A South African Data Service: A Presentation to the ICSU World Data System Members' Forum

Lynn Woolfrey

DataFirst

lynn.woolfrey@datafirst.uct.ac.za

In 2000, member states of the United Nations issued the Millennium Declaration, which pinpointed eight Millennium Development Goals (MDGs) to address global and national inequalities to improve the lives of their citizens by 2015. The original MDG report made no mention of the data that would be needed to implement or measure MDG programmes. In contrast, the post-2015 MDG agenda report released in 2013, highlights the need for measurable indicators of growth. The agenda states that reliable and comparable country data is essential to measure achievement of the MDGs and is also a means of attaining these goals. That development planning should be based on sound data seems obvious. All governments collect statistics on their citizens to inform policymaking. The new development agenda though, is presaged on more microdata— "raw" data at unit record level. Microdata are required for in-depth policy analysis because highly aggregated data can hide local and national inequalities .The 2013 MDG report calls for "a data revolution" that will lead to more and better quality data, more detailed data and, above all, more Open Data.

Making this raw data "Open"—that is, widely and easily available— is vital because often government functionaries do not possess the time or quantitative skills to undertake in-depth data analysis. However, governments can use a data-rich Research Policy Interface to take advantage of the higher quantitative literacy among academics to support appropriate policy feedback. In recognition of this, many countries have established Research Infrastructures to support the collection and reuse of detailed data. This includes Cyberinfrastructure such as high-performance computing and high-speed backbone networks, and institutional infrastructure such as Research Data Centres and Data Archives. Data Archives were originally places to preserve data and make it available to researchers on request, but many now also provide Data Services such as value-added input (e.g., data harmonisation), as well as expert advice and training in data handling and data analysis. In South Africa the South African Data Archive in Pretoria fulfils the former roles, while DataFirst has evolved into a fully-fledged Data Service.

DataFirst was established in 2001 to help researchers access and use microdata collected in South Africa and other African countries. The unit is based at the University of Cape Town in South Africa, but our online data portal enables discovery and access by researchers worldwide. We assist data users via an online support site and we also run data analysis training workshops. This includes skilling-up municipal officials to support fact-based local government planning. The Service trains data managers in South Africa and other African countries. The impact of the service has been an improvement in the quality of South African data through feedback to producers. Training users has also expanded the cadre of African researchers with the skills to use quantitative data. Training African data curators has led to improvements in access to other African data.

DataFirst follows data curation best practice and models its service on the OAIS standard. DataFirst became a Member of the ICSU World Data System in 2013, and was awarded the Data Seal of Approval in October 2014. The latter is an internationally recognised certification for trusted digital repositories, and only six other Data Archives that are members of the International Federation of Data Organizations have garnered this award, at present. New frontiers for DataFirst include the creation of a Secure Data Service to share sensitive, previously unavailable South African data and the growing of our administrative data collection. DataFirst is currently drawing up specifications for a National Data Service at the request of the South African Department of Science and Technology. This could release long-term funding for the work of DataFirst and other data institutions in the country.