Elias, Peter, "Digital Technology and the Conduct of Scientific Research"

Digital technology and the conduct of scientific research

This note sets out some recommendations concerning the ways in which digital technology influences, affects and shapes the conduct of scientific research. 'Digital technology' is defined here to encompass not only the tools (hardware, software and middleware) that can facilitate discovery, access to, processing and preservation of electronic information but also the technical and managerial skills to develop and apply such activities and the communication systems that enable the sharing of scientific knowledge on a global basis. These recommendations are presented in the three functional areas: data access, knowledge access and attribution.

Data access

The extent, volume and variety of electronic data are expanding at rates that currently outstrip our ability to locate, manipulate and preserve such data for research purposes. Older methods of data collection which have traditionally been used in the social and economic sciences (*e.g.* census and survey methods) now face increasing problems of quality and cost, creating pressures which will lead inexorably to plans to make better use of digital technologies relating to data discovery, collection, access, repurposing and curation. However, while great progress in these areas has been made in some scientific areas (*e.g.* astronomy, particle physics), progress has been significantly slower in areas which make use of personal data[1]. Factors which mitigate against such research include legal barriers (*e.g.* legislation explicitly preventing data sharing or the interpretation of legislation which protects human subjects), ethical barriers (*e.g.* data collected for one purpose not being used for another without the need for consent), administrative barriers (data guardians setting up cumbersome access procedures which are inefficient and/or expensive to operate (*e.g.* safe data centres) and obstacles that arise because of perceptions of 'ownership' or 'intellectual property rights' held by data guardians.

Over the past ten years numerous attempts have been made to overcome these barriers to data discovery, access and their reuse for research purposes. Much progress has been made, as is evidenced via the activities of bodies such as CODATA[2], but progress on the sharing of personal data relating to living human subjects (*e.g.* individual income, social security, health, criminal records) has been difficult, mainly because of legal and ethical barriers.

Despite these obstacles, there is one area where significant progress is now being made. More efficient and secure means of access are now evolving in a number of countries, which make use of virtual safe settings – systems through which authorised and authenticated users can gain remote access to data on individuals or organisations whilst preventing copying of data and minimising the potential for abuse of access privileges.

Recommendation 1: National research funding agencies should collaborate to identify best practice in establishing secure systems for remote access to data held by national statistical offices, national government departments, major private sector companies and other agencies acting as data guardians.

[1] Defined here to include data about specific organisations as well as specific individuals.
[2] See http://www.codata.org/

Recommendation 2: National Statistical Offices should be encouraged to locate, catalogue and assist with access to administrative data for research purposes.

Recommendation 3: Research funding agencies should collaborate in promoting social, economic and behavioural research programmes to encourage innovative research uses of new forms of digital data.

Knowledge access

The traditional method by which knowledge was disseminated involves a close relationship between the authors of scientific articles, peer referees, publishing houses, libraries and scholars. Publishing houses set up domain-specific journals with editorial boards (or academics approached publishers with plans for such). Scientists submit scholarly articles for peer review and (hopefully) publication. Scholars (or their institutional libraries) buy or subscribe to the resulting books and journals to gain access to this knowledge.

This system remains in place, though it is now under severe strain. It worked well in an age when access to knowledge was synonymous with access to the paper on which the knowledge was printed. The advent of photocopying threatened this system until controls were put in place to monitor and refund authors and publishers. However, the widespread demand for electronic access to searchable databases of peer reviewed knowledge is having a substantial impact upon this relationship. The rapid growth in popularity of e-books and e-journals is evidence of the shift away from the traditional ways in which knowledge is made accessible. But who pays for 'free' access to scientific knowledge? Who protects the intellectual property of the authors? What is the role of the institutional library in this new environment?

A further issue relates to the demand for easier and more immediate access to research knowledge. Scientists are now more likely to share research findings with colleagues prior to publication in a peer refereed journal. New journals are being created which exist only in virtual format. Wider access to non-refereed or poorly refereed work not only raises issues about the quality of scientific information now accessible on the web but also poses a threat to the demands that the scientific community are making for better access to published material currently protected via copyright.

These are key issues that must be addressed with urgency. While relevant international associations espouse the principles of open access[3] to research knowledge, further international cooperation and action will be required to ensure that twin goals of unfettered access to electronically available research knowledge and a sustainable system of high quality peer reviewing of such knowledge are achieved.

Recommendation 4: Research funding agencies should establish mechanisms through which they can monitor developments in knowledge access and associated peer reviewing. If appropriate, they should agree collectively on measures that will ensure such knowledge is accessible to the global research community whilst ensuring that quality thresholds are achieved.

[3] See for example http://archive.ifla.org/V/cdoc/open-access04.html (visited 9 April 2010).

Attribution

The third issue which requires global action to resolve relates to the issue of attribution – the ways in which the work of scientists is recognised and accredited. Currently, attribution for research efforts arises primarily through the authorship of scientific papers, books and journals. While scientific disciplines vary in terms of the ways in which authorship is defined and interpreted by others, this system forms the basis of traditional attempts to rate scientists in terms of their productivity, prestige, contribution to knowledge and the impact of their research.

A number of issues are now arising as scientific research moves increasingly from the national to the international sphere. The first relates to the fact that scientists have always been heavily involved in the design, funding, operation and management of large scale research infrastructures. Ranging from multiple array telescopes to genetic databases and major national longitudinal studies, these resources require a major career commitment from research scientists if they are to make their optimal contribution to global scientific enquiry. The rapid growth in the number and scale of this infrastructure means that there are now increasing demands placed on the scientific community to engage in design, funding, and operational functions for major research infrastuctures. While the systems are in place (or are being put in place) to allow international sharing of such facilities for research purposes, the current system for attribution fails to recognise the efforts made by those who help develop, operate, manage and sustain research infrastructure. Without effective attribution for these functions, the future supply of skills, knowledge and expertise required to make them effective could be at risk.

Recommendation 5: Funding agencies should cooperate to undertake a discipline-based review of all current and planned major scientific research infrastructures. This should focus upon their future staffing requirements at the most senior levels, particularly on succession planning. In the light of such evidence, funding agencies may wish to establish internationally agreed procedures through which the scientific contributions made by senior scientists engaged in infrastructure management and direction can be attributed to them.

A further issue derives from the demands for open access to research knowledge discussed under the preceding heading. If the quality of electronically available research outputs cannot be effectively controlled and interpreted in an international context, attribution for research outputs is further jeopardised (see recommendation 4).

Peter Elias

Strategic Advisor (Data Resources)
UK Economic and Social Research Council
(Peter.Elias@warwick.ac.uk)

9th April 2010