CODATA (Committee on Data for Science and Technology)

Strategy and Achievement 2015-2016

The CODATA Mission

CODATA was created by the International Council for Science (ICSU) in 1966 with the mission 'to improve the quality, reliability, management, accessibility and use of data of importance to all fields of science and technology.' That mission has remained relevant through the 50 years of CODATA's existence. Rather than being one that can be delivered once and for all, it is a mission that requires perennial re-focusing to respond to changes in technology, in scientific understanding and in the organisation of scientific enquiry.

CODATA's greatest current challenge is to support the scientific community in enhancing efficiency, rigour and creativity in exploiting the digital revolution of the last decade. New processes of digital acquisition have generated an explosive deluge of data, posing severe challenges for data management and analysis, but also unprecedented opportunities for discovery. At the same time, ubiquitous and powerful means of dissemination of data and information offer novel opportunities for enhanced collaboration within the scientific community and creative engagement with stakeholders in broader society, within a framework of joint knowledge creation that has been termed 'Open Science'. These issues are fundamental to the future of science, to attainment of the Sustainable Development Goals, and to resolution of many global challenges.

CODATA is an international community that addresses these issues, and a means whereby national members benefit from and contribute to international action. CODATA activities currently draw in external funding that levers the contributions of its members by a factor of about 10.

The CODATA Strategy

The CODATA strategy[1] focuses on three priority areas that are essential to a coordinated international response to the digital revolution:

1. promoting the implementation of open data principles, policies and practices;
2. advancing the frontiers of data science and its adaptation to scientific research and raising awareness of data issues;
3. mobilising capacity by improving data skills and the functions of science systems needed to support open data (particularly in low and middle income countries - LMICs)

The CODATA strategy is delivered through the work of its officers and Executive Committee, through the activities of Task Groups (elected every two years by the General Assembly) and Working Groups (created by CODATA to meet strategic objectives and opportunities); through its National Committees; through its *Data Science Journal* and through collaboration with other organisations, including ICSU Regional Offices and its scientific union members.

---

[1] Mobilising the Data Revolution: the CODATA Strategy http://dx.doi.org/10.5281/zenodo.50343; Message from CODATA President, Geoffrey Boulton http://www.codata.org/message-from-president-geoffrey-boulton

1   Promoting data principles, policies and practices

1.1   The Science International Accord *Open Data in a Big Data World*

Several national reports have promoted the case for open data[2] both i) as a means of maintaining the rigour of scientific enquiry by ensuring that the evidence (the data) underpinning published scientific concepts is open to scrutiny (the basis of 'scientific self correction') and, ii) as a means of maximising the benefits of the digital revolution. It is now timely that the core principles of open data are articulated by internationally representative bodies of science to encourage coherence between national approaches and to stimulate widespread international adoption. CODATA led the working group for Science International (a collaboration between ICSU, the Inter-Academy Partnership - IAP, the International Social Science Council – ISSC, and The World Academy of Sciences - TWAS) that produced the Accord, which has now been endorsed by over 100 national academies, international scientific unions, university representative bodies and other analogous bodies (http://www.science-international.org/). Science International also supported the case for the Open Data Platforms described in section 1.7.

1.2   Best Practice for Research Data Management Policies

The CODATA Report on *Current Best Practice for Research Data Management Policies* (2015) http://dx.doi.org/10.5281/zenodo.27872, commissioned by the Danish e-Infrastructure Cooperation and the Danish Digital Library, provides the most comprehensive available survey of data policies. It has had considerable uptake and, for example, is being used extensively by Canada's International Development Research Centre's data management policy pilot to develop policies for its projects. This area is a major responsibility (together with 1.3) of the CODATA Data Policy Committee.

1.3 Embedding the Practice of Data Citation

The citation of data is an essential process in facilitating and motivating open data practices. Building on the 2013 CODATA report *Out of Cite, Out of Mind* http://doi.org/10.2481/dsj.OSOM13-043 and the 2014 Joint Declaration of Data Citation Principles http://bit.ly/data-citation-principles, the CODATA Task Group on Data Citation has organized an international series of workshops to promote the citation of data in research literature: http://bit.ly/data-citation-workshops. The Task Group coordinated the workshops to ensure a consistent format, bringing together international experts and national stakeholders (researchers, journal editors, scholarly societies, research institutions, funders, publishers and data repositories). Ten such workshops were held between August 2015 and December 2016. Each event was coordinated by relevant CODATA National Committees and other partners:

*   China, with CODATA China and the Chinese Academy of Sciences, 25 August 2015;
*   Australia, with the Australian National Data Service and the Australian Academy of Sciences, 28 October 2015;
*   Japan, with the ICSU World Data System and Japanese National Institute of Informatics, 29 October 2015;
*   India, with the Indian National Science Academy, 4-5 November 2015;
*   South Africa, with CSIR, DIRISA and the South African National CODATA Committee, 10 December 2015;
*   Israel, with the Israel Academies of Sciences and Humanities and the CODATA National Committee, 25 May 2016;
*   USA, with the National Academies of Science, Engineering and Medicine, 12 July 2016;
*   Russia, with the Geophysical Centre of the Russian Academy of Sciences, 21 July 2016;
*   Finland, with the Finnish National Open Science and Research Initiative, 23 November 2016;
*   Taiwan, with the National Taiwan University Library, 5 December 2016.

---

[2] For example, Royal Society, *Science as an Open Enterprise* (2012) https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/; National Academy of Sciences Ensuring the Integrity, *Accessibility, and Stewardship of Research Data in the Digital Age* (2009).

1.4 Implementation Guidelines for the Legal Interoperability of Research Data

Building on work done in support of the GEO Data Sharing Working Group, a joint CODATA-RDA Interest Group on Legal Interoperability http://www.codata.org/working-groups/legal-interoperability has developed a set of principles and implementation guidelines designed to minimize the obstacles to data reuse, caused by uncertain or conflicting rights, or by inappropriate restrictions on the use of research data. The guidelines have been refined through two strenuous rounds of peer review and have now been published on behalf of CODATA and RDA: https://doi.org/10.5281/zenodo.162241

1.5 Sustainable Business Models for Data Repositories

CODATA leads and coordinates an OECD Global Science Forum project to explore Sustainable Business Models for Data repositories http://bit.ly/OECD-CODATA-Sustainability_Models. The project was launched in April 2016 and builds on a landscape survey and SWOT analysis of *Income Streams for Data Repositories* https://zenodo.org/record/46693 (published March 2016) conducted by an RDA-WDS Interest Group with the CODATA Executive Director as co-chair. The project report will contain a survey of the funding landscape, examples of innovative income streams and options for cost optimisation, sample business models and the results of economic analysis and stakeholder acceptance testing, along with policy recommendations relating to options for sustainable business models for data repositories.

1.6 Open Data Sharing for the Group on Earth Observations (GEO)

CODATA has strongly supported the work of GEO by:

- providing co-chairs for the GEO Data Sharing Working Group, which, in 2015, developed updated Data Sharing Principles (2015) http://www.earthobservations.org/dswg.php.
- contributing to a GEO Task Force that developed a set of Data Management Principles (2015) http://www.earthobservations.org/dswg.php.
- producing a report for the GEO Secretariat on *The Value of Open Data Sharing* http://dx.doi.org/10.5281/zenodo.33830 (published Nov 2015). The report provides a concise but comprehensive survey of current evidence for the diverse benefits of open research data, with particular emphasis on earth observation data. It is designed to facilitate the process of transitioning from restricted data policies to more open policies for government data.

The CODATA Executive Director has been a member of the GEO Programme Board in 2015-16. CODATA commits itself to continue to support the work of GEO. This work is also complemented by the CODATA Earth and Space Science Data Interoperability Task Group which provides usage of modern and open data access standards and tools and helps to develop efficient ways to achieve inter-operability including consolidation of links between CODATA and ICSU WDS.
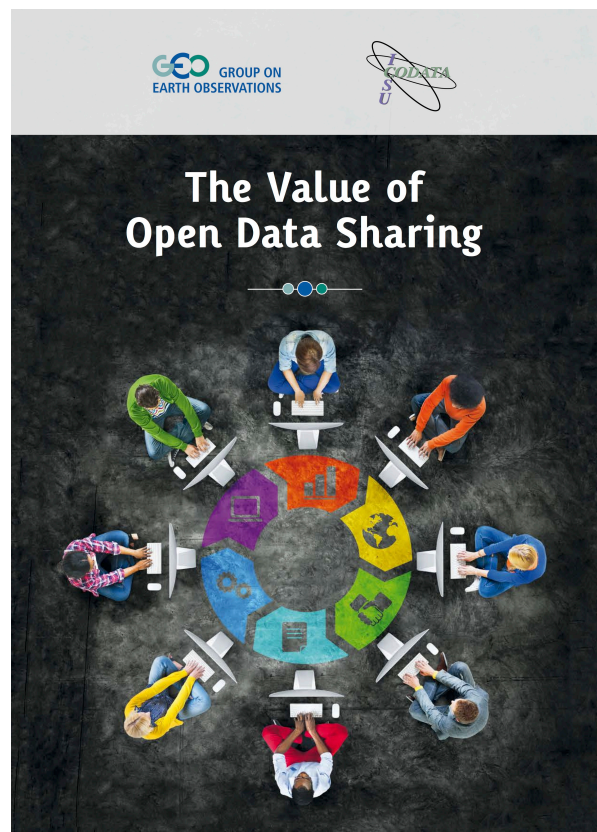
2   Advancing the Frontiers of Data Science

CODATA helps advance the science of data management, representation and analysis. As well as efficient processes of data management and manipulation, it is also vital that the processes of deriving meaning from data are statistically valid.

2.1 Data Science Journal

The *Data Science Journal* has been re-launched on an up-to-date platform, with an innovative Open Access publisher and with a new Editor-in-Chief, Sarah Callaghan. New articles are now appearing regularly and special editions are currently in preparation on 'Polar Data', 'Advances in Data Modeling and Knowledge Representation', 'Open Data and African Research', and '20 Years of Persistent Identifiers' http://datascience.codata.org/announcement/. The Data Science Journal will also publish a major collection of papers from SciDataCon 2016 http://www.scidatacon.org/2016/.

2.2 Major Meetings

2.2.1 SciDataCon and International Data Week

- SciDataCon 2014 (New Delhi, 3-5 November 2014) was the first of a planned series of major biennial conferences organised by CODATA and the World Data System (WDS).
- SciDataCon 2016 (Denver, CO., USA, 12-13 September 2016) is a major conference in the area of data science and data management http://www.scidatacon.org/2016/, attended by over 600 participants, and was embedded in International Data Week, http://www.internationaldataweek.org/ co-organised with WDS and RDA.
- CODATA and WDS have worked with Ubiquity Press to develop a new conference management platform. This will provide SciDataCon with a long-term, stable platform for the management and review of proposals and the hosting of the collection of abstracts, linked to the *Data Science Journal*.

2.2.2 International Symposia on Big Data for Science

In June 2014 an International Symposium on *Big Data for Science* was hosted by RADI (the Institute of Remote Sensing and Digital Earth of the Chinese Academy of Sciences), convened by the then CODATA President, Professor GUO Huadong. It is intended to follow this in 2016-18 with a symposium to consider the priorities for 'data science' and fundamental methodological and statistical issues in relation to Big Data.

2.2.3 Meetings organised by CODATA National Committees and Partner Organisations

CODATA national committees are active both in organising country-specific activities and international meetings. Similarly, CODATA collaborates with or contributes to international meetings organized by a number of important partner organizations.

- In 2014 CODATA China organized the '1st Scientific Data Conference – Scientific Research Big Data and Data Science' http://datascience.codata.org/articles/10.5334/dsj-2015-001/. This has now become a major annual activity of CODATA China.
- In November 2015, CODATA and the CODATA Task Group for Preservation & Access to S&T Data for Developing Countries (PASTD) convened side events at the GEO Plenary in support of the Data Sharing Working Group and the GEO Capacity Building initiative.
- In July 2016, CODATA USA organised an important international workshop on 'Data Citation: Developing Policy And Practice', in Washington DC http://bit.ly/CODATA-US-Data_Citation. It was hosted by the Board on Research Data and Information (BRDI) of the National Academies.

- In July 2016, CODATA was a strong participant in the Digital Earth Summit
  http://www.isde2016summit.org/. The CODATA President gave a keynote and CODATA organized a side
  event involving students from the CODATA China International Training Workshop.
- In July 2016, the Russian CODATA member, the Geophysical Centre of the Russian Academy of Sciences,
  organized a major conference on 'Data Intensive System Analysis for Geohazard Studies'
  http://sochi2016.gcras.ru/. CODATA contributed in a number of ways, including helping to convene a
  workshop on data citation. This event will be followed by a major CODATA Eurasia Conference in 2017.
- In August 2016 CODATA Germany organised an International Interdisciplinary Conference on Digital
  Cultural Evidence, with the objectives of raising awareness of the potential of Digital Cultural Heritage,
  promoting innovative content analysis and knowledge generation from massive digital data and creating
  innovative cross-discipline/cross sector partnerships.

2.3 Issues at Thematic Frontiers

These tend to be issues that have arisen from the
concerns of individual disciplines, but which have
general implications. They include:

2.3.1 CODATA Recommended Values of the Fundamental Physical Constants

A major, longstanding CODATA contribution to
science is its support for the international Task Group
that periodically provides a self-consistent set of
recommended values of the basic constants and

2014 CODATA RECOMMENDED VALUES OF THE FUNDAMENTAL CONSTANTS OF PHYSICS AND CHEMISTRY   NIST SP 959 (Aug 2015)

See: P. J. Mohr, D. B. Newell, and B. N. Taylor, arxiv.org/pdf/1507.07956v1.pdf (2015). A more extensive listing of constants is available in the reference given above and on the NIST Physical Measurement Laboratory Web site: physics.nist.gov/constants.

Quantity	Symbol	Numerical value	Unit
speed of light in vacuum	c, c_0	$299\,792\,458$ (exact)	m s^{-1}
magnetic constant	μ_0	$4\pi \times 10^{-7}$ (exact)	N A^{-2}
electric constant $1/\mu_0 c^2$	ϵ_0	$8.854\,187\,817... \times 10^{-12}$	F m^{-1}
Newtonian constant of gravitation	G	$6.674\,08(31) \times 10^{-11}$	m^3 kg^{-1} s^{-2}
Planck constant	h	$6.626\,070\,040(81) \times 10^{-34}$	J s
$h/2\pi$	$\hbar$	$1.054\,571\,800(13) \times 10^{-34}$	J s
elementary charge	e	$1.602\,176\,6208(98) \times 10^{-19}$	C
fine-structure constant $e^2/4\pi\epsilon_0\hbar c$	α	$7.297\,352\,5664(17) \times 10^{-3}$	
inverse fine-structure constant	α^{-1}	$137.035\,999\,139(31)$	
Rydberg constant $\alpha^2 m_e c/2h$	R_∞	$10\,973\,731.568\,508(65)$	m^{-1}
Bohr radius $\alpha/4\pi R_\infty$	a_0	$0.529\,177\,210\,67(12) \times 10^{-10}$	m
Bohr magneton $e\hbar/2m_e$	μ_B	$927.400\,9994(57) \times 10^{-26}$	J T^{-1}

conversion factors of physics and chemistry based on all of the relevant contemporary data. The latest
adjustment of these values is available at http://arxiv.org/abs/1507.07956v1 and
http://dx.doi.org/10.5281/zenodo.22826. A full-length article detailing the systematic review of this work will
be published later this year.

2.3.2 Uniform Description System for Materials on the Nanoscale v2.0

A complete revised version of the Uniform Description System (UDS) for Materials on the Nanoscale (UDS) was
released in June 2016 http://dx.doi.org/10.5281/zenodo.56720. Version 2.0 of the UDS is the result of
extensive consultation and discussion meetings convened by the CODATA-VAMAS Working Group on
Nanomaterials that substantially updated and extended Version 1.0 (February 2015). The UDS 2.0 contains 19
tables of detailed descriptors and their definitions that are directly applicable for reporting nanomaterials
research results, identifying nanomaterials in regulations and standards, developing formats for
nanoinformatics resources, specifying nanomaterials in commercial transactions, and other uses.

This work grew out of a process of wide
consultation and a series of international
workshops, supported by ICSU and
CODATA. The workshops engaged many
international scientific unions and
standards bodies and provide an example
of how CODATA can act as a convening
and coordinating body. The UDS has
been taken up by a number of
nanomaterials ontology and data
initiatives, most importantly the
European Commission-funded Future
Nano Needs project
http://www.codata.org/nanomaterials:
CODATA will play a major role in
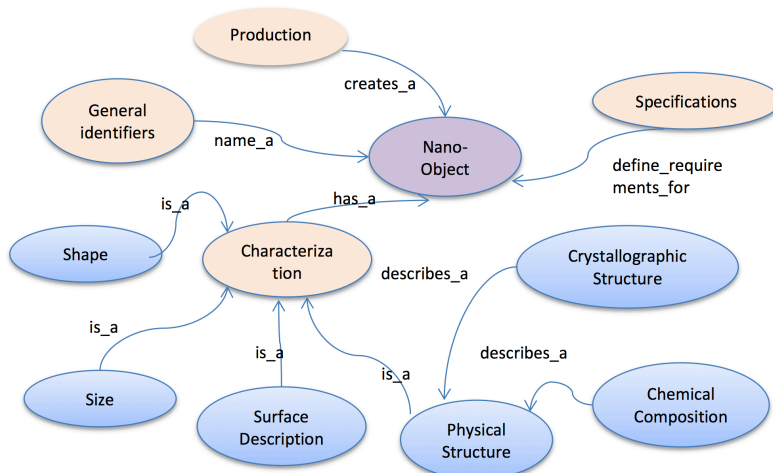organizing the project conference in 2017.

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

2.3.3 Standard Terminologies for Discovery, Use and Interchange of Data for Science and Technology

The need to share data, information and services is becoming ubiquitous and will require the development of standards for the discovery, access, sharing and retention of these data. Many, if not most, of the decisions about what to store, what standards to apply and what are the minimum required metadata lie, or should lie, with the relevant international scientific unions as the international representative bodies of given disciplines. Technical standards for the interchange of scientific data are well advanced, what are missing are domain-specific standards for much of the data collected as part of scientific research. Moreover, the development of compatible standards is vital for linked, interdisciplinary data, for example in programmes such as Future Earth. Whilst the development of these standards should be a major responsibility and concern of the international scientific unions, CODATA wishes to build on the success of the Nanomaterials Working Group, in collaboration with scientific unions, by helping create a Commission to coordinate standards development; to minimising duplication; to provide a web-accessible international standards repository for data models, standards, ontologies, and vocabularies; and to develop an up-dating process. The governance of such a project is under discussion with ICSU and several unions.

2.3.4 Other Thematic Activities

- **Linked Open Data for Global Disaster Risk Research Task Group** was created in response to the post-Sendai framework for disaster risk reduction, which calls for greater engagement of the scientific community, and for which the Task Group has produced a draft White Paper which is currently available for public review http://bit.ly/LOD-Disaster_Research-Gap_Analysis
- **Global Roads Data Development** is an initiative to maps of global roads systems from space on standardized coordinates on an open access basis, for which a version 2 Global Roads Open Access Data Set has just been released http://www.codata.org/task-groups/global-roads-data-development
- **Data at Risk Task Group.** Many sets of scientific data are not in modern electronic formats and are therefore not accessible to researchers that need them. The Task Group addresses these issues and promotes remedial work. A special issue of the Journal of Geophysical Research is dedicated to Data Rescue. See 'When Are Old Data New Data?', R.E.M. Griffin and the CODATA Task Group, 2015, GeoResJ, http://www.sciencedirect.com/science/article/pii/S2214242815000121. The Task Group has also organized a workshop on 'The Rescue of Data at Risk' prior to SciDataCon 2016 and International Data Week http://www.codata.org/task-groups/data-at-risk/dar-workshops

3 Mobilising Capacity for Data Science

Responsibilities for the scientific exploitation of the digital revolution and the implementation of processes for Open Data not only rest on the shoulders of individual researchers but also on those of the other components of national science systems. Capacity building must therefore be 'systemic', being developed at all levels of national science systems. Although science is an international enterprise it is done with the frames of distinctive national systems, such that capacity building needs to be framed by these systems and can only be effective with strong local ownership. This approach will also be implemented through the African Open Data Platform initiative described in 1.7. It also informs the discussions with other national bodies referred to above.

3.1 African Open Science Platform and Analogous Initiatives (see also 1.1)

As an outcome of the *Science International* meeting in Pretoria in December 2015, and in collaboration with ICSU, the Department of Science and Technology in South Africa and the South African Academy of Sciences, CODATA is launching a major open science policy and capacity mobilisation initiative for Africa. Seconded staff are now in place and the initiative will be formally launched at *Science Forum South Africa* in December 2016. Following a scoping survey of African initiatives and stakeholders, the project will, over three years, promote the development and adoption of data policies, address issues of incentives for good practice, foster training and capacity building activities and undertake a far-reaching data infrastructure strategic roadmapping exercise.

Allied to this initiative, CODATA is also involved in discussing analogous initiatives elsewhere:

- with South American institutions and the ICSU Regional Office for Latin America and the Caribbean, a meeting in August 2016 in El Salvador led to the creation of a working group to explore the possibility of a Latin American & Caribbean Open Data Platform.
- with the Polish Ministry of Science and Education in planning an Polish Open Science meeting to be hosted in Warsaw in early 2017, to create a Polish Open Science strategy.
- with the Chinese Ministry of Science and Technology's National Science and Technology Infrastructure Centre the possibility of a platform development.

3.2 Training for postgraduate and early career researchers

CODATA initiatives over the last two years have focused on training activities in data for PG and Early Career Researchers internationally, see: http://www.codata.org/events/workshops. The events include:

- **Training Workshop on Big Data for Science, Beijing, June 2014**, (two weeks, c.25 participants)
- **Training Workshop on Open Data, JKUAT, Nairobi, August 2014** (three days, c.70 participants)
- **Training Workshop on Big Data for Science, Indian Statistical Institute, Bangalore, March 2015** (two weeks, c.70 participants)
- **Training Workshop on Big Data For Science, LIPI, Jakarta, September 2015** (one week, c.50 participants)
- **Training Workshop on Big Data for Science, Beijing July 2016**, (two weeks, c.25 participants).

A policy workshop on Open Data for Science and Sustainability in Developing Countries, organized by CODATA's PASTD Task Group in August 2014 had important outcomes. It led to the creation of a Centre of Excellence for Open Science at the Jomo Kenyatta University of Agriculture and Technology.

The workshops in Bangalore and Jakarta are notable for having been organised by participants from previous CODATA Data Science Training workshops. CODATA works with the organizers to help facilitate the workshop, encouraging an alignment in the teaching materials around foundational data skills.

3.3 CODATA-RDA School of Research Data Science

Building on the above series of successful training workshops, CODATA is working with RDA on a coordinated initiative to establish an international, scalable and sustainable series of 'Research Data Science Short courses': see http://www.codata.org/working-groups/research-data-science-summer-schools. The vision is of a hub school, held annually at the International Centre for Theoretical Physics (ICTP) in Trieste, and a mechanism to support alumni and other partners in hosting short courses on 'Research Data Science' at their home institution. The model has already been tested successfully in previous CODATA schools and is used by Software and Data Carpentry. The CODATA-RDA short courses provide an introduction to the foundational data-related skills that all researchers need in the 21st century. The first CODATA-RDA School of Research Data Science took place in August 2016 and was attended by c.70 students: see the programme at http://bit.ly/School_of_Research_Data_Science-Programme and photos at http://bit.ly/School_of_Research_Data_Science-Photos.

These priorities are also supported by the CODATA Task Group on Preservation of and Access to Scientific and Technical Data in/for/with Developing Countries (PASTD), which promotes worldwide cooperation on research data and in developing open knowledge environments for international science and sustainability in developing countries. In addition to the targeted priorities identified above workshops and training sessions have been held in a number of countries including Brazil, Mongolia and Cuba.

Awareness Raising and Outreach

Awareness raising on key data issues and the role of CODATA has been promoted through:

- **Communication**

  Over the last three years, CODATA has substantially improved its website http://www.codata.org/ and social media presence (Blog http://codata.org/blog/; Twitter https://twitter.com/CODATANews; Facebook https://www.facebook.com/codata.org/) and its use of newslists http://bit.ly/CODATA-International-List

- **Webinars**

  CODATA held a very successful first webinar on the COINatlantic Data Accessibility Benchmark Organizational Self-Assessment Tool http://bit.ly/CODATA-COINatlantic-Webinar. We will continue and extend this series on data policy, data science and capacity building issues.

- **Publications Collection on Zenodo**

  Major CODATA outputs, reports and presentations from workshops etc, will increasingly be archived with Zenodo in order to assure that these resources will have a persistent identification and location as well as their longer term availability https://zenodo.org/collection/user-codata

- **Keynote Presentations**

  The President http://bit.ly/Boulton-Presentations and the Executive Director http://bit.ly/Hodson-Presentations have given a combined total of over 60 presentations in the last 18 months. Other officers and members of the Executive Committee have given at least 15 in the same period, in addition to those given by members of Task and Working Groups.

Return on Investment

Given the size and budget of the organization, the outputs and outcomes described above demonstrate a considerable return on investment.

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 CODATA's leveraging power as it does not include any estimate for contributions in kind (e.g. co-chairs' time).

CODATA President, Geoffrey Boulton; CODATA Executive Director, Simon Hodson with input from the CODATA Executive Committee, Task Groups and Working Groups, October 2016