



# Big Data for International Scientific Programmes: Challenges and Opportunities A Statement of Recommendations and Actions Beijing 9 June 2014



Big Data approaches offer the opportunity to extract information from large and complex data sources from diverse disciplines, and present compelling reasons to embrace new scientific research methods. The international scientific community has a responsibility to examine these opportunities to use Big Data for knowledge discovery that will benefit society and the sustainability of the planet.

Big Data have huge potential and are already benefitting scientific research and discovery. They present particularly significant challenges and notable opportunities for transdisciplinary, international research programmes. Such initiatives aim to produce research results and data in ways that improve decision-making on critical issues for humankind and the environment. To accomplish this requires the integration of increasingly diverse and complex datasets, the design of new forms of data collection, and the extraction and interpretation of knowledge utilizing sophisticated statistical techniques, complex simulation models, and other computationally intensive approaches. The challenges and opportunities of Big Data have significant implications for scientific data services and infrastructure providers.

In recognition of this, the co-sponsors and participants of the *Workshop on Big Data for International Scientific Programmes* join with CODATA in presenting the following statement of recommendations and actions as a call for international collaborations to address the challenges and take advantage of the opportunities of the Big Data age.

## Recommendations to the Sponsors of International Scientific Programmes

1. **Respond to the importance of Big Data for international scientific programmes:** Addressing the combined challenges posed by data volume, complexity and heterogeneity will bring significant benefits to international scientific programmes. Big Data offer notable opportunities for knowledge discovery, particularly in relation to interdisciplinary research questions: this importance should be recognized and acted upon.
2. **Exploit the benefits of Big Data for society:** Sponsors of international scientific programmes need to encourage activities that exploit Big Data to promote applied research for the benefit of society and to make provision to promote and support such activities. Exploiting Big Data will require coordination and collaboration: research funders, national academies, universities, data services and other research performing institutions should formulate coordinated strategies to encourage the development and application of Big Data to meet high priority science needs.
3. **Improve understanding of Big Data through international collaboration:** Research into the methodologies, theories, technologies and – above all – the practical applications of Big Data for international scientific programmes should be strengthened. This must involve a broad range of experts and research disciplines and sustained international collaboration will also be essential.

4. **Promote universal access to Big Data through global research infrastructures:** The knowledge generated by the exploitation of Big Data in international scientific programmes stands to benefit all humanity. Achieving this benefit depends, however, on the widest possible access to and participation in the creation of that knowledge. Where economically feasible, the coordinated development of universally accessible and global research infrastructures capable of handling Big Data should be supported by collaboration between all countries and international organizations. The sustainability of such research infrastructure needs to be addressed by demonstrating the contribution to science and return on investment where appropriate.
5. **Explore and Address the Challenges of Big Data Stewardship:** The principle that data sources should be available for verification and re-use holds true for Big Data, providing an important incentive for selecting, preserving, documenting and further disseminating curated data products. Making the most effective use of Big Data will require curation, quality assessment and the quantification of uncertainty for relevant data sources. However, the challenges and considerable costs involved in the stewardship and long-term preservation of Big Data need further examination and will have implications for the way in which this principle can in practice be realised. The ICSU World Data System (ICSU-WDS), and others, can provide expertise and a coordinating function which will be essential for addressing the specific challenges of Big Data stewardship.
6. **Encourage capacity building and skills development:** The commercial potential of Big Data and data analytics has been much publicized, as has the pressing need for skills development in Big Data science. This undoubtedly holds true in a research context also. We call on the partners to this initiative to collaborate with appropriate national and international organizations to advance an agenda for capacity building and skills development in Big Data science. This involves prioritization of Big Data science in educational regimes and developing the career paths of early career researchers.
7. **Foster development of policies to maximize exploitation of Big Data:** Big Data raise new and more complex management, access and reuse problems. Policies, guidelines, international agreements and protocols should be developed to maximize the collection, sharing and potential exploitation of Big Data for scientific research. Such activities need to be advanced on an international and multi-disciplinary basis. CODATA, together with the stakeholders involved, can play a significant role in support of such policy development to address the specific challenges of Big Data.

### Proposed Actions for a CODATA Working Group

- A. **Establish a CODATA Working Group on Big Data for Scientific Programmes:** CODATA should convene and coordinate a broad-based and expert, international Working Group to examine and promote Big Data issues and opportunities, such as those initially identified below, for international scientific programmes; co-sponsors to the Workshop and other partners are strongly invited to collaborate.
  - a. **Produce case studies in Big Data for international scientific programmes:** Greater clarity is needed about the precise nature of the challenges and opportunities of Big Data. Produced in partnership with relevant international research programmes, a series of case studies with reproducible examples, focusing on use-oriented data for solutions-oriented science, would be an effective means of improving understanding and sharing knowledge.
  - b. **Promote sharing of Big Data solutions across scientific disciplines:** The Workshop identified several possible approaches for technical, infrastructure, and analytic solutions, including activities that draw on and apply reference models, or transpose solutions developed in one research context and discipline to others. The Working Group should collaborate with other partners and with international scientific programmes to facilitate a mechanism for discussing, publicising, sharing and possibly adopting such solutions. A roadmap for exploiting Big Data in interdisciplinary and societally-relevant research could be a possible outcome of this work.

- c. **Research policy, ethical and legal issues for Big Data:** In pursuing a policy agenda for Big Data, specific issues must be addressed. Where Big Data draws on observation or sensor networks, social media and other means of data gathering involving or relating to human subjects, ethical and legal issues need to be addressed. Big Data are not just quantitatively different from smaller data sets; they have characteristics that are also qualitatively different and that raise new science and data policy issues that need to be identified and resolved. For example, licensing and access issues persist, but potentially on a different scale if the potential of vast data integration is to be achieved. The policy agenda should aim to remove inappropriate barriers, perceived or otherwise, to the use and reuse of Big Data sources.
- d. **Research stewardship and sustainability challenges for Big Data:** In collaboration with ICSU-WDS and other partners, the proposed Working Group should help coordinate a detailed consideration of the dual challenges of long-term preservation and the sustainability of data infrastructures in relation to the stewardship of Big Data products for international scientific programmes.

## Background

This Statement of Recommendations and Actions was made at the *Workshop on Big Data for International Scientific Programmes: Challenges and Opportunities*, convened by [CODATA](#), the ICSU Committee on Data for Science and Technology, with a number of international partners, in Beijing, China, on 8-9 June 2014.

The workshop was designed to provide a better understanding of the opportunities and challenges of ‘Big Data’ for international collaborative scientific programmes, including

- ICSU-sponsored programmes such as [Future Earth](#) and [Integrated Research on Disaster Risk \(IRDR\)](#),
- International initiatives such as the [Group on Earth Observations \(GEO\)](#), the [Intergovernmental Platform on Biodiversity and Ecosystem Services \(IPBES\)](#), the [International Society for Digital Earth \(ISDE\)](#), and
- Initiatives sponsored by the [Belmont Forum](#).

The workshop was convened by CODATA, the ICSU Committee on Data for Science and Technology, and co-sponsored by:

- The [ICSU World Data System \(ICSU-WDS\)](#),
- Future Earth,
- Integrated Research on Disaster Risk,
- The Group on Earth Observations,
- The [Research Data Alliance \(RDA\)](#),
- The [International Society for Digital Earth \(ISDE\)](#), and
- The [Institute of Remote Sensing and Digital Earth \(RADI\) of the Chinese Academy of Sciences \(CAS\)](#).

The objective of the workshop was to help set an agenda for collaborative and coordinated activities on to help exploit Big Data for international science, in which CODATA should play a leading role.