Astronomy in Focus - XXX
Proceedings IAU Symposium No. XXX, 2018
M. T. Lago, ed.

doi:10.1017/S1743921319005544

The Digital Revolution, Open Science and Development

Geoffrey Boulton

University of Edinburgh & the International Science Council
emai: G.Boulton@ed.ac.uk

Global society is in the throes of a revolution in the means by which information and knowledge are acquired, stored and communicated, which have always been powerful drivers of human material and social progress. Cost reduction and increased flexibility as consequences of replacement of analogue by digital systems, that permit any device with its own power source to collect non-trivial information about its environment, have been the drivers of this revolution. It has been mediated through four key capabilities: high rates of data acquisition; ever increasing data storage capacity; ever increasing computational power; and the growth of the world-wide-web and the ubiquitous communication devices that interrogate it.

This phenomenal growth in the amount and variety of available data has brought the AI learning algorithms that were developed decades ago into their own, for we are now able to feed their voracious appetites for data at rates that were formerly impossible. They are able to reveal hitherto inaccessible but profound relationships in nature and society on all scales from the molecular to the cosmic and in all areas of human concern, from fundamental business practices to the efficient delivery of health systems to global sustainability. There are few areas of individual, commercial, social or political action to which these developments are not relevant. The high rates at which data can be acquired by sensors and ingested and analysed by machine learning algorithms is now also the basis of autonomous systems able to make judgements once thought to be the exclusive prerogatives of humans, and indeed to act on those judgements.

The greatest novel dimension for science lies in enhancing understanding of complex systems. During the 50-60 years since the development of the modern high performance computer, we have become increasingly adept at analysing the dynamics of complex, coupled systems, such that integrated modelling is now the norm. In principle we now also have the capacity to characterise complexity by integrating data from a variety of sensors that contribute data from different perspectives that relate to the same complex phenomenon. The capacity to model complex system dynamics coupled with identification of deep patterns in system states is a powerful new tool in the scientific armoury. However, except for across relatively narrow disciplinary ranges, such as in modern weather forecasting, this is currently impossible because of the different standards and inadequate or incompatible vocabularies and ontologies used in compiling data. The International Science Council's (ISC) Committee on Data for Science and Technology (CODATA) is currently developing a decadal programme, together with ISC unions and associations, to work on this crucial priority for interdisciplinary data integration. These capacities will also be vital in addressing the UN Sustainable Development Goals (SDGs), such as in understanding the complexities of cities as organisms, where most global population growth is located, in understanding the vectors of infectious disease or in evaluating sustainable routes to increased agricultural productivity.

If science is to exploit this potential, it must break out of the siloes within which individuals and groups tend only to have access to a limited range of data reflecting their own disciplinary focus. We have the choice, either to maintain a system were we only have access to that data which we "own", or to develop one where we have access to a much wider range of the data from many disciplines, and which is the necessary pre-requisite to address many of the complexities of this world. I argue that the latter offers the greatest global public good. However, it is a route that is inimical to deeply engrained practices amongst scientists and their institutions, where data that is generally acquired from publicly funded research is deemed to be "owned" by scientists or by their institutions. The public good argument should take precedence, with scientists and institutions moving from a model of ownership to one of access, as has been argued by Science International†.

As the new paradigm of "Open Science" develops, it begs the question: open to whom? Open data and open access publishing still only represent, scientists talking to other scientists, albeit more efficiently. I argue that science must become a public enterprise that engages actively with business, policymakers, governments, communities and citizens, as a knowledge partner, not as a unique source of wisdom, in jointly framing questions and jointly seeking solutions rather than an enterprise conducted behind closed laboratory and library doors or one that is only found in the publicly incomprehensible pages of scientific journals. Without this development science will fail to contribute its unique insights to a changing society.

The web is a crucial part of the infrastructure for open science. It has created a globally networked society that enables unprecedented access to data and information by a great diversity of public and private actors in the creation and use of knowledge. However, although the web may be ethically neutral, its uses are not. A decade ago, it was expected that the web would be a democratising force, enabling debate and engagement between different social actors and inhibiting the creation of knowledge monopolies. The reality has been different. Major web platforms and social media have fostered the development of "echo chambers" in which we are confronted only with material to which we are personally sympathetic. The web has proven to be indifferent to falsehood and honesty, and a powerful enabler of "alternative facts" in a so-called "post truth society". It has been the essential weapon in an environment where interest groups energetically question the credibility and authority of scientific evidence and the trust that science should be afforded.

Further profound issues for society arise from the digital revolution in the reality and potential of autonomous systems powered by AI for: the future of work, cyber-crime and cyber-warfare, the imminent possibility for brain implants, the dangers of privatisation of data and knowledge as data is increasingly valued as feedstock for voracious machine-learning systems, whether and how we create global governance for the data needed to support management towards planetary sustainability, and in analysing the potential existential threat AI might pose to the human role on the planet. The temptation for many scientists is to take the ethically neutral stance of focussing on discovery without considering use. It is now fundamentally important that we do not take the position that these problematic issues are for government and their agencies alone. We in the natural sciences should be working with our social science colleagues to address these issues in a holistic fashion. The opportunity to do so lies in the frame created by the new International Science Council, and should be grasped.

Scientific exploitation of the digital revolution is inevitably faster in those states with well-funded science systems, with the potential to create yet another global north-south

† Note the principles set out in the Science International Accord, Open Data in Big Data World, endorsed by over 120 major scientific bodies worldwide.

knowledge divide. But there are promising moves to avoid this outcome. Collaboration between ISC, CODATA and major African institutions has led to the concept of an African Open Science Platform, to be launched at Science Forum South Africa in December 2018. The Platform's mission is to put African scientists at the cutting edge of contemporary, data-intensive science as a fundamental resource for a modern society. Its building blocks are:

- a federated hardware, communications and software infrastructure, including policies and enabling practices, to support open science in the digital era;
- a network of excellence in open science that supports scientists & other societal actors in accumulating and using modern data resources to maximise scientific, social and economic benefit.

This may only be the beginning. A parallel development to create a similar platform for the ASEAN group of countries† is being led by the ISC regional office for the Asia-Pacific, and discussions are under way about possible creation of a Latin America and Caribbean Platform, with profound and exciting possibilities of a creative and influential south-south scientific enterprise.

† ASEAN – Indonesia, Thailand, Singapore, Malaysia, Philippines, Vietnam, Myanmar, Cambodia, Brunei, Laos

https://doi.org/10.1017/S1743921319005544 Published online by Cambridge University Press