



Research Data in the Digital Age

Dr Mike Wallace, mikew@elsenburg.com

The last few decades have seen an exponential rise in the volume of data collected and stored across many research domains, yet the current global research data landscape is highly fragmented across (and indeed even within) disciplines from life sciences and health, to agriculture, space and climate. So although we now (in theory) have access to unprecedented volumes of data - often, for various reasons, the datasets are not interoperable and not readily usable in a wider context. In order to facilitate cross-disciplinary research, common data infrastructures, protocols and the building of specific "data bridges" have become necessary to reduce data complexity and enable data sharing to help address many of the grand challenges of the modern age. Concepts such as "big data", the data "cloud", "crop ontologies", "data cubes" and professions such as "computational biologist" and "data scientist" are perhaps not familiar to many of us – but are examples of new vocabulary and careers that are developing within the data and data analysis milieu.

An international organisation known as the Research Data Alliance (RDA) was established in 2013 to facilitate networking and dialogue with the objective of removing barriers that impede data interoperability, data access and data sharing. This is accomplished through focused Working Groups and Interest Groups, formed of experts from many countries around the world – from academia, industry and government. The author (Programme Research and Technology Development Services) and Mr Shaheed Martin (Programme Farmer Support & Development) travelled to the 6th Plenary Session of the RDA held in Paris in September 2015 to participate in the agricultural interest group (IGAD) workgroups and RDA seminar to gain insight into international trends and best practices in the realm of research data management. The outcomes of the small group IGAD workshops were presented by the groups and discussed in the plenary sessions, along with a number of presentations from experts across various disciplines.

Although much of the discussion was obviously of a technical nature, some of the overarching themes and messages of general relevance were as follows:

- In order to move forward and address many of the issues facing us, research data must wherever possible be made “intelligently” open, interoperable and free. Institutional data “brokering” services will be promoted to facilitate this and ensure that data conforms to standards.
- Research data plans have become a prerequisite of many organisations which fund research projects and likewise within universities as a requirement in post graduate study proposals.
- Given the pending Climate Change Conference of the Parties (COP21) in Paris later this year, there was considerable focus on the need for open and free climate data and the need for standards to ensure the interoperability on comparability of both historic and modelled future climate scenarios.
- As 2015 is the International Year of the Soil, the often neglected, but vitally important soil resource data was also a focus, with speakers from a number of countries presenting on the establishment of data portals to enable free access “one-stop shops” to the public via the Web.
- Metadata are critical components in building the “data bridges” referred to in the first paragraph. Metadata are, simply put, small files containing information about a dataset, such as field properties, methods used, scale, units and possibly many other parameters. Metadata files are critical in facilitating interoperability of data. Considerable effort has gone into establishing various international standards for these.
- Closely aligned to the objectives of the RDA, the Committee on Data for Science and Technology (CODATA) is an interdisciplinary committee of the International Council for Science. It seeks to improve the compilation, critical evaluation, storage, and retrieval of data of importance to science and technology. They provide training specifically to address many of the issues discussed above. More specifically, in the fields of agriculture, food and nutrition, a body known as GODAN (Global Open Data for Agriculture and Nutrition) has been established, focusing on developing policy-level public and private sector institutional support for open data (see reference below).

Some agencies stipulate that state funded researchers make their project data open and accessible. One of the fundamental issues relating to such an implementation of “open data” in research is related to the intellectual capital of the researchers themselves, tied up in their

datasets. By making their data open, they risk being “scooped” in terms of publications. Whilst this will always be an ongoing debate, datasets are becoming citeable entities through the use of specific data object identifiers (DOIs). This means that researchers will be able to achieve recognition not just for publications, but also for datasets that they create and curate.

It was very evident from the workshops, presentations and informal discussions that nobody has all the solutions. It is accepted however, that to address the many challenges, we need to promote ongoing dialogue, awareness and trust through formal and social networking, seminars, training and probably above all, a change in the culture of research data creation and management. The latter is of particular relevance in South Africa where some of our parastatals retain historical business models which consider data (gathered at taxpayers' expense) as a revenue source rather than a public good.

Further reading/resources

<https://rd-alliance.org/about.html>

<http://theodi.org/how-improve-agriculture-food-nutrition-open-data>

<http://theodi.org/what-is-open-data> (puts open data in context – compared to “closed data”, that are only accessible to specific people for reasons of privacy, commercial sensitivity and security)

<http://www.codata.org/> (International Council for Science: Committee on Data for Science and Technology)

<http://www.godan.info> (Global Open Data for Agriculture and Nutrition)

For more information, contact Dr. Wallace at 021 808 5088.