# International Data Week 2016:

# From Big Data to Open Data: Mobilising the Data Revolution

Talk of a data revolution is not hyperbole. The IT revolution has now given us the means of collecting huge amounts of data and, with the advent of the Cloud, unprecedented computing power to analyse this data and find new and important discoveries. The ease in which data relating to human behaviour and transactions can be gathered has led to new industries driven by their ability to elicit predictive and commercially advantageous information from masses of data. In academia, new data science courses and multidisciplinary centres are springing up to feed the demand for these analytical and data management skills. The data revolution, the phenomenon of Big Data and advances in data science are everywhere impacting both scientific research and industry.

Many research areas have been transformed in the last twenty years: increasingly scientific breakthroughs depend on an international, collective ability to create, share and analyse vast quantities of data. Likewise, understanding and responding to environmental and societal challenges is now increasingly dependent on our ability to gather data from diverse sources, to identify appropriate indicators, to interpret them correctly and to design intelligent responses. It is no exaggeration to say that a resilient, sustainable and successful society will be one that – by analogy to a successful organism – has the capacity to metabolise data into knowledge.

The data revolution has many dimensions. What has been characterized as ‘Big Data’ or ‘data intensive research’ is an important part of the phenomenon, as are associated means of gathering data through a plethora of devices at unprecedented scale. Also important however are the opportunities for comparatively small data, including that of making bespoke, hard-won data sets more accessible and reusable, enhancing their ‘generative value’, through providing contextual information, linking or integrating with other datasets and through enhanced visualizations and analyses.

Exploiting the opportunities and addressing the challenges afforded by the data revolution and using them to generate wider societal benefit will fundamentally depend on the creation of a complementary ‘Open Data’ environment. Open data is crucial to the maintenance of scientific ‘self-correction’ whereby the data underlying published concepts are open to scrutiny, replication or invalidation. The rapid growth of data makes this crucial principle of research ever more difficult to sustain, and increasingly requires both the data and the code used in data analysis to be open, accessible and useable.

***International Data Week*** brings together data scientists, researchers, industry leaders, entrepreneurs, policy makers and data stewards to explore how best to exploit the data revolution to improve our knowledge and benefit society through data-driven innovation.

## International Data Week

The agenda for the week comprises three complementary events:

1. **SciDataCon 2016:**

A scientific research conference dedicated to addressing fundamental issues and advancing the frontiers of data science and data stewardship, convened by ICSU CODATA and ICSU World Data System.

1. **An International Data Forum:**

This will be an agenda-setting event that will define and advance the state-of-the-art in collaboration between research, industry and education.

1. **The 8th Research Data Alliance Plenary Meeting:**

An assembly of the research data community creating and presenting implementable solutions to enable data sharing.

## SciDataCon 2016: Advancing the Frontiers of Data Science and Data Stewardship

SciDataCon 2016 seeks to advance the frontiers of data science and data stewardship by exploring the fundamental issues concerning recent data-driven transformation of research. In addition, the conference explicitly seeks to relate the issues of data-intensive science with the profound challenges and opportunities associated with environmental and societal change and with data-driven innovation.

These are multi-faceted issues, which cannot be tackled without expertise drawn from many disciplines and practitioners with different roles in the research enterprise. Furthermore, these multidisciplinary research challenges are profoundly international and the response must be genuinely global. Nor can research achieve its objectives without the infrastructure and skills required for state of the art data analysis and the long-term stewardship of valuable data products.

SciDataCon aspires to be \*the\* international scientific conference addressing frontier issues of data science across the research space.

## International Data Forum

The International Data Forum brings together international researchers, industrialists, policy makers and educators to discuss the major opportunities and challenges of the data revolution, such as ‘Big Data’ and ‘Open Data’ with an aim to provide a more profound understanding of the Data Revolution. This is the centrepiece of International Data Week and will feature keynotes from world-renowned speakers, high-level panels and discussion sessions. The target audience is comprehensive, ranging from academic researchers, to government decision makers, and industrial leaders. An exhibition space will be organized to showcase ongoing capabilities and advances, with a focus on visionary prospects in data science.

The Forum will address the national and **international** dimensions of the data revolution and the implications of the data revolution for major research questions and national and international policy. It will explore the linkages between research, industry, government and education that are required for the global community to fully exploit the opportunities offered by the data revolution. Innovations in industry may inform academic and publicly funded research, and vice-versa. Activism around open public data and open scientific data bring the exciting prospect of greater public engagement and citizen science of real value.

Multiple dimensions of the technical and contextual environment in which data are generated will be examined: data ownership, data control, data security, data integrity, data protection and privacy, ‘dataveillance’, data uses (ethics and politics of how data are deployed), data quality and provenance, data integration and interoperability, and data analytics. Further, the administrative, legal, and commercial challenges raised by the concept of ‘open access’ to data and emerging national, institutional, journal and ‘research-discipline’ data policies will be explored.

Data policy and legal interoperability across national boundaries are challenges to be addressed by public officials and a fully engaged community. Both industry and academia are concerned with the development of ‘data scientists’ but definitions of the skills required vary. In order to sustain this rapid evolution, the education and career profile of data scientists must be adjusted to the changing environment. This is of direct concern to the academic community, which must be responsive to societal and economic needs raised by the data revolution.

## 8th Research Data Alliance Plenary Meeting: Implementable solutions to promote data sharing

The 8th RDA Plenary Meeting will provide the RDA member community a unique opportunity to network and collaborate with colleagues and peers in various disciplines, and make concrete progress in technical and social areas on topics related to research data sharing and exchange.

The particular focus of this plenary meeting will be on implementable solutions to promote data sharing and data-driven innovation.

Besides keynote talks by leading members of the community, an important feature of the RDA plenary programme are meetings of the Working Groups and Interest Groups, and joint issue sessions. There are also ‘Birds of a Feather’ discussions exploring new potential working or interest group topics.

## Schedule for International Data Week 2016

Sunday, 11 Sept: CODATA General Assembly and CODATA 50 Years Celebration and Reflection

 WDS Members Forum 2016

 RDA Working and Interest Group Chairs Meeting?

Monday, 12 Sept: SciDataCon 2016: Advancing the Frontiers of Data Science and Data Stewardship

Tuesday, 13 Sept: SciDataCon 2016: Advancing the Frontiers of Data Science and Data Stewardship

Wednesday, 14 Sept: International Data Forum/Summit

Thursday, 15 Sept: 8th RDA Plenary Meeting: Implementable solutions to promote data sharing

Friday, 16 Sept: 8th RDA Plenary Meeting: Implementable solutions to promote data sharing

## SciDataCon 2016: Advancing the Frontiers of Data Science and Data Stewardship

SciDataCon 2016 seeks to advance the frontiers of data science and data stewardship by exploring the fundamental issues concerning ongoing data-intensive transformation of research. In addition, the conference explicitly seeks to relate the issues of data-intensive science with the profound challenges and opportunities associated with environmental and societal change and with data-driven innovation.

**The Frontiers of Data Science and Data Stewardship**

The unprecedented explosion in the capacity to acquire, store and manipulate data and information and to communicate them globally, is a world historical event involving a revolution in knowledge creation far more profound and pervasive than that associated with Gutenberg’s invention of the printing press. This profound transformation, which has led to more data intensive research and to theorizing around ‘Big Data’, poses challenges to the fundamental processes of management and open scrutiny of scientific evidence—the data—on which new discoveries are based. The data revolution offers new opportunities to identify patterns and processes in phenomena that have hitherto been beyond our capacity to resolve. It challenges us to develop new modes of collaboration and coordinated action that are needed to sustain observational and monitoring capacities and maximize scientific and societal benefit. These developments in data-intensive science also challenge deep-seated scientific norms and many of the habits of researchers and their institutions.

These developments have created major new opportunities for science:

* to identify patterns and processes in phenomena that have hitherto been beyond our capacity to resolve;
* to integrate data reflecting a wide variety of coupled processes to obtain much deeper understanding of relationships than has hitherto been possible;
* to improve forecasts of system behaviour by integrating data acquisition and modeling;
* to make cognate datasets and data-integration tools readily available and useable by individual researchers from the rapidly growing number of open databases, while ensuring the quality and reliability of the available data;
* to permit re-use, re-combination and re-purposing of data in ways that make data of perennial, cumulative value, rather than being lost from generation to generation;
* to exploit the opportunities created by communicating, automated sensors - the ‘internet of things’ - in exploring complex phenomena and unraveling complexity.

These frontier issues of ‘data science’ are the core subject of the conference.

**SciDataCon and International Research Challenges**

SciDataCon is motivated by the conviction that the most significant research challenges—and in particular, the pressing and transdisciplinary issues relating to global sustainability in the face of ongoing natural and human-induced changes to the planetary system—cannot be properly addressed without paying attention to issues relating to data. These issues include policy frameworks, data quality and interoperability, long-term stewardship of data, and the research skills, technologies, and infrastructures required by increasingly data-intensive research. These are multi-faceted issues, which cannot be tackled without expertise drawn from many disciplines and roles in the research enterprise. Furthermore, these issues are essentially international and the response must be genuinely global.

SciDataCon aspires to be \*the\* international scientific conference addressing frontier issues of data science and data stewardship across the research space. The conference will also:

* advance discussions of the international policy agenda for Open Data and Open Science; and,
* provide a forum that brings together the researchers who create and use data with the other professional communities involved in data science, those concerned with long-term data stewardship and with data and science policy.

Consequently, three types of paper or sessions are identified:

1. **Research paper sessions**

Quality, peer-reviewed research papers on the thematic issues identified above. Research papers must describe the outcomes and application of unpublished original research. These should make a substantial contribution to knowledge and understanding in the subject matter. Research papers may be submitted individually or as part of a thematic session.

1. **Practice papers and reports**

Practice papers may include reports from significant projects, institutions or groups working on important data issues. Such reports might not include original research, as such, but give a significant and interesting insight into issues of importance to data practitioners of various types (researchers, data scientists, data curators, data managers, database designers etc).

1. **Agenda-setting policy and strategy discussions.**

Participants may submit proposals for sessions which aim to stimulate discussion and debate on topics and controversies of interest to the community. Such sessions may feature short presentations, perhaps as statements of a position, but they should seek to cultivate discussion around important issues.

1. **Poster presentations and demonstrations.**

Poster presentations will be an important feature of SciDataCon and of International Data Week. There will be a poster session and demo session and prizes will be awarded. Participants are encouraged to think of the best means of communicating their research or report. Where there is significant technical detail or the visual medium is important, a poster may be the most effective medium.