

The UK's Research Information Infrastructure: Key Issues and Challenges

by MICHAEL JUBB

INTRODUCTION

Producing, disseminating, and using information is at the heart of any research and innovation system. Like other countries in Europe, the UK faces a number of challenges in seeking to develop an efficient infrastructure of information services to meet the needs and expectations of researchers who are exploiting new ways of operating and communicating with each other. This paper is based on a presentation made at the LIBER annual conference in Uppsala in July 2006, and sets out some of the key challenges, and the role of the Research Information Network (RIN) in seeking to address them.

CONTEXT

In the UK as in many other countries there has been a sharp increase in the volume and scale of research activity over the past decade. Research grants made to UK universities have risen in value by over 80%, and this rise is projected to continue. At the same time, we have seen the increasing integration into all areas of research of some of the conceptual and technological tools developed originally by computer scientists.

One of the results has been the production of increasing volumes and varieties of information and data, with forecasts that over the next five years researchers will produce more data than has been produced in the whole of human history up to now. And there is a dawning realisation of the nature and scale of the new challenges we face in looking after the information outputs from research. At the same time, the shift of publications to digital format has led rapidly to an expectation by researchers that they should be able to get immediate access to publications from their desktop; and there is a growing expectation in some areas of linking between publications and underlying data.

THE UK RESEARCH INFORMATION INFRASTRUCTURE

In response to these developments, there is a growing recognition of how researchers depend on an infrastructure of information services that is highly distributed. Key players in the UK include the national and legal deposit libraries, especially the British Library; other major research libraries, particularly the members of the Consortium of Research Libraries in the British Isles (CURL); the Joint Information Systems Committee (JISC) and its major service providers; major data centres such as the British Geological Survey (BGS), the European Bioinformatics Institute (EBI), or the Arts and Humanities Data Service (AHDS); and the major scholarly publishers and aggregators. The infrastructure of services provided by these and many other organisations is sometimes described as federated, but a federated system perhaps implies a greater degree of clarity and co-ordination than we have at present.

It is thus difficult to get a full picture of the scale of these services, and of the investment in them. If we look just at academic libraries, they typically account for some 3% of the expenditure of UK universities. But it is difficult to disaggregate libraries' expenditure on the support of research from their expenditure in support of teaching and learning. This reflects in part, of course, uncertainties as to what services libraries are, and should be, providing for researchers in a digital world.

If we focus on articles in scholarly journals, however, an interesting picture emerges. The worldwide production of articles has been rising at a rate of about 3% a year. UK library expenditure on serial subscriptions has been rising much faster, and has increased by about 70% over the last decade. But since 2000 in particular, the number of titles subscribed to by UK university libraries has increased much more sharply, by about 100% in just four years, as a result of the coming of big deals. So a simple view of key developments over the past few years is of a rapid rise in the

availability of journal articles, and a fall in the unit cost of subscriptions (but arguably to journals that are not of high priority for many researchers). Clearly, however, the picture is rather more complex than that, and simple talk of a serials crisis on the one hand, or of increasingly easy and cost-effective access on the other, do not give the full picture. We need a better and more nuanced understanding of what is going on.

SCHOLARLY COMMUNICATIONS

Scholarly communications is not a term that would be recognised by many researchers, but it is a convenient label for a process - or rather a series of processes - of central importance for the whole research and innovation system. The processes start from the doing of research and production of information outputs, and include quality assurance and peer review; ensuring appropriate recognition and reward; presenting, publishing and disseminating information outputs; facilitating access and use; assessing and evaluating usage and impact; and preserving outputs so that those of long-term value are accessible for the indefinite future. Developments in technology have brought increasing awareness of the complexity of these processes and the relationships between them; and also a questioning of the functions, roles and responsibilities of the key players, including researchers themselves. Put simply, while the key groups of players share a desire to exploit the potential of new technologies to improve these processes, they no longer agree about who should be doing what, and why and how.

The open access movement in the UK has been given a powerful stimulus by the publication in June 2006 - after innumerable delays - of the Research Councils UK (RCUK) statement on access to research outputs. [1] The statement is based on four key principles, stating that

- ideas and knowledge derived from publicly-funded research must be made available and accessible for public use, interrogation and scrutiny, as widely, rapidly and effectively as practicable;
- published research outputs must be subject to rigorous quality assurance, through effective peer review mechanisms;
- the models and mechanisms for publication and access to research results must be both efficient and cost-effective in the use of public funds; and
- the outputs from current and future research must be preserved and remain accessible for future generations.

Those principles have secured widespread agreement, though there remains a good deal of debate about their implications and how they can best be substantiated. But what the debate has made clear is that we need a much better understanding of how the different roles and functions in the scholarly communications process are currently being performed, at what cost and by whom; of how those functions, roles, and costs are changing; and of the interdependencies between them. There is very little understanding at a system-wide level in the UK, for example, about where and how costs arise in the scholarly communications system, of who meets them, and how. This is one of the issues on which the Research Information Network is trying to shed some light.

RESEARCH DATA

Scholarly communications is now, of course, not just about published outputs, but about data as well. Managing the data that researchers create and collect, as well as the data that they use but which is created and collected by others, is becoming - has already become - an increasingly important and fundamental underpinning for research. The development of skills and expertise in research records management is thus becoming an increasingly important issue for researchers, research institutions, and funders. But there is as yet little sign of the development of clear and consistent approaches to the stewardship of data and other unpublished information outputs.

Librarians, archivists and other information professionals have expertise to offer to researchers in tackling these issues. But such intervention must be based on a clear understanding of the research itself and the different requirements that arise with different kinds of data - from texts and numbers to audio and video streams - generated for different purposes and through different processes:

- scientific experiments, which may in principle be reproduced, although it may in practice prove difficult, or not cost-effective, to do so;
- models or simulations, where it may be more important to preserve the model and associated metadata than the computational data arising from the model;
- observations of specific phenomena at a specific time and location, where the data will usually constitute a unique and irreplaceable record;
- derived data, resulting from processing or combining 'raw' or other data;
- canonical or reference data relating, for example, to gene sequences, chemical structures, or literary texts.

Data may also be generated by different groups of people and organisations, both within the research community itself in the course of research, and in a variety of bodies in the public, private and voluntary sectors for a wide range of purposes, where the data may nevertheless be of value for research. And some kinds of data give rise to issues of privacy and confidentiality.

In the increasingly international context for research and information, policy and practice in the UK, as elsewhere in Europe, must take full account of developments in other countries including the US, with the publication of reports by the National Science Board and the National Science Foundation [2]; in Australia, with the establishment of the Australian Research Information Infrastructure Committee (ARIIIC) with a remit that focuses on improving researchers' access to information; and in Canada, with the publication of a draft policy on access to research outputs by the Canadian Institutes of Health Research (CIHR).[3]

It is also important to participate in relevant international policy forums such as those that have followed the OECD's Ministerial Declaration on Access to Research Data from Public Funding [4], published in 2004, which sets out ten principles to facilitate access to digital research data. An Expert Group has drafted guidelines for governments and research organisations to help guide data sharing among researchers, institutions, and national agencies; and there is a need to ensure that policy and practice at national and local level develops in conformity with these principles and guidelines.

The RIN has thus been working with key partners to raise awareness of these issues and to develop guidelines to facilitate the development of arrangements to help ensure that

- digital research data are created and collected in accordance with applicable international standards;
- the processes for selecting the data to be made available to others include proper quality assurance;
- data are easy to find, and access provided in an environment which maximises ease of use;
- there is credit for and protection of the rights of those who have gathered or created data; and protection for the rights of those who have legitimate interests in how data are made accessible and used.

As with research publications, it is important also that the models and mechanisms for managing and providing access to research data are efficient and cost-effective; and that data of long-term value are preserved and remain accessible for current and future generations. But there is an especial need in this area for clear guidance to ensure that the roles and responsibilities of researchers, research institutions and funders are defined as clearly as possible. This should also help them collaboratively to establish a framework of codes of practice to ensure that creators and users of research data are aware of and fulfil their responsibilities in accordance with principles of the kind set out above.

RESEARCH LIBRARIES AND THEIR ROLE

The role of libraries in supporting research is clearly changing. Most scientific researchers never go near a library, and they probably do not realise that they depend on library funding for the delivery of huge amounts of licensed content to their desktops. Researchers express frustration if they cannot get immediate access to the resources they need, whatever their location happens to be at any particular time; and the complex arrangements to achieve such access require effective collaboration between libraries and between the providers of computer services.

But for arts, humanities and social science researchers in particular, print and manuscript materials (as well as sounds and images) remain of fundamental importance. Most of the human record has not yet been digitised, of course, nor is it likely to be for some considerable time to come; and the volume of printed publications continues to rise.

The persistence of print poses a number of challenges for libraries, and there is an increasing recognition of the need for collaboration at national level in managing and developing collections of printed and microform research resources. Phase 1 of an initiative to develop a collaborative system for preserving and providing access to low-use printed resources - currently termed the UK Research Reserve - began at the end of 2006. The aim is to provide a guarantee of permanent preservation and efficient access for researchers to materials that are currently distributed widely across the UK; and to free up financial, accommodation, staffing and other resources by releasing many libraries from a perceived responsibility to keep copies of such resources 'just in case' someone may wish to consult them. With similar ends in view, the RIN and CURL have established a post to take forward work that has been initiated over the past few years to embed collaboration between libraries in their management of collections of print and microform research resources in specific subject areas, such as Russian and East European Studies.

Of course, even in the arts and humanities, digitisation has proved to be an immensely valuable tool in making research resources more readily accessible and usable; but more work needs to be done if we are to achieve the kind of integration and interoperability between resources that researchers (and others) wish to see.

More generally, we need to stand back and gain a better understanding of

- the impact of recent and forthcoming changes in the ways in which libraries operate in support of researchers,
- how they can best support researchers' needs (for data and unpublished or grey literature as well as to publications),
- what needs can be met effectively at local level, where there is a requirement for collaboration, and where needs can be met only through co-ordination at national level.

As a first attempt to gain some of the evidence we need in order to develop our understanding of these issues, the RIN and CURL have jointly commissioned a major survey of researchers' use and perceptions of libraries and the services they offer, and the results of that will be available early in 2007.

RESEARCHERS AND THEIR NEEDS

At a basic level, researchers' needs are fairly easy to state; but they are rather more difficult to achieve. First, search and discovery are critical and integral parts of the research process, and a recent RIN survey [5] has shown that researchers use a wide range of discovery services - from Google to highly specialised services - to find the resources relevant to their research. As a group, researchers are distinctive in their approach to discovery, and they tend to value comprehensiveness in the results of their searches much more highly than anything else. And there is something of a tension between the frequently-expressed desire for a one-stop-shop for discovery on the one hand, and for the development of even more specialist services for specific disciplinary and subject groups on the other. Somewhat worryingly for librarians, while library portals feature strongly among the discovery services that researchers use and value highly, personal contact with librarians and drawing on their expertise are not seen as particularly valuable, except among some researchers in the arts and humanities. And while librarians tend to think that researchers are inept in their use of discovery services, researchers themselves are confident in their skills, and tend to think that they do not need training from librarians or other information professionals.

Second, researchers want seamless - or at least very well-seamed - access to the resources they discover, with easy-to-use linkages and interoperability between different publications, and between publications and underlying data and other resources. One of the major complaints from researchers is that they cannot always gain access to the resources they identified through discovery services; and there is much evidence that they are frustrated too by having to use a number of different platforms to access the various resources that are relevant to their work. Interoperability is going to be very difficult to achieve, but it is a major issue for researchers.

Third, there is a clear need for more effective and specialised support and training for researchers in handling the information resources that they use, produce and collect. Even if many researchers are confident in their use of

discovery services, many are aware of the need for more support and training to provide the technical expertise that they lack in the handling of digital information resources. And there is a challenge for library and information professionals in seeking to understand more about researchers needs and ways of operation, and to develop themselves the expertise to enable them to provide effective support.

CONCLUSIONS: CHALLENGES FOR LIBRARIES, UNIVERSITIES AND FUNDERS, AND THE ROLE OF THE RESEARCH INFORMATION NETWORK

The complex and rapidly-changing landscape of information provision for researchers presents a number of challenges for all those concerned in the development of the necessary services. There is clearly a need for continuing investment in developing a distributed infrastructure that provides essential services for researchers across a wide range of disciplines, and of institutional settings. But there is also a need for co-ordination to optimise the benefits of that investment, in order to avoid both wasteful duplication and damaging gaps. Co-ordination, of course, needs itself to be built upon a clear understanding of the changes that are taking place; and on greater clarity as to who is – and who should be – doing what, for what purpose and with what impact and interdependencies on others.

The Research Information Network has been established by a consortium of the UK Higher Education Funding Councils (the bodies that provide core block grant to universities and colleges in the four constituent parts of the UK); the Research Councils (the bodies which provide grant funding for specific research programmes and projects); and the British Library, along with the national libraries of Scotland and Wales. It is a small body with an executive team of four. But it has a large mission, to lead and co-ordinate new developments in the collaborative provision of information services for the benefit of researchers in the UK. It is seeking to fulfil that mission by acting as an observatory, gathering evidence on current developments and in particular on researchers' behaviours and needs; bringing key players together; and acting as a broker and an advocate in facilitating and promoting the development of enhanced services.

Researchers are not much interested, of course in who is responsible for the information services on which they depend. But funding bodies, research institutions, and other holders of purse strings do need to address these issues. If we were to be asked at present what would be the best configuration of the different organisations' activities and investments, and whether what we are doing at present represents the most effective use of resources, I do not think that we could give a very good answer.

The fundamental goals that all the key stakeholders in the research and information communities are trying to achieve are to facilitate the advancement of research and innovation; to enhance the efficiency and effectiveness of research; and to maximise the benefits arising from public and private investment in research. Investment in the information infrastructure is clearly necessary in order to achieve those goals, and to ensure that ideas and knowledge derived from publicly-funded research are managed and made available as widely, rapidly and effectively as possible. But if such investment is to be effective, we need to be clear about some key underlying principles.

One of the RIN's aims is to develop a strategic framework for enhancing the UK information infrastructure, and one of the tasks we have set for ourselves is therefore to articulate such principles. Our initial thoughts are that they should include such things as

- the development of codes setting out the roles and responsibilities of the key groups of players, including researchers themselves, as well as funders and research institutions, and the providers of information services;
- standards and quality assurance in creating, collecting and managing data and information arising from and used in research, and in selecting and making them available to others;
- the provision of access in a managed environment which maximises ease of use while protecting the interests of creators and others who have a legitimate interest;
- clarity, efficiency and cost-effectiveness in the use of public funds; and
- operational and financial sustainability over the long term.

Such principles draw on work that the RIN has done in collaboration with others over the past year, in areas including the stewardship of research data and in scholarly communications. We shall be seeking to develop them further over the coming year.

NOTES

- 1 See Research Councils UK' updated position statement on access to research outputs. <http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/documents/2006statement.pdf>
2. Long-lived Data Collections: Enabling Research and Education in the 21st Century, National Science Board, Washington, October 2005, available at <http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>; and NSF's Cyberinfrastructure vision for 21st Century Discovery, National Science Foundation, January 2006, available at http://www.nsf.gov/od/oci/ci_v5.pdf
3. Long-lived Data Collections: Enabling Research and Education in the 21st Century, National Science Board, Washington, October 2005, available at <http://www.nsf.gov/pubs/2005/nsb0540/nsb0540.pdf>; and NSF's Cyberinfrastructure vision for 21st Century Discovery, National Science Foundation, January 2006, available at http://www.nsf.gov/od/oci/ci_v5.pdf
4. The principles in the Declaration are openness, transparency, legal conformity, protection of intellectual property, formal responsibility, professionalism, interoperability, quality and security, efficiency, accountability. See Annex 1 at http://www.oecd.org/document/15/0,2340,en_2649_201185_25998799_1_1_1_1,00.html
5. See: *Researchers and discovery services. Behavior, perception and needs*. RIN, November 2006, available at <http://www.rin.ac.uk/files/Report%20-%20final.pdf>

WEB SITES REFERRED TO IN THE TEXT

- AHDS - Arts and Humanities Data Service. <http://ahds.ac.uk/>
- ARIIC - Australian Research Information Infrastructure Committee. http://www.dest.gov.au/sectors/research_sector/policies_issues_reviews/key_issues/australian_research_information_infrastructure_committee/default.htm
- BGS - British Geological Survey. <http://www.bgs.ac.uk/>
- CIHR - Canadian Institutes of Health Research. <http://www.cihr-irsc.gc.ca/>
- CURL - Consortium of Research Libraries in the British Isles. <http://www.curl.ac.uk/>
- EBI - European Bioinformatics Institute. <http://www.ebi.ac.uk/>
- JISC - Joint Information Systems Committee. <http://www.jisc.ac.uk/>
- RCUK - Research Councils UK. <http://www.rcuk.ac.uk/default.htm>
- RIN - Research Information Network. <http://www.rin.ac.uk/>