A more developmental approach to science

Bhanu Neupane, *Programme Specialist, Communication Sector, UNESCO*

Science 2.0: the data revolution

Science is not only created using data; the principle output of any scientific investigation is also data. The science-led data revolution has allowed Web 2.0 and Science 2.0 to co-evolve. The second-generation World Wide Web (Web 2.0) has made it easier for people to share information and collaborate and, in turn, the second-generation open science movement (Science 2.0) has used these new web-based technologies to share research more rapidly with a greater range of collaborators. This growth in interconnectedness, information-sharing and data-reuse has helped to develop a modern approach to science. As Science 2.0 is maturing, it has gradually begun replacing existing methods of teaching and learning science. Primarily characterized by the exponential generation and utilization of data for scientific purposes, this paradigm shift has both assisted and benefited from this data revolution (IEAG, 2014).

Increasingly collaborative science

Researchers and academics are now sharing their data and research results across web-based platforms, so that the global scientific community can utilize them and further build upon these raw scientific datasets, through collaboration. One example of this type of collaborative science can be seen in the big data generated for climate change projections developed by using global-scale models (Cooney, 2012). Research such as this provides a case for the utilization of large datasets assimilated and compiled in different parts of the world to solve local problems. This type of big data 'downscaling' can bridge the gap between global and local effects by layering larger-scale data with local-level data. Another example is the recently digitized and openly accessible rice breeding project 3K RGP, 2014 which now provides virtual access to the genomic sequence data of 3 000 rice cultivars from 89 countries. Local researchers can use such information to breed improved rice varieties that are locally customized for distribution at farmer level, resulting in higher annual rice yields that nurture national economic growth.

The combined impact of online tools and advocacy for a culture of open science at the institutional and national levels has fueled the accumulation and sharing of big data in virtual knowledge banks. Such sharing of metadata will, for example, allow for the generation of locally relevant projections of weather patterns and the development of cultivars that can best adapt to a particular climatic condition. In this way, studies in various scientific disciplines have become increasingly interconnected and data-heavy. This has made science more dynamic and given rise to two dimensions of scientific practices.

A shift from basic research towards big science

The focus of scientific discovery has shifted from basic research to 'relevant' or big science, in order to solve pressing developmental challenges, many of which have been identified as Sustainable Development Goals by the United Nations. However, basic research is extremely important for any future scientific discovery; one classic example is the discovery of the double helical structure of DNA by Watson and Crick in 1953, which laid the foundations for the subsequent work done in the fields of genetics and genomics. A more recent example is the sequencing of the human genome, which was completed in 2003 within the Human Genome Project. Whereas the identification of the 25 000 genes in human DNA was purely a quest for knowledge, the sequencing of corresponding base pairs within the same project was undertaken to unravel the mysteries of genetic variation, in order to improve the treatment of genetic diseases.

Computer networks and online interactions which facilitate the sharing of scientific information in real time across the global research community have gradually encouraged researchers to access and build upon these results in locally customized ways to solve social challenges. The global research community is no longer pegged on searching for a new element to add to the periodic table or for a molecular base triplet that encodes an amino acid. Rather, its focus is now on the bigger picture and how research can be applied to address challenges that could ultimately threaten human existence, such as global pandemics, water, food and energy insecurity or climate change. This shift in research priorities towards a big science agenda is evident in the amount of research funds allocated to applied science. Researchers are investing more than before in turning a discovery in basic research into a commercially viable and sustainable product or technology with a potentially beneficial socio-economic impact.

Without citizen engagement, no social good can come of open data

Another shift in the focus of science from basic research to an applied and developmental approach fuelled by Science 2.0 technologies is underscored by scientists' easier access than before to big data. Access can be defined firstly in the context of inclusiveness. If basic research is to be used for the betterment of human lives, there is no better way to identify a citizen's needs and challenges and to serve the interests of that person's wider community than to involve citizens themselves in the associated developmental processes. Science can only be inclusive if all parties at all levels (government, academic and general public) are duly involved. Thus, access can be defined secondly in the context of openness. Citizens cannot participate if science is not open and transparent. Without citizen engagement, no social good can come of open data, since there will be no recognition of local needs for subsequent data downscaling and data mainstreaming. For example, a regional scientific project aiming to identify the local impact of an increase in pollution levels can only be successful if citizens are able to report on the state of their health in real time to the scientific surveyors through a virtual platform that makes them active, yet informal participants in the project. Increasingly, discoveries that support early disaster warning – such as three-dimensional simulation models – are being considered more important that those that improve the capability to handle the post-disaster recovery.

Today's interconnected and futuristic approach to science has therefore redefined open and inclusive scientific practices. What used to be a teacher–student interaction in a research laboratory has now become a virtual interaction. These days, there are many scientific experiments in which ordinary citizens are both able to access and contribute to scientific big data in real time across virtual platforms to influence scientific processes – and sometimes, government decision-making processes that affect their daily lives. Engaging citizens in this way enables the general public to take part informally in the collection and analysis of big data and to influence, for example, the local customization of a developmental technology from the West, so that it is adapted to the local needs of a community in the developing world. This kind of public participation will gradually build an educated citizenry and augment the role played by citizens in solving applied scientific problems. The term ciitizen science refers to the public engagement of citizens who actively contribute to science, such as by providing experimental data and facilities for researchers. This fosters greater interaction been science, policy and society and thus more open, transdisciplinary and democratic research.

One example of citizen science is the project on ecosystem services management being implemented by UNESCO and its partners, which has evident linkages to poverty alleviation. The project blends cutting-edge concepts of adaptive governance with technological breakthroughs in citizen science and knowledge co-generation. A set of environmental virtual observatories enable marginalized and vulnerable communities to participate in solving various local environmental problems (Buytaert *et al.,* 2014).

While fostering a culture of open science through the provision of access to big data underpins scientific reproducibility, it also inevitably raises the question of how this type of openness and inclusiveness can maintain accountability for the actions that result from, and affect, these openly accessible data and how the full integration of science and wide participation at all levels can go hand-in hand with respect for intellectual property rights and the avoidance of research duplication or the misuse of data, such as when citation or restrictions on commercial use are ignored.

Researchers are awash with information

With rapidly evolving technologies that range 'from genome sequencing machines capable of reading a human's chromosomal DNA (*circa* 1.5 gigabytes of data) in half an hour to particle accelerators like the Large Hadron Collider at the European Organization for Nuclear Research (CERN), which generates close to 100 terabytes of data a day), researchers are awash with information' (Hannay, 2014).

A recent survey of the research community undertaken by the DataONE project showed that 80% of scientists were willing to share their data with others in the research and education community (Tenopir *et al.,* 2011). Increasingly though, researchers working in data-intensive scientific fields, in particular, are wondering how best to manage and control the sharing of their data and where to draw the line between data transparency for the social good and the risks of an uncontrollable 'data explosion'.

Avoiding the uncontrolled explosion of big data

Global spending on scientific research amounted to PPP$ 1.48 trillion in 2013 (see Chapter 1); the investment made in publishing this research is in the order of billions (Hannay, 2014). Given that interdisciplinary and highly collaborative research fields such as bionanotechnology, astronomy or geophysics are data-intensive and require frequent data-sharing and access, in order to interpret, compare and collaboratively build upon previous research results, resources should be similarly allocated for defining, implementing and communicating about big data governance and for establishing big-data sharing protocols and data governance policies at higher levels of formal scientific collaboration. Even at the level of citizens, the possible implications of 'sharing without control' in an attempt to make science more citizen-friendly could result in citizens being bombarded with an overwhelming amount of scientific information that they can neither make sense of, nor utilize. The creation of scientific big data must therefore go hand-in-hand with big data security and control, in order to ensure that an open and inclusive scientific culture can function properly.

A workshop on data governance organized by the international Creative Commons community in the State of Virginia (USA) in 2011 defined data governance in big science as being 'the system of decisions, rights and responsibilities that describe the custodians of big data and the methods used to govern it. It includes laws and policies associated with data, as well as strategies for data quality control and management in the context of an organization'.[1] Data governance can happen both at the traditional level (universities) and at the virtual level (across scientific disciplines or within large international collaborative research projects).

A code of conduct for digital science?

Big data governance applies to all stakeholders involved in the research enterprise, including research institutions, governments and funders, commercial industries and the general public. Different stakeholders can contribute at different levels. For example, at the more formal levels, governments could create data governance policies in association with affiliated research institutes at both national and international levels. At the level of citizens, people could be provided with tailored educational resources and courses in virtual classrooms to educate them about big data governance. The beneficiaries would be students, researchers, librarians, data archivists, university administrators, publishers and so on. The recent data governance workshop also describes how this type of training could be integrated into the creation of a code of conduct for digital science describing best practices for citizen science, such as data citation and appropriate data description.

By imposing this type of data usage agreement, terms of use clauses and policies targeting funders on open knowledge banks, the way in which these data are globally searched, viewed and downloaded by those interacting with the data archive could be controlled. This would, in turn, shape and differentiate how e-discovery of scientific data takes place both at the formal levels of scientific collaboration and scientific communities, as well at the informal level of citizens.

Big data and openness for sustainable development

With evolving scientific practices nurturing a gradual shift towards virtual science, there is a lot of potential for using and processing openly accessible big data generated from scientific research to help achieve the Sustainable Development Goals adopted in 2015. For the United Nations, 'data is the lifeblood of decision-making and the raw material for accountability. Without high-quality data providing the right information on the right things at the right time, designing, monitoring and evaluating effective policies becomes almost impossible.' The analysis, monitoring and making of such policies will be vital to taking up the challenges facing humanity, as defined by the 17 Sustainable Development Goals and 169 targets comprising *Agenda 2030*.

As a specialized agency, UNESCO is, itself, committed to making open access and open data one of the central supporting agendas for achieving the Sustainable Development Goals.

A mapping exercise[2] undertaken in May 2015 gives a clear understanding of how open science and openness in scientific big data link to the Sustainable Development Goals; this exercise recalls the interconnectedness between the action line on access to knowledge adopted by the World Summit on the Information Society in 2005 and the sustainable delivery of social goods and services to improve lives and alleviate poverty – an interconnectedness that has been the guiding light for the formulation of the Sustainable Development Goals.

REFERENCES

Buytaert, W.; Zulkafli, Z.; Grainger, S.; Acosta, L.; Alemie, T.C.; Bastiaensen, J.; De Bièvre, B.; Bhusal, J.; Clark, J.; Dewulf, A.; Foggin, M.; Hannah, D. M.; Hergarten, C.; Isaeva, A.; Karpouzoglou, T.; Pandeya, B.; Paudel, D.; Sharma, K.; Steenhuis, T. S.; Tilahun, S.; Van Hecken, G.and M. Zhumanova (2014) Citizen science in hydrology and water resources: opportunities for knowledge generation, ecosystem service management and sustainable development. *Frontiers in Earth Science*, 2 (26)

Cooney, C.M. (2012) Downscaling climate models: sharpening the focus on local-level changes. *Environmental Health Perspectives*, 120 (1). January.

Hannay, T. (2014) Science's big data problem. *Wired*. August. See: www.wired.com/insights/2014/08/sciences-big-data-problem

IEAG (2014) A World That Counts: Mobilising a Data Revolution for Sustainable Development. Report prepared by the Independent Expert Advisory Group on a Data Revolution for Sustainable Development, at the request of the Secretary-General of the United Nations: New York.

Tenopir, C.; Allard, S.; Douglass, K.; Avdinoglu, A.U.; Wu, L.; Read, E.; Manoff, M. and M. Frame (2011) Data sharing by scientists: practices and perceptions. *PloSOne:* DOI: 10.1371/journal.pone.0021101

1. See this workshop's final report: https://wiki.creativecommons.org/wiki/Data_governance_workshop

2. See: www.itu.int/net4/wsis/sdg/Content/wsis-sdg_matrix_document.pdf