



CODATA (Committee on Data for Science and Technology)

Strategy and Achievement 2015-2017

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 strategy¹ 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 data² 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 of the CODATA Data Policy Committee.

1.3 CODATA International Data Policy Committee

The CODATA International Data Policy Committee has been reformed and invigorated. Under the leadership of Paul Uhler, the Committee now boasts participation from a team of global experts. The first major area of activity will be to build on the guidance on best practice for research data management policies, updating the framework of data policy elements and testing this with stakeholders, and in particular international research communities. The Committee is also now actively preparing a series of Data Policy White Papers on the following issues: ‘Upstream-Downstream: Non-Profit Research Data and Commercial Innovation’, ‘Data Diplomacy’ and a ‘Model Data Policy Training Module’. The objective is to publish these in September ahead of International Data Week 2018.

1.4 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;



DC¹
Data Citation Principles

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).

- 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;
- Taipei, with the National University Library, 5 December 2016.

1.5 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>

1.6 Sustainable Business Models for Data Repositories

CODATA led and coordinated an OECD Global Science Forum project to explore Sustainable Business Models for Data repositories

http://bit.ly/OECD-CODATA-Sustainability_Models. The challenge of sustaining data repositories (including the majority that are not-for-profit) depends on making a sound value proposition as part of a solid business model that addresses the interests of stakeholders and the sustainability of the service. The final report was published in December 2017

<http://dx.doi.org/10.1787/302b12bb-en> and provides a survey of the funding landscape and sample business models, a discussion of innovative income streams and options for cost optimisation, and the results of economic analysis and stakeholder acceptance testing, along with policy recommendations relating to options for sustainable business models for data repositories.

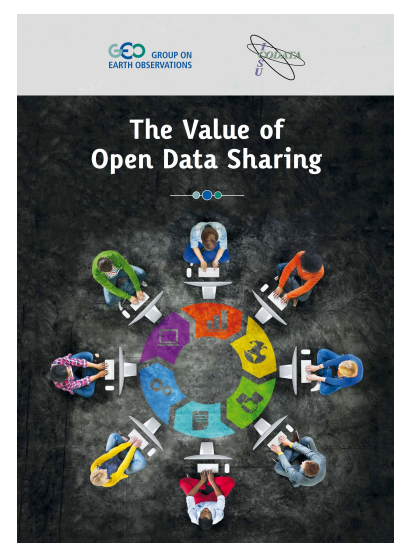


1.7 Open Data Sharing for the Group on Earth Observations (GEO)

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.

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 interoperability.



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* <https://datascience.codata.org/> was 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 collections have been published in a number of areas: 'Advances in Data Modeling and Knowledge Representation', '20 Years of Persistent Identifiers', 'Open Data and Africa' and a selection of papers from SciDataCon 2016 <https://datascience.codata.org/collections/special/>. Further special collections are in preparation.



The Data Science Journal is now indexed in Scopus and the rate of citation is growing steeply: <http://www.scimagojr.com/journalsearch.php?q=4700152809&tip=sid>

2.2 Major Conferences and Meetings

2.2.1 CODATA, SciDataCon and International Data Week Conferences

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 2017 'Global Challenges and Data Driven Science'

<http://codata2017.gcras.ru/> took place in Saint-Petersburg, Russia, from 8-12 October 2017. The Conference explored 'the fundamental issues relating to the availability, (re-)use and scientific analysis of data that relate to the most significant contemporary global challenges'. A special collection for the Data Science Journal is being prepared.

CODATA-Göttingen Symposium on RDM in Universities: In partnership with the University of Göttingen, CODATA is organising a major symposium on 'The critical role of university RDM infrastructure in transforming data to knowledge'. Taking place on 18-20 March 2018, this event will provide the opportunity for an up-to-date and thoroughly evidenced overview of current activity to improve research data management in universities. It will result in a major special collection in the Data Science Journal.

International Data Week 2018 will take place on 5-8 November in Gaborone, Botswana. This landmark event will address the theme of 'The Digital Frontiers of Global Science' <http://internationaldataweek.org/>. Once again the major event combines SciDataCon, the international research conference for all matters to do with data science, research data management and data stewardship, and the 12th Plenary Meeting of the Research Data Alliance. On the agenda will be the role of data in the sustainable development goals, the emergence of FAIR data, the evolution of scholarly communications, the need for data skills in education and - in sum - the many issues related to Open Science / Open Research, globally and in Africa.



2.2.2 Meetings organised by CODATA National Committees and Partner Organisations

CODATA National Committees are active both in organising country-specific activities and international meetings. An important mission of CODATA is to encourage and guide the development of national data activities and bilateral and multilateral collaboration between national members. A number of significant national and regional events have been organised in this way of the past few years. Similarly, CODATA collaborates with or contributes to international meetings organized by a number of important partner organizations: see the CODATA website on workshops and national members' pages for more information.

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 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.

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	$6.674 08(31) \times 10^{-11}$	$\text{m}^3 \text{kg}^{-1} \text{s}^{-2}$
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 a_0	a_0	$0.529 177 210 67(12) \times 10^{-10}$	m
Bohr magneton $eh/2m_e$	μ_B	$9.274 009 994(57) \times 10^{-26}$	J T^{-1}

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 is playing a major role in organizing the project workshops on classification of nanomaterials in Sept and Dec 2017.

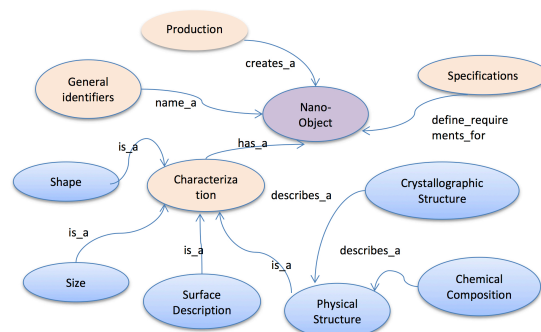


Figure 4. Information categories for describing an individual nano-object

2.3.3 The ICSU-CODATA Commission on Standards for Science

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. The development of compatible standards is vital for linked, interdisciplinary data, for example in programmes such as Future Earth.

To address this need, CODATA – in collaboration with international research programmes, standards organisations and scientific unions – is preparing a major initiative to create a Commission to help coordinate standards development; to minimising duplication; and to provide a web-accessible international standards repository for data models, standards, ontologies, and vocabularies. Scoping and planning workshops were held in June 2017 (ICSU, Paris) and November 2017 (Royal Society, London) with support from ICSU and NSF. **The initiative is now moving forward with three pilot case studies, examining the issues of data availability and interoperability in relation to infectious disease, disaster risk and resilient cities.**



2.3.4 Other Thematic Activities and Task Groups

CODATA currently has eight Task Groups <http://www.codata.org/task-groups> and five Working Groups <http://www.codata.org/working-groups> which actively address various significant data issues. Some Groups are created specifically to support CODATA's most important strategic activities: *Building Foundational Training in Research Data Science* and *Coordinating Data Standards amongst Scientific Unions*. Others address challenges that have arisen from the concerns of individual disciplines, or interdisciplinary research areas, but which have general implications: *Agriculture Data, Knowledge for Learning and Innovation; Citizen Science and the Validation, Curation, and Management of Crowdsourced Data* (with WDS); *Data Citation Standards and Practices*; *Earth and Space Science Data Interoperability*; *Linked Open Data for Global Disaster Risk Research*; and *Preservation of and Access to Scientific and Technical Data in/for/with Developing Countries (PASTD)*.

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 (see also 1.1)

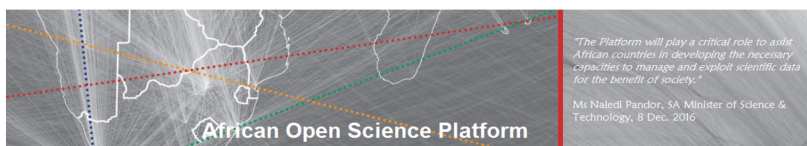
As an outcome of the *Science International* meeting in Pretoria in December 2015, and in collaboration with ICSU, the Department of Science and Technology in South Africa and the South African Academy of Sciences, CODATA is leading a major open science policy and capacity mobilisation initiative for Africa. The African Open Science Platform project <http://africanopenscience.org.za/> was launched at *Science Forum South Africa* in December 2016. Over three years, the project will:

- 1) promote the development and adoption of data policies;
- 2) address issues of incentives for good practice;
- 3) foster training and capacity building activities; and,
- 4) undertake a far-reaching data infrastructure strategic roadmapping exercise.

By the end of 2017, the project will have organised and collaborated a series of workshops to address policy, infrastructure and capacity issues in a number of countries.

- Launch workshop, the African Open Science Platform, Pretoria, Dec 2016.
- Open Science for African Universities Workshop, Association of African Universities Conference, Accra, Ghana, June 2017: http://africanopenscience.org.za/?page_id=52

- Open Data and Open Science for Sustainable Development Goals, Antananarivo, Madagascar, Sept 2017, including a High Level Meeting attended by 11 ministers, an international stakeholders workshop and a training event: <http://africanopenscience.org.za/?cat=6>
- Botswana National Open Data Open Science Forum, 30-31 Oct 2017, with support from the Botswana Tertiary Education Research Science And Technology.
- Open Science Infrastructure Workshop, UbuntuNet Connect Conference, Addis Ababa, Nov 2017.
- Meetings of Technical Advisory Board and Advisory Council, Science Forum South Africa, Dec 2017.



African Open Science Platform

By the end of 2018, the African Open Science Platform project will produce draft reports or white papers on the four pillars of activity as well as engaging with key participating countries and pan-African stakeholder groups (including universities, librarians, NRENs and collaborative research projects).

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).
- **Training Workshop on Open Data for Better Science, Beijing, July 2017** (two weeks, c. 25 participants).

A 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 led the development of 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 CODATA-RDA School of Research Data Science provides an introduction to the foundational data-related skills that all researchers need in the 21st century. The first such school took place in August 2016 and was attended by c.70 students. The programme has continued with a second foundational school in Trieste combined with more advance workshops (July 2017) and with the first regional foundational school at the ICTP Centre in São Paulo, Brazil (December 2017).

The reports, materials, photos, blogs and films (<https://vimeo.com/232209813>) about the School of Data Science are particularly informative about the approach, ethos and impact of the schools.

2018 will see a further expansion of the initiative. Data Schools will take place at ICTP sites in Trieste (August, followed by four advanced workshops), in Kigali, Rwanda (October) and São Paulo (December). The CODATA National Committee in Australia will also run a shortened version of the school in June.

3.4 GO-TRAIN Implementation Network

CODATA and the team behind the Data Schools has led the creation of a GO-TRAIN Implementation Network as part of the GO-FAIR initiative. In February 2017, CODATA convened at workshop for stakeholders and potential partners in the network which made great steps towards preparing an action plan for GO-FAIR training activities with the objective of developing a cohort of data stewards and data scientists as well as improving researchers' data skills. Subsequently, the group has been working with the new GO-FAIR office to prepare the manifesto for the GO-TRAIN Implementation Network: <http://bit.ly/GO-TRAIN-IN>



The Network will now work on the following priorities:

- 1) preparing a core data science and data stewardship curriculum framework and signposting for resources;
- 2) launching a train-the-trainer programme as the key initiative required to build the cohort;
- 3) develop and implementing a mechanism for endorsement and certification of training activities.



Advocacy, Dissemination 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 newsletters <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 80 presentations in the last two years. Other officers and members of the Executive Committee have given at least 20 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 2017