SYNOPSIS

Open Environmental Data in Developing Countries: Who Benefits?

Eduardo Eiji Maeda, Juan Arévalo Torres

Received: 7 March 2012 / Accepted: 13 March 2012 / Published online: 24 April 2012

This synopsis was not peer reviewed.

Should research institutions from developing countries provide open access to their data? Who would benefit? These are highly relevant questions for environmental research. In the backdrop of increasing discussions regarding open data, attention should be paid to guarantee that research institutions from developing countries do not become mere data providers. At the same time, restricted access to data should not become a barrier for advancing our knowledge of critical environmental issues. Open access policies for environmental data must be developed not only for advancing science, but also to provide growing opportunities for researchers in developing countries.

Scientific research plays a crucial role in addressing global environmental problems, such as climate change, biodiversity loss, and water scarcity. Advances in these fields strongly rely on the availability of data allowing researchers to better understand environmental processes. From temperature records obtained using simple thermometers, to complex measurements from satellite sensors, data gathering is a key component of any research project. Nonetheless, data gathering is often complex, expensive, and time consuming.

The importance of environmental data is largely recognized. In fact, it is rather common for research groups or institutions to protect their data, in an attempt to guarantee control of potential scientific findings associated with these datasets (Nelson 2009). However, data protection may lead to situations that are not always consistent with the advance of scientific knowledge or with the coherence of the scientific method itself. One of the most basic characteristics of the scientific method is reproducibility. If datasets used

to achieve scientific discoveries are not accessible by other researchers, then the transparency of the methods and results may unnecessarily be compromised.

Furthermore, in many instances, restricted data access may lead to heterogeneity in terms of metadata, data formats, and storage methods. We may take as an example basic meteorological data in developing regions like Latin America and Africa. Although international projects, such as the Global Historical Climatology Network (GHCN), play an important role in distributing meteorological data from these regions (and also globally), a vast amount of data that are internally produced in some countries hardly reach the international community. Consequently, meteorological records are not distributed properly, and metadata are rather heterogeneous. This has large implications for climate research, increasing uncertainties in identifying current climate trends, understanding global atmospheric dynamics, and simulating future climate scenarios.

In this context, discussions regarding open data strategies have become more intense in recent years (Nelson 2009; Carlson 2011; Overpeck et al. 2011). Herewith, we define open data simply as data that are freely accessible, re-useable, and redistributable by anyone, without restrictions of any kind. Benefits of open data strategies have been broadly discussed (e.g. Reichman et al. 2011). A common argument is that data obtained using public money should be used in a way to maximize the benefits for the general public. This implies that access to data should be unrestricted, enlarging the range of applications and uses for the data. In this regard, the Organisation for Economic Co-Operation and Development has developed standards to facilitate access to research data generated from public funding (OECD 2007). Studies have shown that open access to publicly funded data also generates

wealth through downstream commercialization of outputs and provides decision makers with facts needed to address complex, often transnational, problems (Arzberger et al. 2004). These benefits are gradually being recognized. In some research institutes and funding agencies, open data policy is already considered a standard practice (NOAA 2009; World Bank 2011).

Despite this growing recognition, one remaining question is whether research institutions from developing countries would benefit from the advantages of open data policies. The answer to this question is not straightforward and may not be the same for all the developing countries. With few exceptions, research institutions in Latin America and Africa are likely to face more obstacles in finding financial resources when compared with top institutions from Europe and the USA. An international open data environment could therefore lead to a broadening of the gaps between research institutions in rich and poor countries. Institutions from developing countries could become mere data providers, while wealthy institutions would apply state-of-the-art equipment and well-paid researchers to generate the new scientific breakthroughs.

A frequent strategy for overcoming this problem is the creation of cooperation networks between institutions from developing countries, in an attempt to pool efforts to compete in the international scientific scenario. Nonetheless, this alternative often leads to a data-distribution strategy slightly different from the open data definition stated in this article, given that data are mostly shared among institutions belonging to the networks. This “partially open data” approach can certainly benefit some institutions, but it is not necessarily beneficial for science as a whole. However, even if partially open data initiatives slow down science in the short term, is the strengthening of developing countries institutions not a greater advance for science in the long term? In theory, yes, but even within a context of cooperation networks, data competition and restrictions are likely to lead to even higher barriers to the globalization of data access. It could be one step forward followed by two steps backward.

Hence, data distribution policies must conciliate the advance of scientific knowledge with growing opportunities and conditions for researchers in developing countries. Scientific cooperation should be seen not as a solution for data sharing, but as the first step toward the strengthening of scientific foundations. Open and unrestricted access to data should not only categorically be foreseen by funding agencies and research institutions, but at the same time also ensure that conditions and competitiveness are guaranteed for the researchers from developing regions. This is a complex task, but not an impossible one. Several initiatives from leading institutions in developing countries have proven that when conciliated with a strong scientific

framework, open data policies may highly benefit local science.

A good example is the large scale biosphere–atmosphere Experiment in Amazonia. This international initiative, led by Brazil, is one of the largest scientific cooperation efforts for environmental research in the world. By producing public domain data and creating a solid structure for data sharing, the initiative has produced an output of over 2000 scientific articles in two decades (Artaxo 2012). Furthermore, it contributed to changing the belief that research activities in Amazon are led mostly by scientists from the developed countries (Malhado 2011).

Another case from Brazil is the pioneering satellite data-distribution policy carried out by Brazil’s National Institute for Space Research (INPE). By conciliating free access to satellite imagery and the development of open source software for data handling, this initiative has greatly reduced the barriers for earth observation science in Brazil and Latin America. A large increase in users of remote sensing data has been observed in the region, including highly qualified scientists capable of carrying out activities at an international level. As a result, Brazil is currently considered a reference point in tropical forest monitoring using remote sensing data (Tollefson 2008).

Recent initiatives coming from developed countries have also started to recognize that high quality scientists and institutions in developing countries can take the lead in scientific collaborations. For instance, the RALCEA¹ (Latin American Network of Knowledge Centres in the Water Sector) is a project funded by the European Commission aiming at improving scientific cooperation and fostering information-based policy in Latin America. The project acknowledges the existence of water research centers of excellence in Latin America and direct efforts for strengthening the links among these institutions. This approach allows for the solidification of scientific foundations in developing regions and cultivates a closer relationship among European and Latin American institutions. This closer relationship contributes to establishing a basis for better data sharing leading to open environmental data policies.

Finally, it is evident that the implementation of open data policies will often require efforts for a systematic examination of barriers and best practices, to document the current situation and offer guidelines for further actions. To ensure a fair international scientific environment, much work is needed to better reward data sharing initiatives that contribute to crossing the boundaries of science. Challenges for global open access to environmental data surely exist, but the benefits are likely worth the hard work required.

¹ RALCEA Project: http://ec.europa.eu/europeaid/where/latin-america/regional-cooperation/ralcea/index_en.htm.

REFERENCES

- Artaxo, P. 2012. Break down boundaries in climate research. *Nature* 481: 239.
- Arzberger, P., P. Schroeder, A. Beaulieu, G. Bowker, K. Casey, L. Laaksonen, D. Moorman, P. Uhlir, et al. 2004. An international framework to promote access to data. *Science* 303: 1777–1778.
- Carlson, R. 2011. A lesson in sharing. *Nature* 469: 293.
- Malhado, A.C.M. 2011. Amazon science needs Brazilian leadership. *Science* 6019: 857.
- Nelson, B. 2009. Empty archives. *Nature* 461: 160–163.
- NOAA. 2009. NOAA/National Climatic Data Center Open Access to Physical Climate Data Policy. <http://www.ncdc.noaa.gov/oa/about/open-access-climate-data-policy.pdf>. Accessed 6 March 2012.
- OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data; available at www.oecd.org. Accessed 6 March 2012.
- Overpeck, J.T., G.A. Meehl, S. Bony, and D.R. Easterling. 2011. Climate data challenges in the 21st century. *Science* 331: 700.
- Reichman, O.J., M.B. Jones, and M.P. Schildhauer. 2011. Challenges and opportunities of open data in ecology. *Science* 331: 703–705.
- Tollefson, J. 2008. Brazil goes to war against logging. *Nature* 452: 134–135.
- World Bank—Independent evaluation group. 2011. Access to information policy. <http://ieg.worldbankgroup.org/content/dam/ieg/A2I.pdf>. Accessed 6 March 2012.

Eduardo Eiji Maeda (✉)

Address: Water Resources Unit, Joint Research Centre—European Commission—Institute for Environment and Sustainability, Ispra, Italy.
e-mail: eduardo.maeda@jrc.ec.europa.eu

Juan Arévalo Torres

Address: Water Resources Unit, Joint Research Centre—European Commission—Institute for Environment and Sustainability, Ispra, Italy.