, public data…



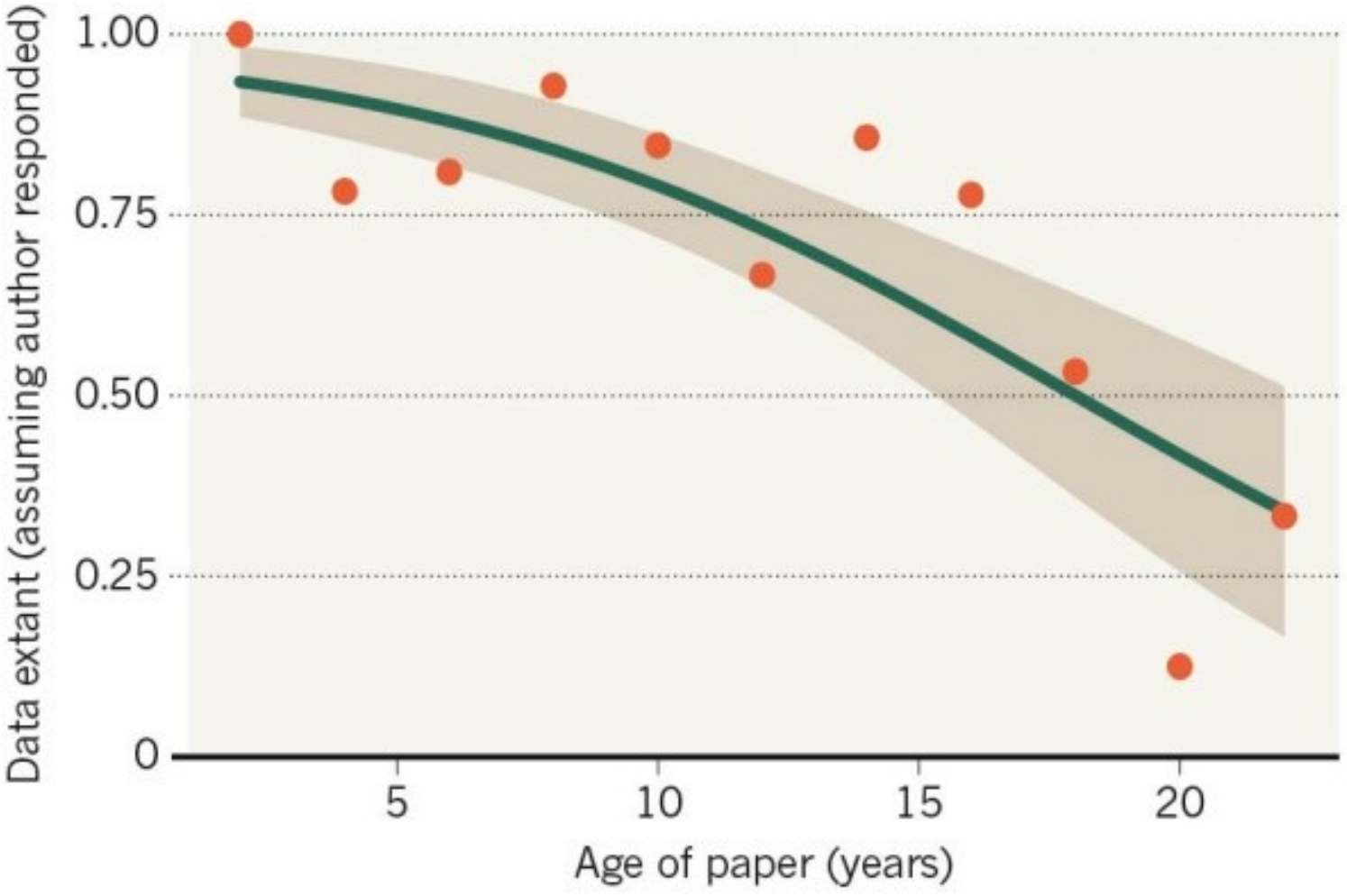


- **Typology of data that are in scope of research data policies:**
  - Data underpinning research conclusions presented the literature must be published.
  - Significant data produced by research projects should be made available where possible.
  - Data resulting from major data creation projects for monitoring and research.
  - Data created by public activities where there are research and development opportunities.
  - Private sector data where there is a public good case or mutual benefit in data sharing.
- Research data should be as open as possible, as closed as possible.
- All research data should be FAIR.
- Framework for development of data policies; will be updated with additional resources in next 12 months by CODATA Data Policy Committee.

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)

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

Slide credit: Dryad

Criteria for appraisal, selection, preservation?

- Need for more work on criteria and guidelines for appraisal, selection, preservation.
- A lot of the criteria has to come from the disciplines, but informed by some general principles.
  - Major data collection/creation exercises with **evident multiple uses** (census, Hubble etc).
  - Unrepeatable observations, measurements in nature or society? **(preserve and publish)**
  - Data created by in vitro experiments that can be reproduced and for which the instruments are being improved **(perhaps very limited reasons for preserving)**
  - Data collected for a given public or private purpose (traffic management, customer relations, ships logs) but which could be used for research…
- **Need for collaborative approach (with researchers) to clarifying the criteria to keeping, publishing and discarding data.**
- **While being mindful of potential for other extra- and inter-disciplinary uses.**
- Qualified by the knowledge that some disciplines have to discard data because of the sheer volume.
- What is important is that we start actively exercising these processes.
- DCC Guidelines: http://www.dcc.ac.uk/resources/how-guides/appraise-select-data
- NERC Data Value Checklist: http://www.nerc.ac.uk/research/sites/data/policy/data-value-checklist/

Interoperability and Reusability

Findable
Accessible
Interoperable
Reusable

- The two most challenging areas of FAIR…
- Major challenges to clarify and unpack **interoperable** and **reusable**.
- **Need to build on knowledge from archival community: what makes a resource usable?**
- **OAIS Reference Model:**
  - '**Preservation Description Information** is divided into five types of preserving information called Provenance, Context, Reference, Fixity and Access Rights'
  - **Independently Understandable:** A characteristic of information that is sufficiently complete to allow it to be interpreted, understood and used by the Designated Community without having to resort to special resources not widely available, including named individuals.
  - **Designated Community**: An identified group of potential Consumers who should be able to understand a particular set of information. The Designated Community may be composed of multiple user communities. A Designated Community is defined by the Archive and this definition may change over time.

CODATA WG on Description of Nanomaterials

Convene ISUs, International Stakeholders and data experts

Form Working Group

Draft Framework for Description of Nanomaterials

Refine/validate in FP7 Future Nano Needs Project

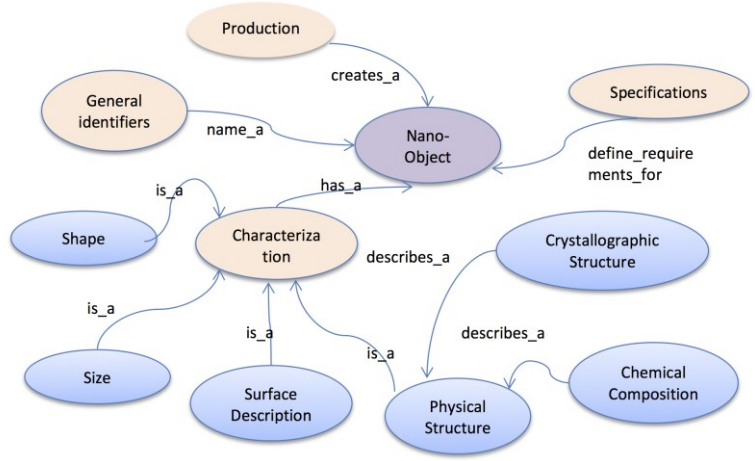


Figure 4. Information categories for describing an individual nano-object

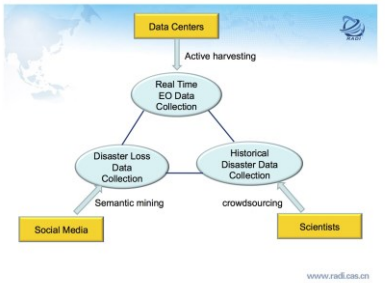
CODATA WG on the Description of Nanomaterials:
http://www.codata.org/nanomaterials

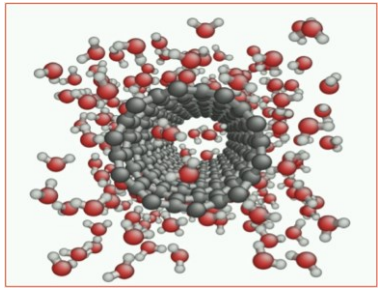
Uniform Description System v.02, May 2016:
http://dx.doi.org/10.5281/zenodo.56720

Future Nano Needs Project:
http://www.futurenanoneeds.eu/

Interdisciplinary Research and Use Beyond the Designated Community

- Major transdisciplinary research issues depend on the integration of data and information from different sources.
- Fundamental importance of agreed vocabularies and standards.
  - Fundamental to integration of social science, geospatial and other data
  - Essential to effective interface of science and monitoring (e.g. Sendai, SDGs, sustainable cities)
  - LOD for Disaster Research, Nanomaterials Uniform Description System
- Huge opportunities but significant challenges.
- The ICSU and ISSC, any merged Council, and international scientific unions could have a major role to play to encourage and accelerate these developments.





ICSU-CODATA Commission on Data Standards for Science

- **'Inter-Union Workshop on 21st Century Scientific and Technical Data Developing a roadmap for data integration', Paris, 19-20 June: http://bit.ly/codata_standards_workshop**
- Representatives of International Scientific Unions: IUCr for CIF; IUPAC for chemical terminologies; IUGS for GeoSciML; etc.
- Representatives of Standards Organisations: e.g. Darwin Core, for biology, biodiversity; DDI for social science surveys; OGC for geospatial data; W3C for the web.
- Position paper available at http://bit.ly/data-integration-paris-workshop-position-paper
  - Directory of activities involving international scientific unions.
  - Maturity model for vocabularies and standards.
  - Case studies of applications of vocabularies and standards for transdisciplinary research.

ICSU-CODATA Commission on Data Standards for Science

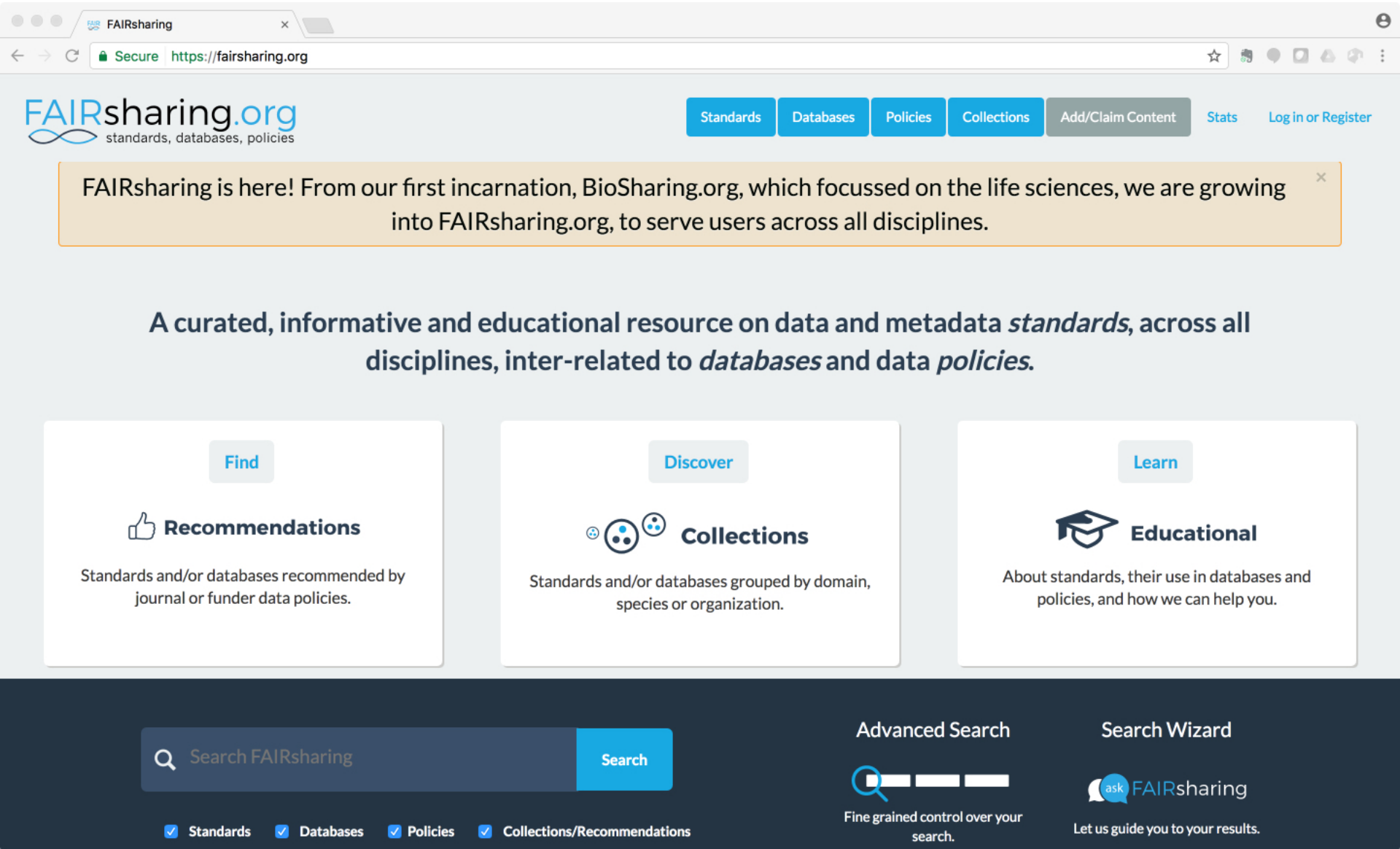
- **Larger follow-up workshop 13-15 November, Royal Society, London.**
- Identified three pilots to explore interoperability and reuse in interdisciplinary research: **reducing disaster risk; infectious disease; sustainable and resilient cities.**
- Vision of a decadal initiative to advance science through integration of data and information.

Metrics for Standards and Vocabularies

- Someone's text list (brain fart?) on the web
- Machine-readable lists
- Non-proprietary lists of words with simple definitions
- Concept-based, community definitions, RDF, governed
- Concept-based, RDF, linked, endorsed, multilingual

Slide credit: Lesley Wyborn

FAIRsharing



The Challenge: Business Models for Sustainable Data Repositories

- Research increasingly relies on digital repositories, curated databases and services over data.
- Research funder policies increasingly mandate data stewardship of data produced by funded projects.
- 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.
  1. **Essential that data stewardship and FAIR data is considered an essential part of the cost of doing research.**
  2. **Essential to demonstrate the value proposition of data repositories / data services.**
  3. **Sustainability is not just about whether something is funded, but how it is funded: what are the most effective and sustainable mechanisms for funding?**

The Background and Context

- Relatively little work has been done to understand the economics and business models of data infrastructure, let alone the funding mechanisms.
  - Blue Ribbon Task Group Report on Sustainable Digital Preservation: http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf
    - **Need to understand value proposition for communities.**
  - Sustaining Domain Repositories for Digital Data: A White Paper (ICPSR): http://datacommunity.icpsr.umich.edu/sites/default/files/WhitePaper_ICPSR_SDRDD_121113.pdf
    - **Need to understand how repositories are funded.**
- OECD Project builds on previous work of RDA-WDS Interest Group: 'Income Streams for Data Repositories': http://dx.doi.org/10.5281/zenodo.46693

Project Questions and Team

- How are data repositories currently funded?
- How do the different mechanisms for funding (income streams) affect sustainability?
- What innovative income streams are available to data repositories?
- What means of optimising costs are available?
- How do income streams match willingness/ability to pay of various stakeholders?
- How do income streams/willingness to pay fit together into a sustainable business model?

- Co-Chairs of OECD Project: Ingrid Dillo, DANS; Simon Hodson, CODATA.
- Consultant and writer: Paul Uhlir, previously NAS.
- Economics consultant: John Houghton, author of many studies of economics of data infrastructure.
- OECD GSF Expert Group.
- Invited participants of November 2016 and March 2017 Workshops.

Project Activities and Programme

- **Detailed interviews** with 47 data repositories funding and business models, conducted by members of the Expert Group and by OECD staff.
- **First workshop, Nov 2016:** SWOT analysis of different business models, discussion of cost restraint.
- **Economic analysis, Nov 2016-Feb 2017:** conducted by John Houghton and discussed with the project team, co-chairs and Expert Group. Identified strengths and weaknesses of particular revenue streams, frameworks for business models and recommendations. Draft report circulated prior to workshop.
- **Second workshop, March 2017:** presentation of economic analysis, discussion of strengths and weaknesses of different business models and of draft recommendations with stakeholders.
- **Redrafting of report, April-July 2017:** final iterations and refinement of report and recommendations.

Final Report: Business Models for Sustainable Data Repositories

- **Report summarizing findings and analysis, key features:**
  - Improved understanding of mechanisms for funding research data repositories.
  - Criteria to assess the sustainability of business models for research data repositories in different circumstances.
  - Summary of the pros and cons of different business models.
  - Insights into the development of business models for research data repositories.
  - Recommendations that primarily target policymakers and funders in OECD member states, as well as repository operators and managers.

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.
- **Springer Nature initiative to standardise policies: http://www.springernature.com/gp/group/data-policy/policy-types**

- **FAIRsharing https://fairsharing.org ;**
- **RDA Interest Group to encourage development and adoption of journal data policies**

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**

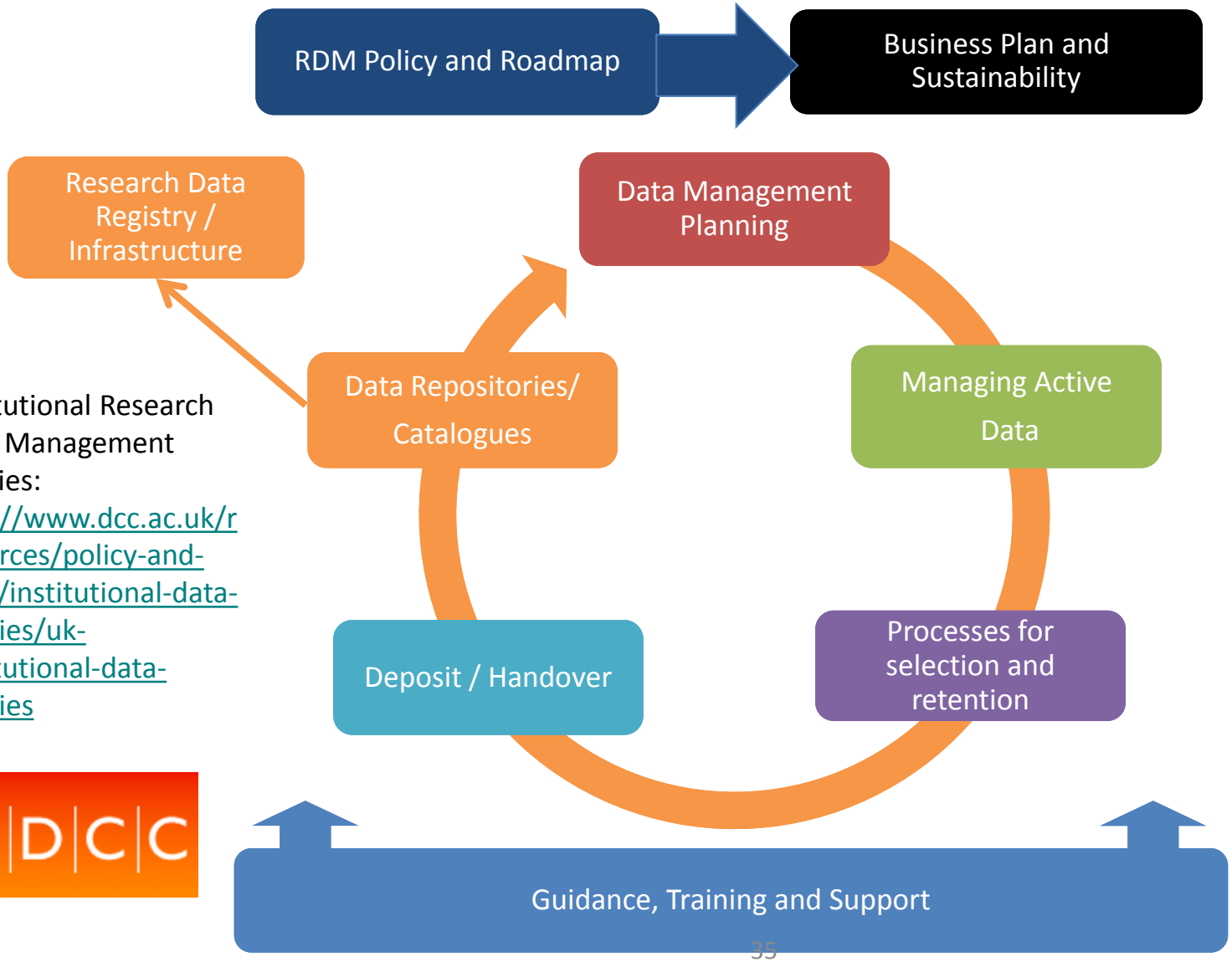
Framework for Regional, National and Institutional Data Strategies

- **National / Institutional Open Science and FAIR Data Strategy**
  - Consultative forum, stakeholder engagement.
- Open data **policies and guidance** at national and institutional level.
  - Clarify the **boundaries of open** (particularly privacy, IPR).
  - Clarify the data in **scope**, guidelines on selection.
- Develop **incentives and reward** systems.
  - 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**:
  - Essential **data skills for researchers**.
  - Develop skills and competencies for **data stewards, data scientists**.

Framework for Regional, National and Institutional Data Strategies

- Scope, roadmap and implement data **infrastructure**.
  - Key **components of national and regional infrastructure** (network / NREN, economies of scale for storage and compute).
  - Development **of regional, national and institutional infrastructure(s)** for research collaboration and data stewardship/RDM, generic research platforms/environments, trusted digital repositories.
  - **Collaborative infrastructures** for certain research disciplines, nationally, regionally to pool expertise and lower costs.
  - **International infrastructure / data ecosystem components**: permanent identifiers, metadata standards.

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

Göttingen-CODATA Symposium
18-20 March 2018

- **The critical role of university RDM infrastructure in transforming data to knowledge**
- http://conference.codata.org/2018-Goettingen-RDM/
- An opportunity to share experiences, research and insights in the development implementation of RDM services in research institutions.
- Special collection of Data Science Journal.
- Themes: services and solutions; strategy; measuring success; skills and support; sustainability; shared services and outsourcing / consortiums; service level, trust and FAIR; champions and engaging with researchers.
- Announcement of call for papers in the next week or so.
- Deadline for abstracts, 15 December.

GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

DataTrieste Film on Vimeo
https://vimeo.com/232209813

The Abdus Salam International Centre for Theoretical Physics

INTERNATIONAL DATA WEEK
IDW 2018

Gaborone, Botswana: 22–26 October 2018

CODATA

ICSU WORLD DATA SYSTEM

RDA RESEARCH DATA ALLIANCE

UNIVERSITY OF BOTSWANA

Digital Frontiers of Global Science

Frontier issues for research in a global and digital age.

Applications, progress and challenges of data intensive research.

Data infrastructure and enabling practices for international and collaborative research.

Data, development and innovation: data as an interface between research, industry, government, society and development.


Botswana, Africa and the World!

Stable, safe, modern and exciting country

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

Data is difficult: motivations and reward

- Open and FAIR data is essential for transparency and reproducibility; to take advantage of analysis at scale; to tackle major interdisciplinary challenges that require integration of data from many resources; has significant economic and other societal benefits…
- **But…**
- Research funders and research performing institutions will have to invest in data infrastructure.
- Essential to consider the cost of data stewardship and dissemination as part of the total cost of doing research.
- Data description, definitions and ontologies, data management require significant effort.
- Requires data skills, motivation and reward.
- Data should be integrated more with the process of scholarly communication and recognition of research contribution: **data citation** and journal availability policies; recognition for making available major datasets.
- **RPOs and research groups will increasingly build prestige on the basis of their data collections: research intensive institutions will be data intensive institutions.**

What do we mean by data citation / referencing

1. **Citation as reference: a citation is the means by which we indicate in scholarly work that a given assertion relies upon a given source. It points the reader or the readers' code to the evidence.**
   a) Pointing to the date, where appropriate accessible via a trusted digital repository, is good research practice.
   b) Assists verification, reproducibility, reuse.

- Desideratum: incorporate data more effectively into the process of scholarly communications so data can be found, accessed, verified, reused…

2. **Citations as a means of crediting research contributions.**
   a) Fundamentally a form of attribution.
   b) Citations as credit exert considerable influence over researcher motivations both for recognition and reward (career advancement).
   c) Data citations are not currently used in metrics around research contributions.

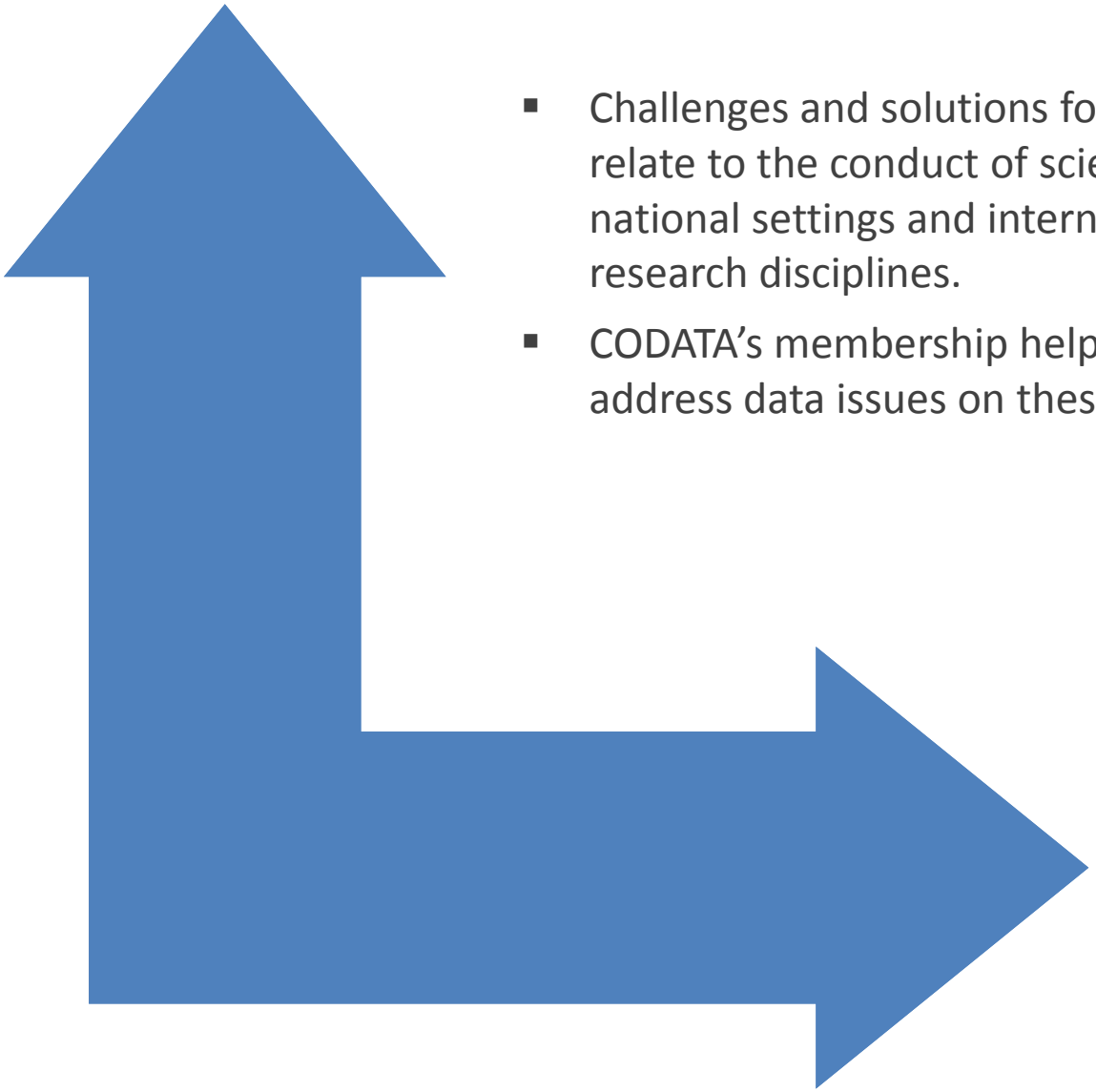
- Desideratum: encourage data citations in order better to give credit to those who produce, share, annotate data

Research Data: challenges and stakeholders

National Research Systems

CODATA National Members

National Academies of Science or Data Organisations



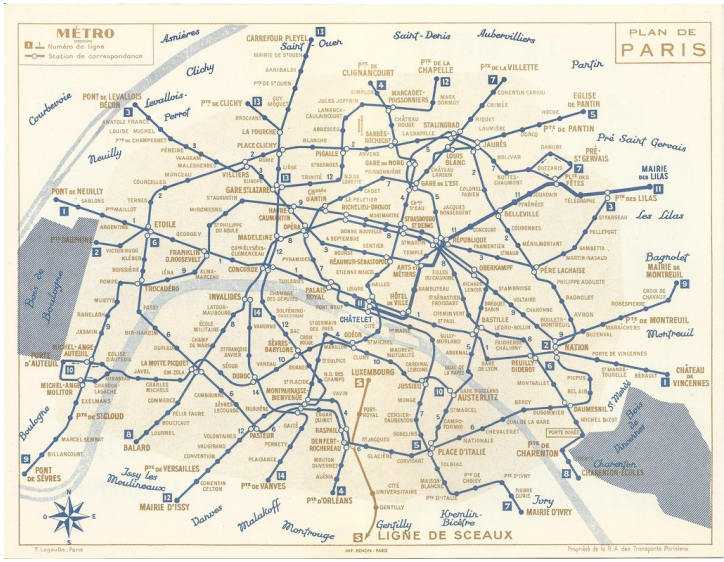
- Challenges and solutions for data issues relate to the conduct of science in national settings and international research disciplines.
- CODATA's membership helps us to address data issues on these two axes.

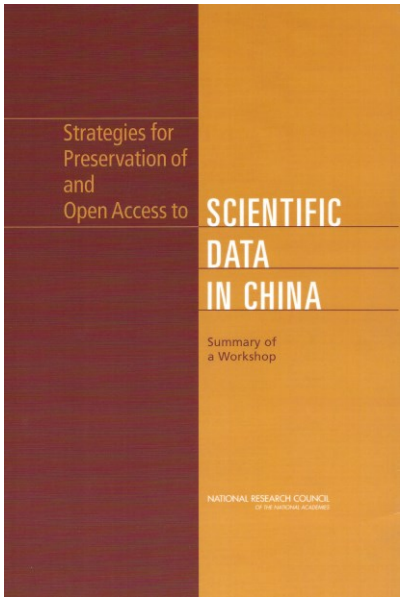
Scientific Disciplines

CODATA International Scientific Union Members

Role of CODATA National Committees

- **Join CODATA and form a National Committee**.
  - CODATA membership dues are aligned with GDP.
  - CODATA National Committees are composed of national stakeholders and data experts.
- **What are the benefits of having a CODATA National Committee?**
  - **Engage:** point of contact with CODATA;
  - **Influence:** contribute to CODATA strategy;
  - **Coordinate: forum by which national stakeholders may advance data agenda in step with international developments;**
  - **Collaborate:** propose Task Groups, host or participate in international workshop series, engage with Early Career Data Professionals Group;
  - **Partner:** undertake activities with other National Committees, bilaterally or in groups.





Leverage

- **From September 2015 to September 2016, the annual income from membership fees of c.€205K leveraged further investment in activities to a total of over €1.9M: a leverage ratio of over 9.6:1**. This estimate includes external contributions to events, Task Groups and similar activities, sponsorship obtained as well as host and participant investment in events. As a specific example, in August 2016, the CODATA-RDA School of Research Data Science was held at ICTP in Trieste. CODATA's own investment in the event totals c.€10K in travel and student support. The event as a whole leveraged an additional c.€270,000 in support, comprising international and local travel and accommodation for experts and students as well as sponsorship and local expenses. **It should be noted that this estimate considerably undervalues the CODATA's leveraging power as it does not include any estimate for contributions in kind (e.g. co-chairs time).**

- **Next two years, concerted outreach to expand membership and to engage more with National Committees.**