Jisc

International advances in digital scholarship

Jisc and CNI meeting, July 2016, Oxford University, UK

October 2016

Authors
Sheridan Brown, Chris Keene, Rachel Bruce and Clifford Lynch

"International advances in digital scholarship. Jisc and CNI meeting, July 2016, Oxford University, UK"

Authors
Sheridan Brown, Chris Keene, Rachel Bruce and Clifford Lynch

Foreword

Rachel Bruce, Jisc and Clifford Lynch, CNI

Jisc and CNI enjoy a long-term partnership and collaboration. This July was a chance to come back together with leaders concerned with digital research and learning from the UK and US and reflect on the direction of travel in digital scholarship. It was also an opportunity for a new generation of leadership in both nations to join the conversation, and for US representatives to meet the new leadership of Jisc.

One of the main themes was collaboration. We really think that the event built on prior connections to see where there are opportunities for collaboration between the UK and US. At the previous Jisc CNI meeting in 2014 we focused on open scholarship and open access; themes of fundamental importance to global scholarship, we returned to these in Oxford and it was striking how much we had learned and how much had changed. Notably, in the US the emergence of detailed funder mandates has reshaped institutional strategies and priorities.

This year, we also explored new, sometimes emergent, areas and in particular areas where practice, policy and technology are advancing in very complex interplay. One example is analytics. Learning and student outcomes analytics can help to enhance the student experience, but must be carefully balanced with issues of transparency and respect for privacy. Analytics and metrics related to research are seeing wide adoption, for better or worse, and these pose numerous challenges and opportunities for improvement; here, again, transparency and reproducibility are emerging as very important issues. Reading analytics and reader privacy are emerging as an interesting new dimension of the analytics developments. Discussion was also lively around the issues of preservation and particularly techniques and practical solutions to achieve long-term, large-scale, management and stewardship of research data and other digital assets.

We noticed in all discussion topics participants were keen to learn from each other. Often there were concrete, practical steps that could be pursued in both the UK and the US to move ahead with the issues.

This report describes the diversity of the discussions, conclusions, and potential follow-on actions, however perhaps for us three of the most compelling points are:

1. There are new concerns around openness and transparency when it comes to analytics, whether for research or learning; this is about the openness of algorithms and being transparent about the data used to drive any analytics or draw insights and metrics.

2. There is a common desire and need for consistent and open identifiers, for example for digital outputs, people, places, organisations, etc. Progress in this area has been made but there is much further work to do. We need a commitment to global solutions and serious commitments to interoperability here.

3. Shared infrastructure and services is a very live issue on both sides of the Atlantic and how to create sustainable infrastructure is key, but also there was, we think more strongly than ever before a desire not to re-invent the wheel and to genuinely think about sharing services more widely, while recognizing the real organisational challenges of finding ways to do so.

We think you will find this report interesting. Let us hear your thoughts and reactions. We look forward to continued dialogs and collaborations between our organisations and our member institutions in both nations. We believe that we are fortunate to have this partnership and channel, it helps to find better solutions for higher education and research in the UK, the US and beyond.

This report would not be possible without the valuable contribution of all those who spoke and participated at the conference, we would like to thank all for their insight. We would also like to thank Joan Lippincott and Neil Jacobs for their contribution to this document.

Contents

Introduction

The 2016 Jisc and CNI one-day conference brought together leaders from the UK and the US academic and research communities concerned with scholarly communications.

A report on the proceedings of the entire event is available separately but this document is designed to bring together into seven themes the key points from the conference. Following the conference a group of experts from the UK and US were invited to spend a second day to consider and identify opportunities for greater collaboration through detailed discussions on two themes: first, creating, sustaining and using the scholarly record; and second, improving the ability to re-use the scholarly record both for research purposes and more broadly to advance the enterprise of scholarship broadly. Outcomes from those discussions have been woven into this analysis where appropriate. In particular, ideas about opportunities for collaboration between the UK and US have been distilled from both days and are presented towards the end of this document.

Jisc and the CNI have been working together for over a decade to look ahead, plotting a course through an environment increasingly characterised by uncertainty and discontinuity. Paul Feldman, Jisc's Chief Executive, believes there is an opportunity for both Jisc and the UK scholarly communications sector working with collaborators globally to move forward with the development of a number of high impact programmes in fields such as open access, open science and open data, acting as an effective agent for change in the research and education sectors.

The Executive Director of CNI, Cliff Lynch, regards these joint events not only as opportunities for building relationships and learning from each other, but for identifying common challenges that might be best addressed through a collaborative process. This is the context for the conference: the sector is faced with long established and newly emerging challenges.

Infrastructure

Research Infrastructure refers to facilities, resources and related services used by researchers to conduct research. As the complexities of the research environment grow, so too does the need for supporting infrastructures to enable researchers to operate effectively.

A continuing problem is how to ensure the sustainability of scholarly communications infrastructure, and how best to ensure it is governed in the interests of the research community.

There are a number of examples of important developments in the infrastructure arena, though many of them face ongoing sustainability challenges as they transition to operational status.

The team behind SHARE[1] is building a free, open dataset about research and scholarly activities across their lifecycle in support of the SHARE team's mission to maximise research impact by boosting accessibility, discoverability and re-usability of research data in line with the FAIR[2] Guiding Principles for scientific data. Jisc's Monitor service offers two separate but integrated services for tracking information about open access publications, one at the local institutional level and one at the national level. Monitor integrates with other infrastructure such as CORE[3] and Crossref[4]. The service will go live in autumn 2016 for the 23 universities that were involved in the pilot phase.

Open access advocacy and policy has progressed and matured at a quicker pace than open access infrastructure and implementation. As the community moves forward with implementation it is thought desirable for it be open– in terms of metadata, standards and identifiers and code, platforms and APIs. In some areas, however, notably with CRISs provided by commercial suppliers, there is clear divergence from the ideal of open platforms and code. Institutions that have not yet gone down the route of procuring a commercial CRIS could potentially use shared infrastructure such as that which would be provided via the proposed national research information infrastructure should this vision move to implementation. Throughout the meeting there were many hopes for the future: Examples: a strong hope that metadata does not become privatised; a desire for consensus in the development and adoption of broader indicators of quality; funding for infrastructure development that facilitates interdisciplinary and international collaboration, and grid of global pre-print servers to enable text mining across disciplinary boundaries.

From the researchers' point of view the infrastructure, and the ways in which different systems interoperate, should be invisible in much the same way that most people rarely give a second thought to how power, water and telephony services are provided to their home or office. In the research data space Jisc has been developing a shared service[5] to enable universities to meet the requirements of funders' with respect to research data management. A shared service approach offers cost savings and related efficiencies, common approaches and practice and the prospect of standardisation and interoperability, the benefits of which extend far beyond the service itself. The Alpha development phase begins this summer and, assuming a positive trajectory in terms of feedback and acceptance of the business case, the shared service should begin to be rolled out in April 2018.

In tandem with the development of research data management solutions there is an important focus on digital preservation. In many cases funders expect data to

be retained for a minimum of ten years and potentially much longer. Barriers to access and re-usability include the inability to read files in software formats that have been superseded or because of expired software licenses. Archivematica was developed to address the issue of digital preservation at the technical level. York University discovered that of the research data deposited over the course of a year nearly two thirds of the files were unidentifiable using DROID[6], a file format identification tool; files that cannot be identified are difficult to preserve. This points to the need for more community-wide research into file signatures to improve the PRONOM[7] technical registry.

One important strategy for digital preservation is the use of emulation of a software environment, including the underlying operating system, as opposed to migration, which looks to update the data files to be compatible with modern software, which can lead to compatibility problems. The emulation approach comes with the overheads of bundling a large amount of software with the original data, though can be seen as a more secure way to ensure the data remains accessible unaltered in the future. In the repository space Hydra-in-a-box is an interesting new development and example of a next-generation repository system, in a space that has mainly seen incremental developments over recent years.

Among the many interesting ideas to gain ground during the event was that of data as infrastructure. It is postulated that some components are so important that they act as nucleation points for other data. For example, open linked data from the Land Registry or Companies House, both in the UK, provide important connection points. There are calls for more registries of useful information to play similarly important roles, backed up by relevant sets of persistent identifiers.

In the US, while national initiatives exist and provide highly valuable services, the nature and structure of funding makes such initiatives, and hence shared infrastructure, relatively rare on a national level. To complement this, we are seeing important state and regional developments; and a variety of membership organisations have also become a way of approximating national infrastructure at least for the major research universities. There is also a great deal of development at the institutional level, or by very small groups of universities working together and disciplinary-oriented infrastructure development, led by funders, is also gaining momentum. In the UK, Jisc has made a bold proposal for the development of a new National Research Information Infrastructure (NRII) based on open protocols. The goal is to enable the efficient collection and dissemination of richer and more reliable information than is currently available allowing for better benchmarking and business intelligence, a simpler process for making REF[8] submissions and a straightforward route to compliance with funders' policies. This is believed to be a development on which the UK and US can work together, sharing technology, ideas and expertise.

[1]

1 share-research.org
2 force11.org/group/fairgroup/fairprinciples
3 https://core.ac.uk
4 crossref.org
5 jisc.ac.uk/rd/projects/research-data-shared-service
6 http://ji.sc/preserving-digital-records-droid
7 nationalarchives.gov.uk/PRONOM
8 ref.ac.uk

Analytics for learning and research

Metrics for research outputs have long been established, if controversial, in the scholarly world. However, we are seeing the rise of data driven metrics, indicators and analytics growing rapidly across academia both in research, and teaching and learning.

There is a view that measurement leads to more objective and, therefore, presumably, better decisions. In the research sector governments use metrics to demonstrate value for money; funders use metrics to help with the efficient allocation of research funding and demonstrate public accountability; and universities are keenly aware of rankings and REF-related metrics. There is an alternative view that many commonly used research metrics are meaningless – or worse – without proper understanding of how they are derived and without proper statistical caveats. They can create distorting incentives. The development and publication of research metrics looks to have become a significant source of profit and control for commercial organisations involved in delivering measurements using those metrics. This particular worry is symptomatic of a wider set of concerns various metrics and the data needed to compute them: lack of transparency, lack of reproducibility, biasing of metrics to specific (non-transparent and managed) universes of content or interactions taking place within so-called "walled gardens". This has led to calls for the research sector itself to curate and utilise better the basic data it produces and from which more transparently based metrics could be derived.

There is a desire to see the development of open collections of basic citation-related data. Currently the bulk of citation data is held in proprietary commercial databases controlled by companies that sell data back to the community; in addition, the data is often not open for computational analysis by third parties. CORE is a UK-based aggregation of full text open access papers that currently offers around four million full text items. CORE is building a dashboard for repositories to expose citation counts but the number of bibliographic records is currently insufficient to enable it to be used as a substitute for commercial citation indexes. In the US, it is hoped that over time the SHARE effort will produce a valuable database of research outputs, but there are currently no plans to try to build this retrospectively. Services that do have large aggregations of bibliographic records such as LOCKSS and Portico are typically governed by licenses that restrict secondary use. In any case, there is an argument for investing effort in developing new, alternative metrics that rather than trying to replicate existing data and citation-based metrics. A number of organisations and researchers, including Jisc, currently have initiatives to help explore potential new metrics.

Meanwhile, it was noted that researchers don't like metrics or understand why they are being measured; there is a fear that understanding of what researchers do may be lost in reductive number crunching. Researchers tend to prefer a wider range of "indicators" rather than the narrow set of publishing-related metrics that are most commonly used. Metrics can leave some researchers, such as those who take a non-traditional approach, at a disadvantage, and reward those who game the system. They lead to citation clubs, research slicing and an unbalanced focus on popular or non-controversial research topics. Commonly used research metrics are based on opaque data sources and questions have been raised about the veracity of systems that seek to quantify the qualitative outputs of the peer review process. The interest in, and concern about, research metrics led HEFCE[9] to commission a

review, chaired by Professor Wilsdon, of the role of metrics leading to the publication in 2015 of the Metrics Tide report[10] . More recently the Stern review[11] of the REF has been published, both reports upholding the primacy of peer review over the wholesale use of research metrics.

Metrics need to be used appropriately and responsibly and that institutional and funders' policies and practice must reflect this. The worldwide Declaration on Research Assessment[12] (DORA) was established to stop the uninformed use of journal impact factors to assess the contributions of individual researchers and their promotion prospects. DORA is struggling for traction in the UK where so far only six institutions have signed up, though those that have, have found it useful to review their research reward practices. At a more fundamental level, there perhaps needs to be more emphasis on what we are trying to measure and why. Institutions commonly state that they strive for "excellence" without a common understanding of what that means. Implicit assumptions about excellence need to be made explicit: we need to know what are the qualities of research that deliver desirable results. Excellence may include benefits in the form of economics, health, environment, culture, education and further research. The pursuit of beneficial outcomes that deliver for society should eclipse the simplistic assumption that the number of citations is a proxy for excellence.

University teaching has seen the rise of learning analytics, exploiting and aggregating the large amount of data collected on learners and their activities, both in the digital and physical space. This has given new privacy concerns from several perspectives. Issues about data reliability and security, access and licensing requirements are evolving in complex ways. Efforts by libraries to protect users' personal information to guard against surveillance may, in fact, be driving users towards alternative content platforms that have fewer data protection barriers and are therefore easier to use – but where there is also a potentially greater prospect of the misuse of personal information. Much more work is required on the concept of "informed consent" for student-related analytics as part of a broad policy framework in this area. One approach, advocated by the Open University in the UK, is to enable students to become active participants in the collection and management of their own personal data. Learning analytics service providers have a responsibility to ensure the security of data collected about students; Jisc's learning records warehouse, part of its national learning analytics architecture, contains highly sensitive information and maintaining the security of that data has been a top priority in the design of the system.

[1]

9 hefce.ac.uk
10 http://bit.ly/metric-tide
11 http://ji.sc/research-excellence-framework
12 ascb.org/dora

Researcher culture, incentives and skills

The research environment, in its broadest sense, continues to develop rapidly. A researcher returning to the sector after a decade away would find themselves in a relatively alien landscape, from practice to policy, from technology to funding, and with the many additional requirements that are expected of them.

As these change continue, there is need for much more effective communication among researchers, librarians, and policy makers. In particular, researchers must find a more central engagement in policy discussions, particularly as policies are developed which will force further shifts in research culture.

The requirement to drill down into the question of how the policy environment shapes digital scholarly behaviour in the UK and the US has been recognised, as has the need to focus on equipping researchers at all levels – but especially new and early stage researchers - with the capabilities needed to flourish in the digital world.

It has also been recognised many of the hardest and most important problems are complex and multidisciplinary in nature, and call for collaborative teams of researchers. Thus, there is a need to better reward collaborative research. At present, however, metrics are extracted at an individual level and the most commonly used measures reward citations. Metrics are rarely designed to register the network benefits that flow from collaborative work. Some projects attempt to award fractional credit but it can be difficult to assign a true value to an individual's contribution to a research team.

In terms of incentives, there was a concern over researcher promotion boards using, sometimes disproportionately, metrics such as the Impact Factor of journals a researcher has published in and their articles citation counts, and that this may unduly influence researchers' promotion prospects. Though it is unclear how wide spread this problem currently is. In reality the decision about promotion is complex, and must consider a wide range of factors beyond publication records. They include the impact of the research on wider society, public engagement, leadership skills and quality of teaching. The relative importance that has been placed on narrow citation measures has largely been driven by researchers themselves (although administrators have willingly embraced their use) and there is a call for senior academics and faculty leaders to openly object to the weight commonly given to citation metrics in order to protect early career researchers from the effects of their injudicious use.

As the policy and compliance landscape evolves apace, so the ways in which researchers are equipped and incentivised to produce various types of research outputs should likewise be reviewed.

Researchers are focused on research, and while they are generally sympathetic to open access to their publications – so long as the impositions on their time are limited and does not call for major changes in their publishing practices - they are largely unconcerned about the costs of scholarly publishing since they are not typically exposed to them.

Views on open access to researcher-produced data are much more complex and varied; they are quite sensitive

also to disciplinary and sub-disciplinary cultural norms, and in a very substantial number of cases, constrained by issues involving human subjects. Even in cases where the researcher is willing to share data (or is mandated to do so by funder policy) there is a substantial amount of work involved in curating the data into a form suitable for sharing and effective re-use, and this is creating a growing need to build up research data management support infrastructure and staffing.

Research increasingly requires a broader skill set, especially around software code and data science. Many develop their own software despite often not having any formal training. These skills are rarely explicitly recognised or rewarded through the evaluation system (though REF does reference software as an output). There is no credit given for releasing software, making it easy to use or supporting others in its use. There are similar issues around the development and deployment of network-based services that are widely used within scientific communities. This underpins the claim that the current research culture present barriers but few incentives to properly equip researchers for the modern age of software-enabled research. International initiatives such as software carpentry, aim to develop software skills of researchers. The Software Sustainability Institute[13] provides a national facility to support and foster the UK research software community. However, the question of credit and reward for research software and data remains unanswered, along with the issue of career paths for non-faculty professions such as software engineers who play many key roles in the modern research enterprise.

Discussions about open access, open data and the more general cultural shift that puts software and systems near the centre of the research enterprise tend not to include the patent system. Universities have a commercial incentive to lock up its research outputs using patents. There is a balance of interests between the outputs of publicly funded research to be open and the desire of institutions to capitalise on their researchers' endeavours. There is no resolution at present, though the question of the extent to which patents impede or support innovation is open to discussion in the community. Institutions should perhaps think about other ways of extracting value. It has been argued that while the world of protected intellectual property is not going away, the speed of innovation is such that there may be more value in not hindering innovation by applying the patent brake.

[1] 13 software.ac.uk

Advancing data sharing, re-use and openness

Openness is more than just an ideal but as a means to increase research quality and efficiency. We have become accustomed to consider open access as representing a binary state: something is either open or closed at any given point in time.

While this has some resonance in the world of journal articles, the world of data is more nuanced, not all data is suitable to be made openly available. The Open Data Institute[14] has introduced the idea of shades of openness from closed, through shared, to open.

At the closed end of the spectrum, "internal access" data, only available to members of an organisation, includes both enterprise data, such as salary information, as well as some types of research data. These include active research, and sensitive data containing personal or confidential information, as well as data that is restricted due to technical reasons. Moving along the spectrum sees increasing access to the data, either to specific authenticated groups or publicly available, and finally open data with respect to rights as well as access. Policy needs to be developed, and technical and legal barriers overcome, to increase the amount of research data which is on the open end of the spectrum, where this is appropriate.

The power of open data is wide scale participation and more needs to be done to encourage data creators to provide access to their data. There are many examples of value that has been created by making data open, from human genome data to data on mortality rates during the Ebola crisis where access to data enabled huge global effort to be focused on finding a solution. The arguments for the benefits of open data need to be repeated often to meet the challenge of the predisposition of many public and private owners of data to privatise or monetise access or simply to keep it closed.

[1] 14 http://theodi.org

The changing role of the library

Libraries have been adapting to external changes for many years with shelving for printed books and journals making way for a variety of learning spaces, resources and technology.

It is widely acknowledged that the university library has important roles to play in terms of the following:

- Providing a gateway for locating information for research
- Acquiring the resources required
- Archiving and preserving materials
- Supporting and facilitating teaching activities
- Supporting the research process
- Helping undergraduates develop research, critical analysis and information literacy skills

This list encapsulates the basic business of a university library but the ways in which things are done today compared with two or three decades ago are starkly different. The evolution of the role of the library may be witnessed through the lens of modern job roles: digital scholarship editor; designer for online publications; science data librarian; web content specialist; digital preservation librarian; digital humanities librarian; digital repository manager; data visualisation coordinator – and so the list of roles involved in supporting digital scholarship goes on. The digital scholarship centres that have become fashionable are often located in the library providing a focus for this type of activity. More so than ever the library is more than a place to meet and study – it is the embodiment of a broad set of services to support research. Space alone no longer defines the library: it needs to be valued both in its own right and as a complement to the new research environment.

Evidence of the changing role of the library is found in the US where increasingly university presses are being reorganised under the management of university libraries. This is leading to a significant re-thinking of the funding models for these university presses, and indeed the ways in which the historical need for presses to be financially self-supporting, or close to it, had distorted thinking about the purpose and agenda of these presses. This is closely connected to a broad conversation and a set of experiments and initiatives that are taking place, primarily in the US, to rethink the financial models that support the publication of scholarly monographs. Open access is connected to this discussion, but is not a given.

The library has a clear path to readers who often rely on library-mediated discovery mechanisms. Libraries also have relationships with authors and are increasingly providing support for developmental editing for digital works including putting effort into design, layout and readability. This shift to the library serving as the agent of an author in terms of helping develop the digital work and make it available – through the university press or otherwise – is very strategic and significant. It offers a potential resolution to the situation where scholars are having difficulty disseminating the results of their digitally based research to modern readers and getting credit for it, while ensuring the these results can, and will be, preserved.

Another important development that is emerging, particularly in the US, is a growing set of collaborations and institutional focus on the stewardship and accessibility of collections of all kinds, held in museums, archives, libraries and even academic departments within the institution.

Sustainability and costs

Sustainability is a topic of interest in research and scholarly communications. How do we ensure the services, infrastructures and processes are a secured long term future for example in terms of funding and governance, especially with the change in business models bought about by open access and the move to digital.

In the UK the government's goal is for almost all scientific output to be open access by 2020 where realistic and affordable. It has been described as broadly favouring the gold route, though the recent REF open access policy steers towards the green route of repository deposit, and should lead to a significant increase in open content. In the US a number of Congressional bills and national initiatives, including FASTR[15], are leading to a mixture of different routes to open access.

In a situation where Article Processing Charges (APCs) costs are rising – which tend to be greater for better quality journals - there is likely to come a point at which the gold route preferred in the UK becomes unaffordable despite negotiations with publishers for offsetting deals. There is a suspicion that a "market" based on APCs, one where publishers set the APC level wherever they choose, will be as dysfunctional as the subscription-based market on the well founded assumption that authors tend to choose to publish in the best quality journal they can in their discipline. It is generally thought that current library budgets would not cover the additional costs of switching from the current subscription model to one based solely on APCs to purchase the same portfolio of publications. One of the key conclusions of a cost modelling project, Pay It Forward[16], is that giving authors discretionary funds introduces APC price competition without interfering with author choice in where to publish. This is based on the study's finding that authors are price sensitive when they have discretion to choose where to publish based on cost and perceived quality and reputation of the journal.

Studies in the UK depict notable increases in the total cost of publication as institutions pay for both subscriptions and for open access publications using the gold route plus the attendant additional administration costs. This is primarily an issue driven by the use of hybrid journals, subscription journals in which some articles are open access due to an APC having been paid. Offsetting agreements are designed to prevent publishers being paid twice for publishing the same article to reduce the administrative burden on institutions and to speed the transition to full gold open access. Despite the offsetting agreements costs are rising inexorably and the revenues of the top ten scholarly publishers in the UK continue to increase. The Springer Compact agreement has "flipped" the publication payment model such that the bulk of payments from institutions to Springer are in the form of APCs with relatively little being spent via the traditional subscription fee route. This has simplified the administration process for institutions and helped with policy compliance. However, there are warnings that even with these initiatives, the costs may not be sustainable for research-intensive institutions in the UK and US.

Looking ahead, several steps have been suggested to help promote the sustainability and affordability of the publishing process: the use of research funding for publishing in hybrid journals may be somehow restricted; funders' policies should give greater support to the green route to open access; greater support could be given to smaller society publishers, university presses or other innovative business models. Importantly, there needs to be international

cooperation to avoid the creation of a multiplicity of different policies and models that will lead to confusion and hamper the wider research community's ability to constrain the growth in publishers' prices for subscriptions and APCs. This cause is not helped by the main government research funders in the UK and the US officially preferring very different routes towards OA. It may be that these positions moves closer as the impact of funders' policies, particularly the REF-related open access policy in the UK, begin to have an impact. It may also be that APCs are transitional and that membership models could become more attractive. We are living through a period of profound change in the scholarly publishing arena and it is not yet clear what the outcome will be, but more work is needed on both sides of the Atlantic to find a path to an equitable and sustainable situation.

Turning to the sustainability of the infrastructure that underpins open access, there is still much debate about how such infrastructure should be funded. Even if funding programs are established, the research grant system is not a solution in terms of infrastructure sustainability. Since grants tend to have short time horizons (except in the case of large equipment purchases), they tend to fund one-off projects and make it difficult to build interlocking infrastructure ,and short funding horizons make it hard to retain good people. Grants are designed to fund innovation on the whole whereas there is a need for money to maintain infrastructure and to allow those services to continuously invest in the development required to adapt to changes in the environment. At best they can be a good source of venture capital for infrastructure prototypes or research that may inform or lead to the development of infrastructure, or perhaps for parts of the funding for initial infrastructure construction. They cannot sustain effectively.

There's an additional problem with using grant programs to fund the development of infrastructure, or even research projects that are intended to form the basis of, and evolve into, infrastructure: infrastructure calls for scale, including scale of adoption. Programs that fund proposals based in intrinsic scientific or scholarly merit do not necessarily lead to scale, but often fragmentation – multiple systems arising out of various communities that may take different approaches, and almost always different code bases, to accomplish similar functions. There is a point at which efficient allocation of resources demands that someone choose winners and losers.

[1]

15 http://ji.sc/fair-access-to-science
16 http://icis.ucdavis.edu/?page_id=713

Areas for collaboration

During the course of the event a number of opportunities for collaboration between stakeholders in the scholarly communications sectors in the UK and US were identified. A summary of these ideas is presented here.

Preservation and curation

- Curation is often confused with, and equated to, long-term preservation. Curation is not just the process of keeping data, but preparing them for re-use as appropriate, and defining and managing the life cycle. Having said that, there is still a serious problem with the creation and support of the infrastructure that will deal with genuinely long-term preservation

- There is a growing realisation that it is not possible to curate all the data produced by the research sector and that both general and disciplinary level models need to be developed to identify which data need not be kept (perhaps because it is just experimental "noise" in the active research phase) and which data has a limited life span in terms of its usefulness and can be deleted after a period of time

- One of the key challenges for institutions in dealing with data curation is how to reasonably accommodate disciplinary differences. There is work to do at the international level to develop disciplinary standards where they do not currently exist or where they are inadequate for the task of data curation

Open access publications

- The process of complying with evolving funders' requirement has been putting a strain on institutions' technical and administrative capabilities especially in terms of managing APCs. Shared approaches to achieving compliance through perhaps technical development, national level agreements with publishers and sharing best practice could be useful in reducing the burden on individual institutions

- The community needs to continue to investigate what the future market for publications should ideally look like and dare to think beyond marginal changes to the current system. Might APCs just be a temporary fix as the sector transitions to full open access? How sustainable are APCs for hybrid OA journals? How might stakeholders in the UK and US come together to present a united front in the face of the prospect of increasing APC costs?

- In the US and the UK different agencies have been producing policies and mandates that differ from one another to a lesser or greater extent and which make the process of compliance burdensome. Collaborative work to both persuade agencies to bring policies into closer alignment and to make it easier for individual institutions and their researchers to understand what is required of them and help them achieve compliance would be very useful

- The question of open access for monographs (and indeed the broader related question of funding monographs) is still largely unsettled; to date, funders have avoided this topic, by and large, and there does not seem to be a strong consensus among researchers that monographs should be open access from the start (though there is perhaps more consensus about opening up older and out of print works). This is an opportunity for joint work leading to a common vision and roadmap

- It would be beneficial for UK and US HE colleagues to work on a set of joint principles and requirements on the basis that speaking with one voice increases the chances of achieving the desired outcome

New publishing models

» In light of the fundamental changes facing the publishing sector it would be useful to look again at the potential role of university-based publishing and more broadly that of the library in the dissemination of scholarship

» New born-digital long form publications are providing value beyond that available through basic pdf versions of standard monographs but there is work to do on standards, licensing and preservation

» Some institutional libraries in the US are breaking new ground by providing editorial and other publishing services to faculty members. This disruptive and potentially transformative development merits close attention and could lead to a wider scale adoption of this approach

Metrics and openness

» There is considerable concern about proprietary algorithms that compute measures of impact. Equally troubling, however, are algorithms – proprietary or open – that rely upon proprietary (and often non-transparent) collections of data in order to deliver their measures. The great historical example of this was citation databases; as new measures are being used, however, they also rely on transaction data from proprietary systems or "walled garden" environments for sharing scholarly work

» The problems stemming from the extent to which bibliographic/citation data is mainly proprietary and from which opaque metrics, which are difficult to replicate, shape the actions of individuals and institutions were widely discussed. Work towards building a common open bibliographic and citation database on which both the public and the private sectors would be able to build services is required. Further, such a database should have clear and broad criteria for inclusion, and might well encompass monographs as well as journals, and open access journals and preprint servers as well as the traditional, primarily paid subscription based scholarly journal literature

Learning and research infrastructure

» There is the very real prospect of synergistic development at the infrastructure level between, for example, services like SHARE in the US and CORE in the UK

» There has been a lot of work done on the sustainability of scholarly infrastructure in support of open access to publications and latterly data, but the idea for a white paper focused on this topic and encompassing requirements in the UK and US received support. This would include things such as how best to develop and maintain the infrastructure components, a review of the rules of engagement and further examination of the relative merits of the range of support and governance models that were discussed during the second day of the event

» Many of the discussions involved to some extent the importance of standards, accurate and appropriate metadata and persistent identifiers - all in the service of interoperability. Solid progress has been made on identifiers for authors (though there is still more to do), and there is hope that an organisational identifier is within reach but there remains a lot of work to do in this area and probably will be for the foreseeable future

» In the US there is a diverse approach to innovation in research infrastructure, whereas the UK has focused on shared national infrastructure, both have their advantages in this critical area. The sharing and joint development of new and innovative infrastructure should be explored

Share our vision to make the UK the most digitally advanced education and research nation in the world

jisc.ac.uk

Jisc
One Castlepark
Tower Hill
Bristol, BS2 0JA
0203 697 5800
info@jisc.ac.uk