Towards Better Access to Scientific Information

A pre-workshop review and synthesis of related literature

European Coordination Workshop

on access to and preservation of scientific information

Paul Stokes : JISC

Overview

This synthesis is an attempt to condense approximately 500 pages of densely worded discussion, conclusions and recommendations into something that can be skimmed and understood in less than an hour. As a result there are omissions and some loss of the original authors nuances. However, links have been provided in the bibliography for those who wish to get their information from the original source. The papers are presented in chronological order.

In simplistic terms the current situation could be summed up as:

- There's a LOT of data out there[1]
- There are many, many scientific papers extant[2], a significant number of which are behind pay-walls
- Making this knowledge and data freely available has positive intellectual and economic benefits
- Keeping it is difficult and expensive

Reality is, of course, much more complex than that.

Recognising these factors and the drivers behind them the European commission has long been encouraging member states to take steps to maximise any benefits that could accrue from "the rising tide of scientific data" and minimise any detrimental factors. The current wave of initiatives could arguably be said to stem from the paper "Riding the wave—How Europe can gain from the rising tide of scientific data" published in 2010. In this paper the vision of a seamless, trans-European scientific e-infrastructure by 2030 to ensure a competitive Europe was outlined.

The Knowledge Exchange group responded in 2011 with "A Surfboard for Riding The Wave—Towards a four country action programme on research data" (taking the nautical metaphor one stage further) concentrating on the data infrastructure. This paper presented an action plan for the four countries involved, first outlining the current situation and then outlining a series of mid-term and long term objectives which would allow the "Riding the Wave" vision to be realised.

The nascent collaborative efforts surrounding data preservation are addressed here by the inclusion of the 2012 journal "Aligning National Approaches to Digital Preservation". The essays contained within arose from discussion at the May 2011 "Aligning National Approaches to Digital Preservation" conference and subsequent exchanges between participants. Not surprisingly, being a rapidly developing discipline, there were a large number of recommendations developed.

The question of Open Access (OA) as a possible sustainable business model has mainly been addressed here in the form of the 2012 Knowledge Exchange paper "Sustainability of Open Access Services" in which a series of use cases are examined with a view to establishing their viability (or otherwise).

This year the European Commission added to the current debate with the July communication "Towards better access to scientific information: Boosting the benefits of public investments in research" and the associated recommendations—the key driver for this workshop. These outlined the steps the commission intends to take to improve access to scientific information (papers and the underlying data), in particular the information that has been paid for by the public purse.

[1] Estimated at 1.8 zettabytes (1.8 trillion gigabytes) in 2011—IDC Digital Universe Study, [http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf] June 2011—and doubling every 2 years. The current 2012 estimate from IDC is 2.7 ZB.

[2] Arif E. Jinha from the University of Ottowa estimated the number as 50 million in 2010 in his paper "Article 50 million: an estimate of the number of scholarly articles in existence" [http://www.stratongina.net/files/50millionArifJinhaFinal.pdf]

Overall the recurring themes in all the referenced papers are:

- Public data should be freely available, freely usable and preserved
- Policy/enforcement/collaboration is required to ensure that availability
- Interconnected, fit for purpose e-infrastructures are required
- Piecemeal approaches will delay intra European integration (and intra global)
- It's expensive (but there are significant benefits)
- All stakeholders need to be made aware of the benefits

Riding the wave—How Europe can gain from the rising tide of scientific data

The major message/vision from this paper is that, to ensure that Europe remains competitive, a seamless, scientific, trans-European e-infrastructure needs to be in place by 2030.

Key aspects of this vision include:

- Raising the awareness of *all* Stakeholders as to the critical importance of conserving and sharing data produced during the scientific process.
- Researchers/practitioners being able to find, access, & process the data they need. They need to be confident in their ability to use and understand data, and be able to evaluate the degree to which that data can be trusted.
- Producers of data prefer to deposit their data in reliable repositories that are built using international standards.
- Funding bodies have confidence that their investments in research are paying back dividends to society through increased use and re-use of publicly generated data which should in turn lead to an increase in funding.
- Efficient exchange of data between private and public sectors allows both to benefit, especially when innovating
- The public can use, contribute to enrich the vast data store.
- Policy makers can make decisions based on solid evidence, and can monitor the impacts of these decisions.
- Global governance promotes international trust and interoperability.

The role for the European Commission and government(s):

1. Develop an international framework for a Collaborative Data Infrastructure—a broad, conceptual framework identifying how different companies, institutes, universities, governments and individuals would interact with the system.
2. Earmark additional funds for scientific e-infrastructure—for example the EU Structural funds
3. Develop and use new ways to measure data value, and reward those who contribute it—if we can't measure it we can't develop reward mechanisms.
4. Train a new generation of data scientists, and broaden public understanding—new policies are required to foster the development of advanced-degree programmes for the emerging field of data scientist. Data management and governance considerations need to be included in the curricula of secondary schools, as part of a European wide IT familiarisation programme.
5. Create incentives for green technologies in the data infrastructure—the e-infrastructures needed to deal with the rising tide of data will impact upon CO_2 management plans. Steps need to be taken now to mitigate that impact.
6. Establish a high-level, inter-ministerial group on a global level to plan for data infrastructure—The European Commission needs to identify and resource an international group to manage e-infrastructure planning.

A Surfboard for Riding The Wave

The *Riding the Wave* report calls for a collaborative data infrastructure that will enable researchers and other stakeholders from education, society and business to use, re-use and exploit research data to the maximum benefit of science and society. The "Surfboard" report is the collaborative response from the Knowledge Exchange (KE) partners; Denmark, Germany, the Netherlands and the United Kingdom. It presents an overview of the present situation with regard to research data, shows how the partners have embraced this vision, and offers broad outlines for a possible joint action programme to realise the envisaged collaborative data infrastructure.

In order for such a plan to succeed it needs the involvement of all stakeholders from the scientific community. Four key drivers have been identified:

- incentives
- training in relation to researchers in their role as data producers and users of information infrastructures
- infrastructure, and
- funding of the infrastructure in relation to further developments in data logistics.

Incentives

Four main areas of incentives to share and publish datasets are identified:

1. Re-use and recognition—inadequately catered for by current systems. Published data sets need to be included in the "bottom line" metrics (those that influence future funding) of funders. Systems and metrics need to be developed around the citation of data.
2. Principles of science, reflected in rules and codes of conduct,—codes of conduct have historically dealt with data retention rather than sharing. More recent codes are now being adopted that cover sharing and re-use, but these need to be more widely adopted.
3. Requirements by funding organisations—funder requirements in some jurisdictions are now incorporating sharing and re-use policies, and, in some cases, co-ownership policies. Policing/monitoring of policies appears to be as issue.
4. Journal data availability policies—many journals offer "supplementary" data files, but these are often limited in size/format and the data is not curated. Some Journals now mandate data availability. This is an area ripe for intervention.

Training

New skills are needed for creating, handling, manipulating, analysing, and making available large amounts of data for re-use by others. Three protagonists have been identified:

1. researchers—basic skills with regard to data handling should be a core academic competency;
2. 'data scientist'—responsible for computing facilities, storage and access in their discipline; and
3. 'data librarian'— a relatively new role responsible for data curation, preservation and archiving. Formal training programmes are few and far between.

The current situation in the four KE countries is rather diverse and very much in development.

Data infrastructure and its funding

This paper distinguishes institutional data infrastructures from disciplinary (inter)national infrastructures, describing the situation in the four KE partner countries in this respect. It is observed that in "big science" data tends to be stored in specialised, discipline oriented data centres. In contrast, "small science" practitioners tend to generate and store data locally over the lifetime of a project, after which it may or may not be stored in an institutional or journal repository. This leads to two challenges: gaps in the present data infrastructure and connectivity issues.

The funding of data infrastructure and funding models (both traditional and new) are also discussed along with reporting the results of cost benefit studies and describing the present situation of funding of data archives and data centres. The cost was previously described in other documents as "expensive" ("…research data

infrastructure is estimated to be 10-15% of the overall cost of the research infrastructure."), but this paper suggests that economies of scale and tool developments will keep costs down. Spin off benefits include *very* cost effective[3] generation of papers developed by re-using existing data.

Toward a four country action programme

Three long-term strategic goals are identified:

- Data sharing will be part of the academic culture
- Data logistics will be an integral component of academic professional life
- Data infrastructure will be sound, both operationally and financially.

This report presents the broad outlines of an action programme with mid term objectives and actions designed to achieve the long term goals above.

The plan addresses each of the four key drivers originally identified, offering strategies to resolve the problems identified in the main body of the report and highlighting areas where KE partners have already initiated applicable programmes.

[3] Figures indicating more than 60 times as many papers when compared to "original" research were quoted.

Aligning National Approaches to Digital Preservation

This 340 page volume provided some contextual background on the emergence of the digital preservation and introduced a model that might be used to identify milestones as the community progresses towards alignment. The six alignment chapters demonstrate both the need to define common goals and the benefits to be reaped by having an on-going discussion. The intent of the conference and the volume was to contribute to the next phase of community development for digital preservation and encourage international collaborations.

Each of the essays identified opportunities for alignment and provides a useful frame for community discussion. The themes fall broadly into the following categories:

- Legal
- Organizational
- Standards
- Technical
- Economic
- Education
- Gaps
- Convergence

In each of these various opportunities for alignment within the "memory community" are outlined below.

Legal

Raise awareness about legal deposit—There is resistance to legal deposit for digital material. To overcome this there is a need to identify, articulate, and disseminate case studies demonstrating the benefits and impact of legal deposit (and the impacts of not depositing) to different stakeholder groups.

Pursue cooperative agreements—Distributed preservation infrastructures are becoming the predominant models. These models need appropriate, cooperative, supportive agreements and governing laws.

Investigate collective licensing—Extended, collective, cross border licensing could help address issues of orphan works and cross-border access.

Develop a multi-national view on challenging issues—Bring together problem solvers from a broad international pool.

Organizational

Foster good practice—a more deliberate exchange of lessons learned and case studies documenting the use of emerging tools, workflows, and techniques across national and continental boundaries is required.

Encourage collaboration—Collaborations across institutions and nations should seek to extend the scope of content that is preserved by sustainable digital preservation programs.

Shift from projects to programs—Digital preservation must move from project-driven activities to being a core activity of memory organizations.

Build on strategies within nations—Align efforts intra-nationally and consider benefits of lateral collaborations at local and regional levels between nations.

Leverage interdependence between organizations—Evaluate choices about replications at all levels to inform challenging negotiations about interdependence.

Standards

Delineate interoperability standards—"Interoperability" needs to be explored and defined along the whole chain of steps that form the lifecycle of an object.

Express platform-agnostic digital preservation requirements—Concentration upon functional requirements that can be implemented in a variety of information systems allows a non-repository centric approach.

Standardize requisite skill-sets—Codes of practice that rely on clear requirements for skills and know-how should be better defined and standardised.

Engage the users of standards—The value of standards in digital preservation can best be demonstrated if we engage appropriate user communities in the discussion.

Monitor external standards—Be aware of the numerous standards pertaining to digital content that are developed outside of the digital preservation community.

Develop readiness for re-standardization—Practice will evolve and standards will need to be adapted.

Technical

Develop evaluation protocols and benchmarking—Common test data should be developed and made available to allow for evaluating and comparing technical benchmarks across solutions.

Approach interoperability rigorously—A better understanding of the value of inter-system communication and interchange is needed.

Link testing of software and implementation to economics—Economies of scale can be achieved by working cross-nationally on costly efforts like establishing operational benchmarks.

Avoid too much alignment—Multiple approaches are beneficial for technical development and the absence of diverse approaches can lead to a monoculture perspective that will be too limited.

Accept a long learning curve—Accept that there is a long learning curve and that in a hundred years the community will really know about preserving over long periods of time.

Address the need for a commodity bit storage layer—Embrace the common near-term need for a bit layer as an opportunity for alignment.

Protect against vulnerabilities of collections as examples of digital cultural heritage—Develop defences against attacks to the security and integrity of cultural heritage collections that would jeopardize content as well as confidence and trust in repositories.

Economic

Raise awareness about sustainable digital preservation—A coordinated international campaign could help to make Library/Archive/Museum (LAM) directors and administrators aware that long-term digital preservation is necessary and that it requires stable resources.

Establish a digital preservation resource centre (DPRC)—Provide LAM decision-makers with a single place for current information on digital preservation solutions.

Develop case studies—Make available case studies of digital preservation costs in order to promote better understandings of where costs accrue and where cost savings may be possible.

Define selection criteria—Develop a matrix of selection criteria for digital preservation to help build a common framework for digital content selection decisions.

Study and promote community-sourced solutions—Identify viable, community-driven business models (including cross border), and study how these models can be reapplied in other contexts.

Explore opportunities for public-private partnerships—Identify ways to cooperate by standardizing the preservation needs of public-sector institutions thus allowing private companies to compete to meet those needs.

Define core services—Identify key services, coordinate initiatives, promote common standards, implement policies and recommendations to stimulate competition among technology providers, leading to lower prices.

Support research and development—Support inter-institutional R&D across national borders to identify tools and services that yield the best return on investment.

Distinguish between investing in digitization versus digital preservation—Digitization is a short-term investment requiring long-term investment in digital preservation.

Benefit from a collective understanding of cost models—It will be easier to develop a sound understanding of cost models collectively rather than individually.

Accept instances where the sustainability strategy identifies long-term collections as public goods—A good deal of digital materials is sustained by using general public funds.

Prepare for disinvestment in cultural heritage—Work together on strategies to respond to massive disinvestment in cultural heritage globally due to economic challenges.

Balance costs of physical risk management strategies against opportunities of digitization—Recognizing the value of digitized content as a special kind of insurance for physical collections.

Education

Develop an international certificate program—Develop a common understanding of digital preservation concepts by developing an international certificate program in digital preservation.

Develop accredited curriculum, providers, and metrics—Establish a means for benchmarking and foster cooperation between international education/training providers.

Address supply and demand for qualified trainers—Develop a pool of qualified trainers capable of delivering high quality training both within and beyond the cultural memory sector.

Engage with employers and professional bodies—Use employers and professional bodies to act as reviewers for current training offerings and associated learning.

Improve cooperation in defining skill-sets—Develop a coherent way to classify education and training options to facilitate effective comparison of offerings and to enable professional development planning.

Confer about curricula internationally, then apply nationally—Many curricula and certifications are national, so translate international outcomes into national action.

Extend training to people outside the community, including the general public—Address the growing needs of people who are preserving their own content as an opportunity to raise awareness more broadly through training and as a counterpoint to rapid commercial obsolescence.

Gaps

Develop holistic strategies for data-intensive scholarship–Develop strategies that embrace the full range of the cultural and scientific record and that make policy and technical linkages to identify candidates for common and shared infrastructure.

Devise strategies to address the analogy legacy in audio-visual portions of the cultural record—Address the costs and challenges of preserving fragile and at risk audio-visual resources.

Extend strategies to new born-digital formats—Address content that is less familiar and may not be publicly accessible to determine appropriate approaches.

Convergence

Engage in outreach efforts to make the case to the public and to policy makers—Educate the public and policy makers to understand the importance of digital preservation.

Focus attention on the archival scope reflected in collecting and stewardship policies—Enable effective decisions about selection and retention that encompass new/evolving content through shared collecting & preservation strategies.

Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions
Towards better access to scientific information: Boosting the benefits of public investments in research

This Communication sets out the action that the Commission intends to take to improve access to scientific information and to boost the benefits of public investment in research focussing on the Horizon 2020 initiative. Its progenitors include: 'Digital Agenda for Europe'[4]—which covers an open data policy for publicly funded data; 'Innovation Union Communication'[5]—innovation programmes and policies; 'Communication on scientific information in the digital age'[6]; and 'Communication on ICT infrastructures for e-Science'[7].

The underlying philosophies driving this communication are:

- Wide, fair, sustainable and easy access to publicly-funded scientific information and its sustainable preservation for re-use can make a significant contribution to Europe's economic growth.
- Information already paid for by the public purse should not be paid for again each time it is accessed or used.
- Open access to underlying research data is important.

Why does it matter?

Fuller and wider access to scientific publications will help to:

- accelerate innovation (faster to market = faster growth);
- foster collaboration and avoid duplication of effort (greater efficiency);
- build on previous research results (improved quality of results);
- involve citizens and society (improved transparency of the scientific process).

Lack of speedy access to scientific information can add years to development/production cycles.

The commissions stance

Rising journal prices[8] are placing library budgets under pressure. This price increase has led to increased calls for open access publication, a stance reinforced by the increased number of funding bodies that require researchers to provide open access to publically funded research results. This requirement is mainly satisfied by the Green model (60% of the 20% of shared articles are available via the green model). Open access doesn't affect author choice regarding publishing (or not), patenting and commercial exploitation.

Sharing data openly is not as far down the line. There aren't well established sharing practices—only 25% of researchers share their data openly. Much of the data is locked away, an inefficient use of research funds. Some funders have also started to require the deposit of data in open repositories but the practice is not yet widespread.

Preservation has been shown to be cost effective (with a fourfold return on investment), but there are potential problems regarding the preservation of the software and hardware needed to keep it accessible.

[4] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0245:FIN:EN:PDF

[5] http://ec.europa.eu/research/innovation-union/pdf/innovation-union-communication_en.pdf

[6] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2007:0056:FIN:EN:PDF

[7] http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0108:FIN:EN:PDF

[8] Increasing by 3.5% above inflation per year (this is partially explained by the increased number of articles published)

Barriers

Overall the key issues are the level of investment lack of co-ordination between states.

For scientific publications the problem is one of a too rapid movement to open access potentially leading to a destabilisation of the scientific publication sector which, until recently, has essentially *been* the scientific information system. The publication process costs money and this needs to be taken into account (by using the gold pay to publish model for instance). However, pay to publish must lead to a reduction is subscription costs. When it comes to complying with open publication mandates there are also issues of education–knowledge of how to go about self-archiving, lack of infrastructure—nowhere to put the publication, contractual problems, and lack of enforcement.

E-infrastructures for access to data are emerging, but many lack long term financing models. Interoperability is also a problem. Researchers appear to be reluctant to share 'their 'data, and because there is no immediate benefit (reputational or financial) they are reluctant to expend the time and money.

Preservation is perhaps the most affected by financial issues. Preservation support structures are created for particular projects and In many cases funding is so short lived the data doesn't survive beyond the lifetime of that project. Legal requirements regarding data depositing are being adapted to cove digital data in many states but these are inconsistent.

European level actions

So far the commission has developed policy, but progress across states has been uneven. Mandated open access to research funded by the commission was introduced as a pilot scheme in 2008 as part of the FP7 programme. In addition the commission has supported the development of e-infrastructures for science[9].

Next steps include:

- Mandated open access (green or gold) for publications generated as part of Horizon 2020
- Encouraging authors to retain their copyright
- Pilot schemes on open access to and reuse of research data
- Guidance for researchers and institutions
- An e-infrastructure for the commissions own publications
- Continued dialogue with stakeholder groups
- Funding infrastructures and data preservation projects
- Co-ordination beyond the EU

[9] For example OpenAIRE http://www.openaire.eu/

Commission recommendation of 17.7.2012
On access to and preservation of scientific information

Having set out it's stall in the July communication paper (above) this paper goes on to make firm recommendations for member states to adopt. Key principles/information covered in the introduction to the recommendation are as follows:

- A knowledge based economy is a priority
- All publicly funded research (papers and data) should be widely disseminated through open access. This will reduce duplication of effort and speed up scientific progress.
- Open access —free access as early as possible, use and reuse—should be implemented taking into account intellectual property rights. This should be complemented with robust preservation strategies.
- Businesses—in particular small/medium sized enterprises (SME)—will benefit
- Research and funding policies need to reflect the new linked/open world
- Preservation strategies are under strain because of the sheer volume of data now being produced.
- Policies should be defined at national or sub-national level
- Robust, fit for purpose, inter connected e-infrastructures will improve access to scientific information.
- Publishing is in transition. Suitable new solutions for the scientific process need to be established

Recommendations to member states
(paraphrased)

Open access to scientific publications

1. Define clear policies for the dissemination of and open access to scientific publications resulting from publicly funded research including:

- concrete objectives and milestones;
- implementation plans;
- associated financial planning.

Ensure that, as a result of these policies:

- there is open access to publications immediately, as soon as possible, or no later than six months after the date of publication, (twelve months for social sciences and humanities);
- licensing systems contribute to the open access to scientific publications without prejudice to the applicable copyright legislation; researchers should be encouraged to retain their copyright;
- academic career systems support and rewards researchers who share the results of their research, through open access by developing and using new, alternative models of career assessment;
- transparency is improved, in particular in relation to agreements between public and publishers;
- small and medium-sized enterprises and unaffiliated researchers have the widest and cheapest possible access to scientific publications.

2. Ensure that research funding institutions responsible for managing public research funding and academic institutions receiving public funding implement the policies by:

- defining institutional policies for the dissemination of and open access to scientific publications; establishing implementation plans at the level of those funding institutions;
- making the necessary funding available for dissemination (including open access and other alternative channels);
- adjusting the researcher recruitment and career evaluation system and the evaluation system for awarding research grants to researchers to benefit those who share results and data;
- giving guidance (particularly relating to intellectual property rights) to researchers on how to comply with open access policies;
- conducting joint negotiations with publishers to obtain the best possible terms for access to publications, including use and re-use;
- ensuring that results of research that receives public funding are easily identifiable by appropriate technical means.

Open access to research data

3. Define clear policies for the dissemination of and open access to research data resulting from publicly funded research. These policies should provide for:

- concrete objectives and milestones;
- – implementation plans;
- – associated financial planning.

Ensure that, as a result of these policies:

- research data is publicly accessible, usable and re-usable through digital e-infrastructures taking into account concerns in relation to privacy, trade secrets, national security, legitimate commercial interests and intellectual property;
- datasets are easily identifiable and can be linked to other datasets,
- additional information is provided to enable their proper evaluation and use;
- public research funding institutions and publicly funded academic institutions put in place mechanisms enabling and rewarding the sharing of research data;
- advanced-degree programmes of new professional profiles in the area of data handling technologies are promoted and/or implemented.

Preservation and re-use of scientific information

4. Reinforce the preservation of scientific information, by:

- defining and implementing policies, and responsibilities for the preservation of scientific information, along with associated financial planning, in order to ensure curation and long-term preservation of research results and data;
- ensuring that an effective system of deposit for electronic scientific information is in place, covering born-digital publications and, where relevant, the related datasets;
- preserving the hardware and software needed to read the information in future, or by migrating the information to new software and hardware environments on a regular basis;
- fostering the conditions for stakeholders to offer value-added services based on the re-use of scientific information.

E-infrastructures

5. Further develop e-infrastructures underpinning the system for disseminating scientific information by:

- Supporting scientific data infrastructures for dissemination of knowledge with a common look and feel to data discovery across disciplines;
- supporting the development and training of new cohorts of data-intensive computational science experts;
- building on existing resources to be economically efficient and to innovate in the areas of analysis tools, visualisations, decision-making support, models and modelling tools, simulations, new algorithms and scientific software;
- reinforcing the infrastructure for access to and preservation of scientific information at national level, and earmarking the necessary funds;
- ensuring the quality and reliability of the infrastructure;
- ensuring interoperability among e-infrastructures at national and global level.

6. Ensure synergies among national e-infrastructures at European and global level by:

- contributing to the interoperability of e-infrastructures;
- supporting transnational cooperative efforts that promote the use and development of information and communication technologies infrastructure for higher education and research.

Multi-stakeholder dialogue at national, European and international level

7. Participate in multi-stakeholder dialogues at national, European and/or international level on how to foster open access to and preservation of scientific information.

Participants should in particular look at:

- ways of linking publications to the underlying data;
- ways of improving access and keeping costs under controls;
- new research indicators and bibliometrics encompassing not only scientific publications but also datasets and other types of output from research activity and the individual researcher's performance;
- new reward systems and structures;
- the promotion of open access principles and implementation at international level.

Structured coordination of Member States at EU level and follow-up to the Recommendation

8. Designate by the end of the year a national point of reference whose tasks will be:

- coordinating the measures listed in the Recommendation;
- acting as an interlocutor with the European Commission on questions pertaining to access to and preservation of scientific information;
- reporting on the follow-up to the Recommendation.

Reviewing and reporting

9. Inform the Commission 18 months from the publication of this Recommendation, and every two years thereafter, of action taken in response to the different elements of this Recommendation.

Sustainability of Open Access Services

Open Access appears to be vulnerable with many (although not all) services running on recurrent project funding or goodwill. The Knowledge Exchange (KE) Open Access Working group commissioned this study as part of a larger programme of work to look at the issue of sustaining key services into the long term. This document reports on Phase One—a scoping exercise of what is currently freely available, what should be available and roles—and Phase Two—stakeholder consultation and engagement—of that study.

The main use cases are developed in terms of the services used by the research-related community and the key free-to-use services are described as a series of value curves depicting the current and future value of each service type. For each use case scenarios are used to map; Actors, Workflow, Components, Current need, Future need, and services available.

The use cases are:

- Establish repository
- Run repository
- Publish research findings
- Read and use research findings
- Build and manage academic profile
- Manage research (institution)
- Manage research (funder)
- Publish Open Access journals
- Promote Open Access

There are some overlaps or duplications in service provision—not only between commercial services, but also between free-to-use services—as well as complementarities and gaps (also mapped). In the context of a set of plausible Open Access scenarios for the future (and predicated on the premise that encouraging developments already taking place is an optimal strategy) the key service basis for each of these scenarios are described as follows:

(i) 'Gold' Open Access[10]
- OA journal publishers
- OA journal publishing technologies and platforms
- APC payment mechanisms/agencies
- Preservation services
- QC and standards
- Search services
- Advocacy

(ii) Fully 'Green' Open Access[11]
- Repositories
- Quality control service (peer review management, editorial work) payment mechanisms/agencies
- Preservation services
- QC and standards
- Search services
- Advocacy

(iii) 'Green' Open Access supplementing subscription access as 'Gold' OA grows (the current scenario)
- OA journal publishers
- OA journal publishing technologies and platforms
- APC payment mechanisms/agencies
- Preservation services
- QC and standards

[10] Gold Open Access—authors publish in an open access (OA) journal that provides immediate OA to all of its articles on the publisher's website, in some cases this involves a fee paid to the publisher.

[11] Green Open Access—authors publish in any journal and then self-archive a version of the article for free public use in an OA repository/website

- Search services
- Repositories and deposit services
- Usage and impact services

The roles different stakeholder groups—libraries, research and infrastructure funders, and service providers—play in future scenarios are discussed. Libraries are already contributing directly (financial donations and in kind) to sustaining some key services such as arXiv[12], the Directory of Open Access Journals[13] and Bielefeld Academic Search Engine[14]. Funders in general acknowledge they have a role to play in helping to sustain key components of the system. Some universities have committed funds to pay directly for Open Access publication charges.

Three strategic areas are identified as having particular potential for future work. These are: embedding business development expertise into service development; consideration of how to move money around the system to enable Open Access to be achieved optimally; and governance and coordination of the infrastructural foundation of Open Access.

A number of recommendations for further work are made—primarily addressed to infrastructure funders—around these strategic areas. Issues arise here: coordination between funders, between projects and between funders and projects; emergence –the unexpected emergence of services from projects that were not expected to produce one; and the business development aspects of service building.

Recommendations:

1. Develop best practice guidelines for new projects/services that all parties can sign up to—a set of practical, focused and relevant questions of the type that would underpin a normal business planning process.
2. Scope the development of a system of formal business appraisal for services in development.
3. Develop guidelines for service providers to use in monitoring the business aspects of their services throughout the development phase and beyond
4. Establish systems for business monitoring and evaluation of projects—review and revise plans for the nascent services.
5. Provide information for research funders on sustainability issues
6. Address the gaps and redundancies
7. Scope and design a coherent governance system—bringing interested parties together to develop a shared vision of the overall Open Access infrastructure and to gain their on-going support and commitment

[12] http://arxiv.org/
[13] http://www.doaj.org/
[14] http://www.base-search.net/about/en/index.php

Bibliography

High Level Expert Group on Scientific Data, “Riding the wave, How Europe can gain from the rising tide of scientific data” [http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf] Pub: European Union, Oct 2010

Maurits van der Graaf, Leo Waaijers “A Surfboard For Riding The Wave: Towards A Four Country Action Programme On Research Data” [http://www.knowledge-exchange.info/default.aspx?id=469] Pub: Knowledge Exchange, Nov2011

Nancy Y. McGovern (ed), Katherine Skinner (ed) “Aligning National Approaches to Digital Preservation” [http://www.educopia.org/sites/default/files/ANADP_Educopia_2012.pdf] Pub: Educopia Institute, 2012

European Commission “Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions: Towards better access to scientific information: Boosting the benefits of public investments in research” [http://ec.europa.eu/research/science-society/document_library/pdf_06/era-communication-towards-better-access-to-scientific-information_en.pdf] Pub: European Commission July 2012

European Commission “Commission recommendation of 17.7.2012on access to and preservation of scientific information” [http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf] Pub: European Commission July 2012

Alma Swan “Sustainability Of Open Access Services” [http://www.knowledge-exchange.info/Files/Filer/downloads/Open%20Access/SustainabilityServices/Sustainability_OA_services_12.pdf] Pub: Knowledge Exchange Sep 2012