

9-1-2012

ICSU and the challenges of big data in science

Ray Harris Prof
University College London, UK

Follow this and additional works at: <https://www.researchtrends.com/researchtrends>

Recommended Citation

Harris, Ray Prof (2012) "ICSU and the challenges of big data in science," *Research Trends*: Vol. 1 : Iss. 30 , Article 5.

Available at: <https://www.researchtrends.com/researchtrends/vol1/iss30/5>

This Article is brought to you for free and open access by Research Trends. It has been accepted for inclusion in Research Trends by an authorized editor of Research Trends. For more information, please contact r.herbert@elsevier.com.

Section 4:

ICSU and the Challenges of Big Data in Science



Professor Ray Harris
Department of Geography
University College London

The Fourth Paradigm

The enormous amounts of data now available to science and to society at large have stimulated some authors to say that we are in the Fourth Paradigm of data-intensive science (1). The First Paradigm was the period of observation, description and experimentation characterised by early scientists and explorers such as Ptolemy and Ibn Battuta. The Second Paradigm was that of the development of theory to explain the way the world works such as in Maxwell's equations and Newton's theory of gravitation and laws of motion. The Third Paradigm developed the earlier theories to create extensive simulations and models such as those used in weather forecasting and in climatology. The reason for the step change to a new paradigm, the Fourth Paradigm, is that the volume of data available to us is so large, now often termed Big Data, that it is both presenting many new opportunities for analysis as well as requiring new modes of thinking, for example in the International Virtual Observatory Alliance (<http://virtualobservatory.org>) and in citizen science.

Big Data

One clear example of Big Data is the Square Kilometre Array (SKA) (www.skatelescope.org) planned to be constructed in South Africa and Australia. When the SKA is completed in 2024 it will produce in excess of one exabyte of raw data per day (1 exabyte = 10^{18} bytes), which is more than the entire daily internet traffic at present. The SKA is a 1.5 billion Euro project that will have more than 3000 receiving dishes to produce a combined information collecting area of one square kilometre, and will use enough optical fibre to wrap twice around the Earth. Another example of Big Data is the Large Hadron Collider, at the European Organisation for Nuclear Research (CERN), which has 150 million sensors and is creating 22 petabytes of data in 2012 (1 Petabyte = 10^{15} bytes, see Figure 1). In biomedicine the Human Genome Project is determining the sequences of the three billion chemical base pairs that make up human DNA. In Earth observation there are over 200 satellites in orbit continuously collecting data about the atmosphere and the land, ocean and ice surfaces of planet Earth with pixel sizes ranging from 50 cm to many tens of kilometres.

In a paper in the journal *Science* in 2011, Hilbert and Lopez (2) estimated that if all the data used in the world today were written to CD-ROMs and the CD-ROMs piled up in a single stack, the stack would stretch all the way from the Earth to the Moon and a quarter of the way back again. A report by the International Data Corporation (3) in 2010 estimated that by the year 2020 there will be 35 Zettabytes (ZB) of digital data created per annum.

International Council for Science

The International Council for Science (ICSU) is the coordinating organisation for science and is taking a leading role in developing further the capability of science to exploit the new era of the Fourth Paradigm. The members of ICSU are the 121 national scientific bodies such as the Australian Academy of Sciences and the Royal Society plus the 31 international science unions such as the International Astronomical Union and the International Union for Crystallography. ICSU has always been committed to the principle of the universality of science and in its vision (4) it sees:

"... a world where science is used for the benefit of all, excellence in science is valued and scientific knowledge is effectively linked to policy making. In such a world, universal and equitable access to high quality scientific data and information is a reality ..."

Because of its desire to make a reality of universal and equitable access to data, ICSU established three initiatives to address how ICSU can encourage better management of science data (5).

- Panel Area Assessment on Scientific Information and Data, 2003–2004.
- Strategic Committee on Information and Data, 2007–2008.
- Strategic Coordinating Committee on Information and Data, 2009–2011.

World Data System

One of the main outcomes of these ICSU initiatives is the establishment of the World Data System (WDS). In 1957 during the International Geophysical Year (IGY) several World Data Centres were initiated to act as repositories for data collected during the IGY. The number of these data centres increased over time but they were never fully coordinated. The World Data System is now in the process of rejuvenating these data centres by establishing an active network of centres that practice professional data management. The objectives of the WDS are as follows:

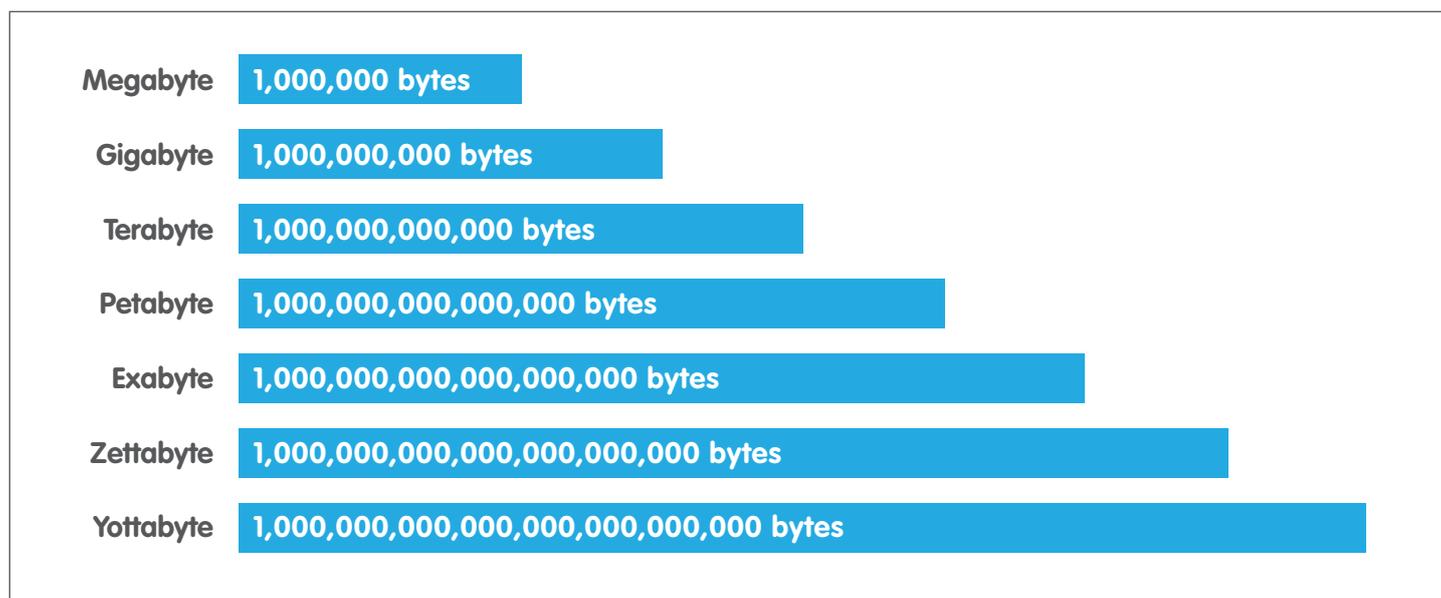


Figure 1: Overview of data scale from megabytes to yottabytes (log scale).

- Enable universal and equitable 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

By early 2012 over 150 expressions of interest in the WDS had been received by ICSU, resulting in over 60 formal applications for membership. Approved members of the World Data System so far include centres for Antarctic data (Hobart), climate data (Hamburg), ocean data (Washington DC), environment data (Beijing) and solid Earth physics data (Moscow) plus the International Laser Ranging Service and the International Very Long Baseline Interferometry Service. By 2013 it is anticipated that the WDS will comprise over 100 centres and networks of active, professional data management.

Further actions

There is still much to do in developing a professional approach to data management in science. The main outstanding issues were addressed by the ICSU Strategic Coordinating Committee on Information and Data noted above and include the following: better guidance for best practice on data management; improved definitions of the various terms used in the phrase "open access"; greater recognition of the publication of data by scientists as well as the publication of journal articles and books; practical help in data management for less economically developed countries through partnership with members of the ICSU family and others; and cooperation with commercial companies for mutual benefit.

Conclusion

Big Data presents science with many challenges, but at the same time presents many opportunities to influence how science grows and develops for the better, not least by adding data-driven science to hypothesis-driven science. Improvements in professional data management will result in better science.

References:

1. Hey, T., Tansley, S. & Tolle, K. (2009) The Fourth Paradigm. "Data-intensive scientific discovery", Microsoft.
2. Hilbert, M. & Lopez, P. (2011) "The world's technological capacity to store, communicate and compute information", Science 332, 1 April 2011, 60-65.
3. IDC (2010) "IDC Digital Universe Study, sponsored by EMC", May 2010, available at <http://www.emc.com/collateral/demos/microsites/idc-digital-universe/iview.htm>
4. ICSU Strategic Plan 2006-2011, International Council for Science, Paris, 64pp
5. All the reports are available at the ICSU website, www.icsu.org