Continuing to Build Trust in Scientific Data: An Update from the World Data System

Rorie Edmunds
WDS Acting Executive Director

Third International Polar Data Forum
Hosted by the Finnish Meteorological Institute
Dynamicum, Helsinki
18–22 November 2019
International Science Council

- On 4 July 2018, ICSU and ISSC merged as the 'International Science Council', a global non-governmental organization representative of both the Natural and Social Sciences
  - 141 Member Organizations (National scientific bodies)
  - 39 Member Unions and Associations
  - 30 Affiliated Members
World Data System of the International Science Council

Objectives

• Enable universal and equitable (full and open) access to quality-assured scientific data, data services, products, and information

• Ensure long-term data stewardship

• Foster compliance to agreed-upon data standards and conventions

• Provide mechanisms to facilitate and improve access to data and data products
1. Improve the trust in and quality of open Scientific Data Services

2. Nurture active disciplinary and multidisciplinary scientific data services communities

3. Make trusted data services an integral part of international collaborative scientific research
WDS Members

Trustworthy Scientific Data Services

- Capture and storage
- Curation
- Long-term preservation
- Discovery, access, retrieval
- Aggregation, analysis and/or visualization
- Associated legal frameworks
## WDS Members (Nov 2019)

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>81 Regular</td>
<td>Data stewards and/or data analysis services: certified Trustworthy Data Repositories</td>
</tr>
<tr>
<td>11 Networks</td>
<td>Umbrella bodies of data stewards: accredited Trustworthy Data Networks</td>
</tr>
<tr>
<td>11 Partners</td>
<td>Contribute support to WDS membership</td>
</tr>
<tr>
<td>20 Associates</td>
<td>Interested in the WDS endeavour</td>
</tr>
</tbody>
</table>
Geographical Coverage (Nov 2019)

WDS Regular and Network Members (10/2019)

Regular Members
Network Members*

* Note that Network Members often act as international organizations. Only the location of the Member's secretariat is shown here, and WDS coverage extends to regions not marked.
Disciplinary Coverage
(Oct 2019)
Regular Member Accreditation

CoreTrustSeal Trustworthy Data Repository Certification

- Expression of Interest
- **CoreTrustSeal Certification**
- Agree to WDS Bylaws
- Regular membership approved by WDS-SC
- Sign Letter of Agreement
- Renewal every three years
CoreTrustSeal

Data Seal of Approval Certification of Trusted Data Repositories

Research Data Alliance Repository Audit and Certification DSA–WDS Partnership WG

WDS Certification of Regular Members

WORLD DATA SYSTEM
CoreTrustSeal 101

- Minimally intensive process via an online application containing 16 Requirements
- Evidence a data repository is trustworthy given by supplying URLs of (ideally public) documents
- Applications reviewed by two community peers, taking into account repository’s specific aims and context (1000 EUR administration fee required)
- Successful applications are made publicly available
- Renewal is every three years
- 70 Seals Awarded as of 15 November 2019
CoreTrustSeal Requirements

• Background information

• Organizational infrastructure
  o Mission/scope, Licenses, Continuity of access…

• Digital object management
  o Integrity and authenticity, Appraisal, Documented storage procedures…

• Technology
  o Technical infrastructure, Security
CoreTrustSeal Review 2019

• Define the Requirements for 2020–2022
• Focus on feedback from applicant, communication and outreach activities, and an open comment period from 1 March 2019 to 31 May 2019
• Each suggestion examined and accepted, rejected, or deferred, with a brief explanation added
• Accepted feedback incorporated into revisions of the Requirements, Extended Guidance, and Glossary
• Comments often indicated overlap or the need for repeated evidence. Actual redundancies were removed or further guidance provided for clarity
CoreTrustSeal Review 2019

- Cross-references among Requirements were reviewed, validated, and simplified, with text standardized to reduce ambiguity
- Number, structure, and content of the Requirements remain fundamentally stable for 2020–2022
- Draft revision of the CoreTrustSeal Requirements made available on 6 August for comment and final version released on 15 November
- Applications against the revised Requirements will be accepted from 1 January 2020
International Technology Office

- Established April 2018 and hosted at University of Victoria by Canadian consortium of WDS Regular Members:
  - Ocean Networks Canada
  - Canadian Astronomy Data Centre
  - Canadian Cryospheric Information Network/Polar Data Catalogue
- ITO Associate Director, Karen Payne, appointed November 2018
ITO Remit

• Manage the contribution of WDS to operational systems that form part of Global Research Data Infrastructure (GRDI)

• Coordinate the development and integration of components of GRDI with other major players with operational roles

• Coordinate WDS contribution to technical working groups with partners
ITO Projects

- Survey of Members: Online with follow-up interviews
- Harvestable Metadata Services Working Group
- Schema.org: Semantic markup tool with built-in socialization mechanism for mappings; training on best practices
- Gov of Canada RDM mandate: $572.5 m to initiate Digital Research Infrastructure Contribution programme
WDS Achievements

• Conferences & Workshops
  o International Data Week
  o Asia–Oceania Conference
  o Latin America & Caribbean Workshop

• Task & Working Groups, and Projects
  o WDS–CODATA Citizen Science TGs
  o WDS–RDA SCHOLIX WG
  o WDS–RDA Data Fitness for Use WG
  o PARSEC
ISC Action Plan 2019–2021

Challenge Domains:

• The 2030 Agenda for Sustainable Development
• The Digital Revolution
• Science in Policy and Public Discourse
• The Evolution of Science and Science Systems
2.1 DATA-DRIVEN INTERDISCIPLINARITY
(Project in progress)

Many of the major contemporary problems faced by science and society are inherently complex. They concern the operation of systems that exhibit emergent behaviour as a consequence of interactions between their component parts. Some examples include the operation of cities, the human brain, the dynamics of infectious diseases, climate change and pathways to sustainability. Researching these challenges almost invariably requires interdisciplinary collaboration. The tools of the digital revolution, now enhanced by the techniques of artificial intelligence, have created unprecedented opportunities to exploit such collaboration by integrating relevant data from disparate disciplinary sources. The prospect of realizing Stephen Hawking’s prediction that “the next century [the 21st] will be the century of complexity”.

Yet our ability to combine datasets from heterogeneous sources and across disciplines remains limited in many instances and, at best, is excessively resource-intensive. The adoption of new data-intensive techniques across scientific communities and practices is uneven, and the annual effort required to prepare and clean the dataset before uses is a considerable diversion of scientific resources. Ontologies and vocabularies are often incompatible and sometimes quite inadequate to the task.

Addressing these problems is crucial; if we are to best extract the increasing quantities of diverse data to understand the complex systems that are at the heart of global challenges. Doing so will require the widespread adoption of replicable, generic approaches to data integration and FAIR (Findable, Accessible, Interoperable and Reusable) data standards in more science disciplines and interdisciplinary research areas. This is a crucial effort and its success will depend on active participation and engagement from all disciplines, including the social and human sciences, and by scientists from all parts of the world, including countries whose data science capacities may be limited.

ANTICIPATED IMPACT
More effective, evidence-based solutions for complex global challenges based on interdisciplinary collaboration enabled by data integration policies and practices across scientific fields and disciplines.

NEXT STEPS
Working with the support of the ISC, the Council’s Committee on Data for Science and Technology (CODATA) has been developing technology and semantic good practices for data interoperability and integration. Based on this initiative, a three-year-long programme is planned, comprising:

- Underpinning technologies and good practice for data integration that are applicable across a wide range of disciplines.
- Interdisciplinary case studies in global challenge areas: infectious disease, resilient cities and disaster risk reduction designed to contribute value in these areas but also to act as demonstrators of the value and importance of the approach in all areas of complex interdisciplinary science.
- Engagement with scientific unions and associations in programmes of work designed to promote progress across the disciplines of science that will enable interdisciplinary data interoperability.

Programme development will be led on behalf of the ISC by CODATA, working in partnership with the Council’s World Data System (WDS) and the Research Data Alliance (RDA). The ISC will work to promote membership engagement in ways that extend this approach to new communities of scientists and stakeholders, including from developing regions, where the open science platforms described under Domain Four could be key agents for these processes.

2.2 GLOBAL DATA RESOURCES AND GOVERNANCE
(Project in development)

It is 50 years since Tim Berners-Lee’s vision of universal connectivity and openness became the World Wide Web open and accessible to all. The culture of sharing and collaboration that it embodied has progressively expanded from open source software to open access publishing, open data and entirely open analysis, and now to the growing open science movement. This modern offering has much wider dimensions than open access publication and simply making data available. It extends to providing information on how to repeat or verify an analysis, exposing results that can be reused by others for comparison, comparison, or deeper understanding and inspiration. But it has even more ambitious and radical targets:

- Firstly, if we are to understand the complex systems that are at the heart of most global challenges (see project 2.1) by analysing the vast diversity of data that this involves, we need to have access to such data in an interoperable form. To achieve this would require a widespread ethos and practice of data sharing, not just within the publicly funded scientific community, but across the public and private sectors, including government, scientific publishers and international agencies.

- Secondly, there is now an opportunity, and arguably an obligation, to systematically open science and its outputs to a wider range of societal actors and to citizens in addressing shared problems and in the joint creation of actionable knowledge. For both purposes, stimulating the widespread availability and use of data resources, and securing their effective governance, are vital issues for 21st century science and for addressing today’s global challenges.

This is therefore a timely moment to consider the global data ecology, its governance, ownership, accessibility and usability in the data universe; to identify principles for the emergence of systems, protocols and commons from these typologies; and to explore how a federated global commons might develop and operate to the benefit of science. It is also vital to understand how global society’s data patrimony may be conserved when many scientific datasets are supported by short-term funding in the absence of a sustainable business model, maximizing affordable access to well-structured data in the interest of all scientific fields and communities, as well as funders, publishers and companies handling large amounts of data.

ANTICIPATED IMPACT
A global cross-sectoral coalition of support for principles and processes of data access, for the adoption of priorities for its federated governance, and for sustainable business models for key scientific datasets in a way that addresses the global scientific data market.

NEXT STEPS
The ISC will convene a group of technical experts and representatives of the principal data holding sectors to determine the scope and ambition of the project. An exploratory phase would consider
Thank you!

www.world-data-system.org

ipo@icsu-wds.org