Jisc

June 2014

Bringing the UK's open access research outputs together

Barriers on the Berlin road to open access

Neil Jacobs (Jisc) and
Nicky Ferguson (Clax Ltd)

Contents

Introduction

In 2003, the Berlin Declaration on Open Access (OA) proposed that research outputs should be made openly available for use and reuse - with appropriate attribution - via repositories using established technical standards. In 2013 Jisc, Research Libraries UK (RLUK) and the Society for College, National and University Libraries (SCONUL) undertook a feasibility study into the development of an "open mirror", which would bring together the UK's open access research outputs and so make them easier to use and reuse.

The work is described here, and it has identified significant barriers hampering the creation of the open mirror. This report considers why, ten years after the Berlin Declaration, and with significant amounts of the UK's research output being (at least nominally) open access, it is still so difficult to build an open mirror. Finally, it recommends work to make the process easier.

Context

Like many countries, the UK is moving towards open access for the publications of its researchers, for a variety of reasons and driven by factors rehearsed extensively elsewhere. It seems natural, therefore, to be able to see (and look after) this growing corpus of open UK publications in a single place, perhaps one that is within the ambit of the academic community that produces those publications. That idea was the starting point for the Open Mirror feasibility project, which ran from June 2013 to February 2014. This document summarises where we are at the end of the project, based on extensive consultation, horizon scanning, technical prototyping, legal review and a dedicated stakeholder workshop that was held in January 2014.

What is an Open Mirror?

As originally conceived, the Open Mirror would be an aggregation of OA content, building upon the network of institutional repositories in the UK. It would cover all UK OA publications[1], both "green" and "gold", and might therefore require significant changes in the interoperation between institutional repositories and other initiatives such as publisher platforms and subject-specific repositories.

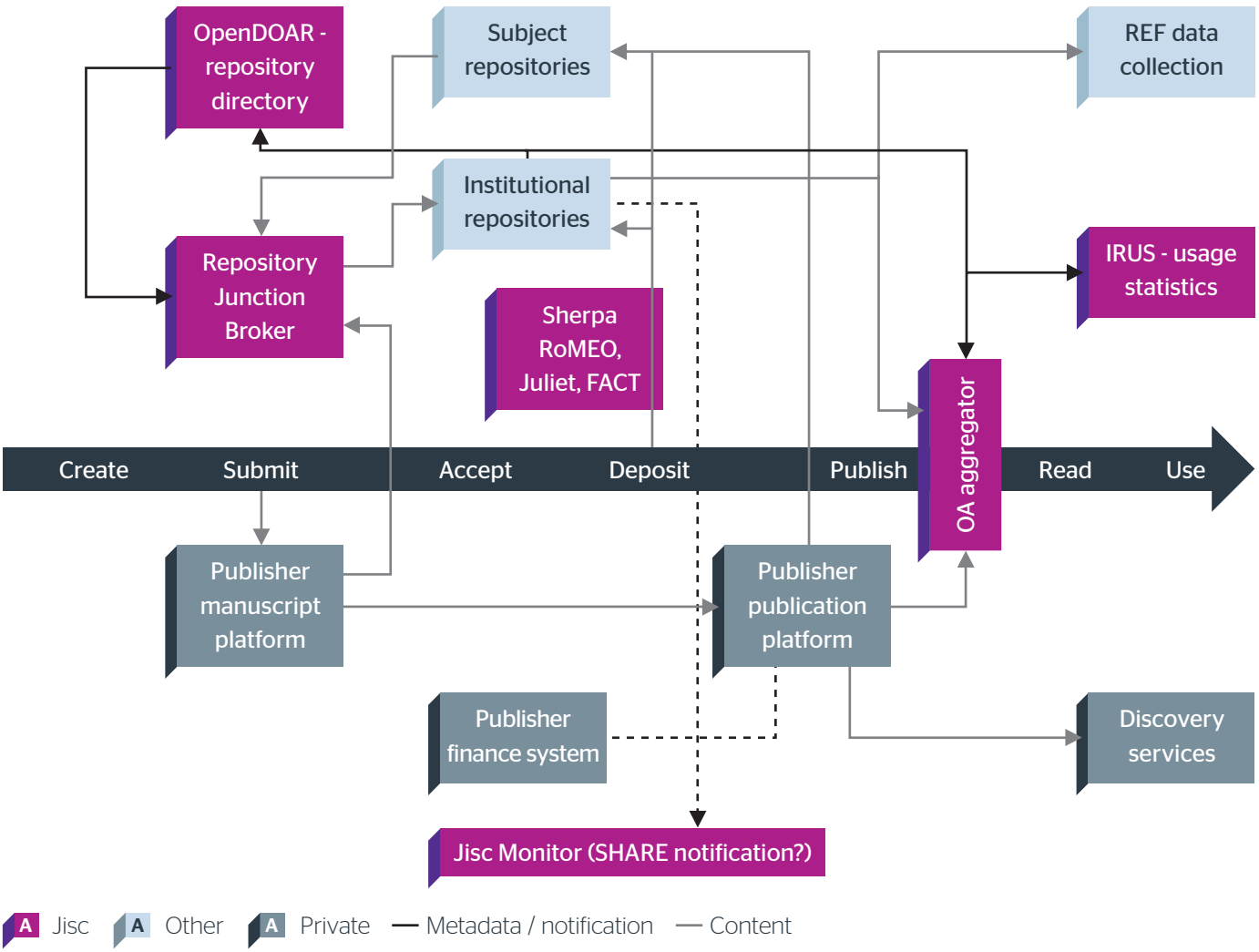
Environment

Services similar to the Open Mirror already exist internationally – for example, the Clearinghouse for the Open Research of the United States (CHORUS) is a publisher initiative that is based on existing infrastructure and arrangements such as publisher platforms CrossRef and Portico. Its proponents argue that, by restricting published papers to journal websites while making them readable, it both preserves trust in the scholarly record and expands access to it. Also in the US the Shared Access Research Ecosystem (SHARE) is a higher education-based initiative to strengthen efforts to identify, discover, and track research outputs. In some ways a rival and in some ways complementary to CHORUS, it is based on institutional and other academic infrastructure, such as repositories. There are several initiatives under way in European nations, including the National Academic Research and Collaborations Information System (NARCIS), a national aggregation of OA publications in the Netherlands. One of its roles is to provide the European OpenAIRE system with Dutch OA material in an efficient way.

In the UK there has been considerable investment in repository and related infrastructure, and the schematic in Figure 1 presents this, related services, and how these fit in a typical scholarly communications workflow, from authoring to using research findings. The services, content and metadata flows shown in the diagram are illustrative; not every research paper is represented by all of them. Nevertheless, the diagram suggests both that the skeleton of an infrastructure is in place and that more work is needed to streamline the content and metadata flows. Might Open Mirror be a part of that work?

[1] Our working definition is: any Open Access publication with an author based in the UK.

Figure 1: The UK environment



Identifier / registry services

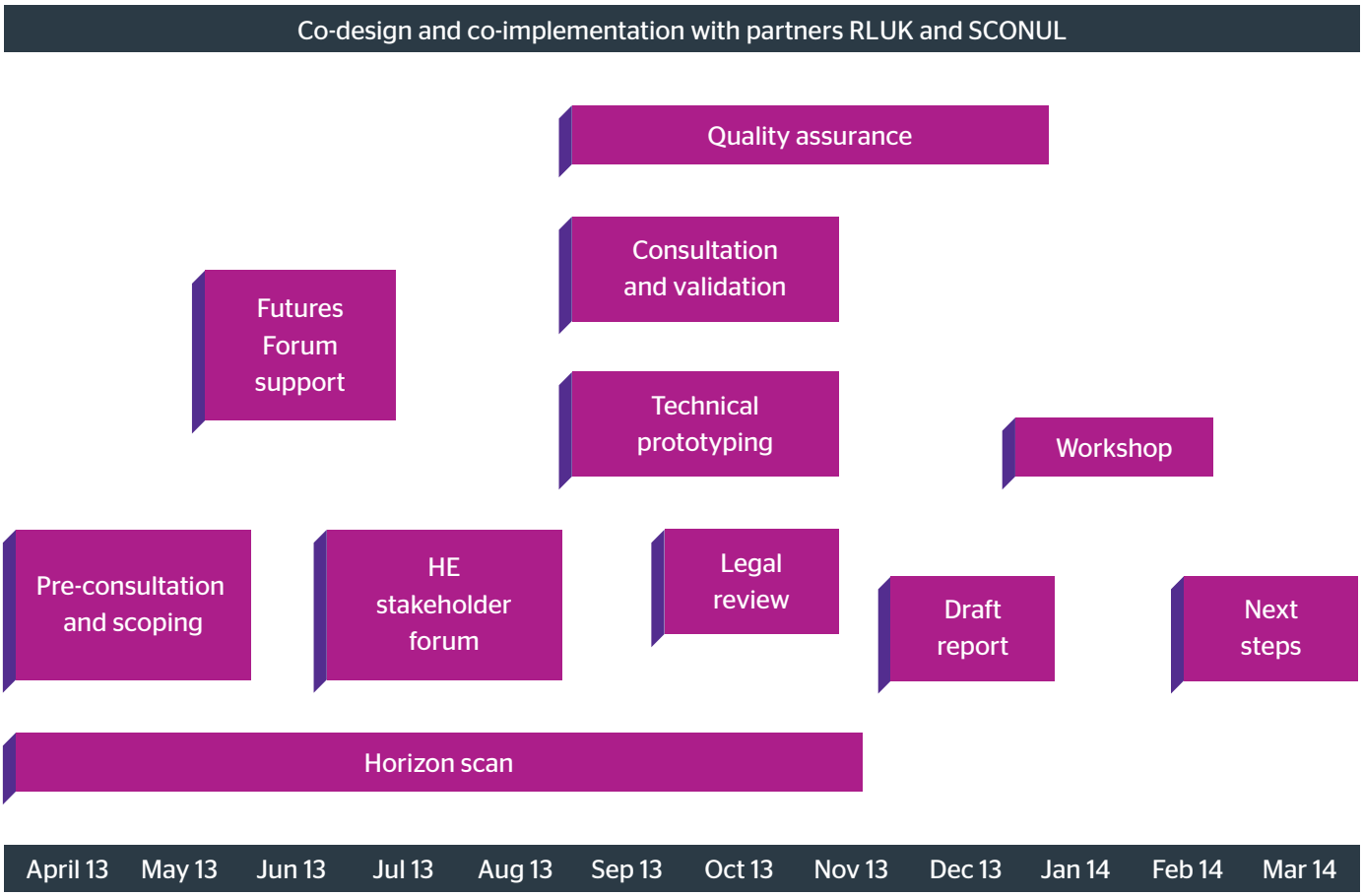
Ringgold ISSN crossref fundref DOAJ ORCID

A feasibility study

Jisc was asked by RLUK and SCONUL as a part of its first "co-design" programme to investigate the value and feasibility of the Open Mirror. The report from that study for the final stakeholder workshop is available as an appendix to this paper, together with the various report annexes.

The broad shape of the study is shown in Figure 2.

Figure 2: The feasibility study - Method



As will be clear from Figure 2, the study consulted widely and developed a robust evidence base from which to draw conclusions. Even so, consultation and partnership building work remain to be done.

Current high level use cases

Innovation tends to be iterative, identifying problems to solve and opportunities to exploit in parallel with development to see what is possible and interesting. The UK's repository infrastructure has been built in this way and a range of high level use cases have crystallised as being core to its existence.

These use cases are also highly relevant to the Open Mirror. They are, in no particular order, to:

- Report on research, e.g. to Research Councils and the Research Excellence Framework (REF)
- Preserve research outputs, either to ensure business continuity or in the long term
- Ensure the visibility and searchability of UK OA content
- Showcase institutions' research assets
- Increase research impact
- Justify UK expenditure on research for government and citizens
- Provide management information for funders, HEIs, government, etc. (e.g. for planning, benchmarking)
- Enable the development of new software applications and services relying on UK OA content, by providing harmonised, programmable access to it
- Enable more accurate and evidence-based analysis of OA content, such as for the purposes of assessing the state of the "mixed economy" as posited in the Finch Report
- Increase the quality of the metadata and the quantity of full text (both formal publications and so-called "grey literature") in institutional repositories (IRs)
- Enhance the persistence and accuracy of access to resources held in IRs through the consistent use of identifiers (for example, for outputs and for authors)

The challenges faced by different stakeholders in meeting the use cases

Unsurprisingly, the high level use cases noted on page 5 were strongly represented in the feedback in the feasibility study. Many of those consulted noted considerable challenges in meeting those use cases.

Managers and staff of individual IRs (over 150 IRs at the last count) face a number of challenges:

- Curating and maintaining the quality of their metadata for a rapidly expanding and difficult to track down corpus of papers, publications and other literature
- Correctly identifying and interacting with authors to obtain the initial metadata and - particularly difficult - the full text
- Improving the visibility of content within IRs
- Interoperating with funders, subject repositories and journal publishers
- Preservation

Research managers need to:

- Track existing collaborations and identify potential new ones
- Identify potential research collaborators and funders, benchmark and measure research performance

Funders find that:

- The poor coverage of some IRs makes interacting with institutions through their IRs patchy and difficult
- Text mining such a diverse range of sources is highly problematic technically and legally
- There is a need for international coordination over standards, initiatives and funding
- There is an urgent need for UK HE to be able to meet monitoring and compliance requirements for RCUK and Wellcome and to manage efficiently the research outputs which HEFCE will require to be available via repositories for the next REF

Publishers see IR contents as unreliable and poor quality. Conversely, improving the quality through use of subject repositories or aggregations is seen as a threat to their subscription income.

Commercial intermediary services would welcome aggregation activity as they find that the increase in green OA repository sources makes access and discovery more difficult. They are also concerned about preservation.

Researchers and research users are frustrated by:

- Barriers to text mining, especially licensing
- The lack of visibility and impact of some IRs
- The absence of a jargon-free route to access research and particularly the researchers who produce it

The challenges expressed above give us a shopping list of possible tasks, the accomplishment of which would assist the UK repositories infrastructure in meeting the high level use cases and becoming more interoperable, in line with the Finch Report:

- Increase access to all open access material, whether green or gold, self-archived or publisher-archived
- Improve the quality and exposure of repository-held content
- Aggregate metadata for open access content
- Recognise the value and importance of text mining and the need to aggregate full text content for this purpose, where legally possible
- Improve exposure and visibility on established resource discovery services
- Ensure preservation of OA content in line with the Finch report recommendation 7.6
- Encourage use of identifiers and develop systems of guaranteed persistent access so that OA content and associated citations persist over time and reduce web link "rot"
- Enable UK higher education (HE) to manage efficiently the research outputs to be available via institutional repositories for REF post-2014

A full table of problems/requirements, benefits, risks and possible solutions appears as an appendix. It is unlikely that a single programme of work would address all of these in a realistic way. We are therefore at a crossroads, where each path would address only some of the tasks outlined above.

Finch report
researchinfonet.org/publish/finch

The barriers and how to overcome them

The barriers and how to overcome them

The last phase of the work was to subject the evidence from the feasibility report to review by the main stakeholders, including publishers, librarians, research funders, repository managers and experts in scholarly communication[2]. One outcome has been to identify more clearly the main barriers on the road to full OA, as described in the Berlin Declaration. These are outlined in the table below, in no particular order.

Table 1: The barriers on the Berlin road, and some ways to get over or around them

#	Barrier	Implication	Possible solutions
1	Metadata in repositories is incomplete and/or inaccurate.	Management information or resource discovery use cases are not well supported - research is hard to find and hard to count.	» Well-defined and widely implemented metadata profile[3].
2	Repositories do not always hold, or even point to, an OA version of the full text.	Research users cannot read or reuse the outcomes from university research.	» Repositories should have an incentive to address this from the Higher Education Funding Council for England (Funding Councils) REF policy. » Support initiatives such as Repository Junction Broker[4], and repository use of the CrossRef API.
3	The licence conditions for OA content are unclear.	Research users do not have confidence to use some OA material, especially from repositories.	» Work with funders, publishers and libraries to agree a licence for green OA material. » Encourage funders to monitor compliance with this aspect of their policy. » Support tools such as "how open is it" that surface these issues.
4	The licence conditions for OA content are too restrictive.	Research users and curators cannot do their job.	» Encourage universities to support authors in retaining the rights they need.
5	It is technically difficult to extract OA content from some platforms.	OA material, e.g. from some repositories and publishers, is not as widely visible and re-used as it might be.	» Ensure repositories and publishers support propagation of OA content by meeting their concerns (e.g. about loss of usage data). » Improve technical interfaces, perhaps building on the new ResourceSync protocol.
6	Poor visibility of existing material in IRs.	IRs seen as poor investment.	» Shared services support IRs to optimise their visibility in discovery services, including search engines, and associated troubleshooting.
7	The number and variety of non-interoperating systems recording research publications, leading to duplication of effort, inconsistent metadata, etc.	The scholarly record is fragmented and unreliable. IRs are seen as an inefficient solution to (inter)national requirements.	» Increase interoperation between IRs, and also with subject repositories and publishers and CrossRef. Repository Junction Broker, possible CORE and associated work and repository support services may all have roles to play.
8	Lack of easy machine access to UK OA materials.	Potential lost for innovative services to be created by other players.	» Ensure API access to any publicly funded aggregation efforts.

[2] Research managers were also involved in the review, but unfortunately could not attend the workshop.
[3] The RIOXX profile is a simple schema designed to meet funder requirements, and will have a supported roll-out in the UK through 2014-15.
[4] The Repository Junction Broker is now called "Publications Router".

Which path to take next?

This was a complicated study, with many cross-cutting opinions, constraints and options. A summary options appraisal appears as an appendix. Though none of them are universally supported, the following broad approaches emerged from the analysis and are discussed in the report appendices.

a. **Don't build anything new, at least for the moment. Map the current provision first**
Simply undertaking further consultation and consensus-building has the value of laying a secure basis for service development, but risks protracted discussion without changing the facts on the ground when, arguably, these need to be changed over the next couple of years in response to strong policy drivers. A mapping exercise matching requirements against national and international projects and services is an essential groundwork and will be a prerequisite for any near-term decisions.

b. **Focus on metadata, provide full-text wherever legally possible**
Useful services can be built based on metadata records alone, but they are not as useful as those based on metadata records plus the full text. Nevertheless, noticeable improvements to resource discovery and research management reports could be expected from a service that aggregated and improved the metadata held by universities about their research outputs. The risk profile here is relatively low.

c. **Collect everything, enable potential services for a range of users**
An open collection of all the UK's research outputs could support services for academics, students, librarians, funders, developers and others including those working outside universities. These could be different services, created by others on the basis of the collection, for example a specific or comprehensive portal, a preservation service, sophisticated management reports, or text-mining services. There will be significant challenges as well as notable legal risks given the complexity and lack of clarity around rights to many research outputs, but collection and experimental pilots could start now.

d. **Support the existing repositories, look to the future**
Many of the barriers to an Open Mirror can be reduced by enabling the UK's institutional repositories to operate and interoperate more effectively and efficiently, and by better use of standards and services that already exist, or are in late development. One challenge here would be the variety of systems, including both repositories and more sophisticated research information systems, being deployed by universities. However, national policy drivers might provide a business case for some harmonisation. We should also be aware of the high costs to the sector of running more than 150 separate IRs and look to a future where a national service might supply the IRs with expertise, records and routes to compliance.

Following further reflection by Jisc, RLUK and SCONUL, we recommend several complementary ways forward which will encompass these approaches. We recognise that, whatever path is agreed, the activity will need to operate in a diverse and dynamic ecosystem of services in scholarly communication and research information management.

Conclusions and recommendations

Conclusions

The feasibility study has found a range of views on the Open Mirror, from stakeholders and experts. There is a sense from the UK higher education community and others that action is needed to improve several aspects of scholarly communication and research information management, and there is some support for a number of initiatives that might form part of an "Open Mirror" as a contribution to that improvement.

It seems clear that some well-defined actions could be valuable in this area, and the summary requirements table outlines the problems faced by a range of stakeholders that might be alleviated by such actions. Barriers and possible solutions are summarised in Table 1 on page 8

It is absolutely essential that any action is well integrated into other local, national and international activities[5] which it could complement, to benefit UKHE. We recommend below a mapping exercise to ensure any new initiatives fit with current activities. At least two issues arise from this:

- While such a mapping exercise was beyond the scope of this feasibility study, it seems clear that two initiatives, Connecting Repositories (CORE) and Repository Junction Broker, have the potential to be key to any potential aggregation activity. The latter has Jisc funding for two years, and its development plan should be influenced by this study and any further scoping work. CORE does not have funding in place beyond March 2014 and so is specifically cited below
- Many elements of the infrastructure operate on a commercial basis. Open Mirror should not compete with these but, instead, either use them to provide valuable services to UKHE that are not otherwise delivered, or enable commercial services to offer better value to UKHE

No single path or combination of paths is likely to meet everyone's requirements, or overcome all the barriers – nor does it need to do so. Nevertheless, it is possible to identify several complementary ways forward from the evidence above.

Recommendations

1. One of our interviewees noted that it was important to see the Open Mirror as part of "an overall repository and scholarly infrastructure for the UK". A systematic mapping exercise and review of the potential of elements of this infrastructure, national and international, should be undertaken before further development work begins. In considering the options and ways forward, consideration should be given to timescales, perhaps best indicated by "feasibility" in the options appraisal, to ensure that any further work balances a need for quick wins, developing a sound basis for any high-profile service and the need to ensure that any such service is not superseded by events.
2. In the near term, it might be necessary for Jisc to carry the full costs of CORE. However, Jisc should actively seek international support for something like CORE and in the meantime CORE should focus on:

    a. Aggregating materials from UK IRs and from publishers and subject repositories of outputs with UK-based authors to ensure that UK resources are well represented in CORE; this should be accompanied by an active programme of support for UK repositories to enable them to participate (see 5 on page 11).
    b. A number of user-focused, small pilot projects should be planned or commissioned to demonstrate to end users[6] the potential benefits of different uses of the aggregation. Possible topics for the pilot might include:

        i. Populating IRs with consistent quality data
        ii. Providing material suitably licensed for text mining
        iii. Improving or providing metadata using automation

3. Jisc should consider developing a managed consultation, partnership and scoping phase for further aggregation effort, to include a cost benefit analysis of the comparative costs, benefits and risks of a shared service to support HE institutions (HEIs), particularly with some of the functions currently duplicated across more than 150 IRs. This consultation should aim to develop a consensus vision of the potential of an aggregation to supplement other existing services, while recognising that a variety of actors may build new and innovative services from ("on top of") this aggregation.

4. Metadata standards and formats were mentioned throughout this study. It should be possible to tell immediately from standard metadata if an output is OA and if so which licence is attached, what the embargo period is (if any), and also who funded the effort on which this output is based. This is one particular activity which Jisc should raise with international partners and aim to address in collaborative activities, for example by encouraging and enabling the adoption of both the National Information Standards Organisation (NISO) and the V4OA proposals on this topic and the workflows that will enable the relevant metadata fields to be populated accurately.
5. Jisc has supported the Repositories Support Project over a number of years and also some specific additional work around standards and interoperability. It is highly desirable that this kind of technical support be provided to some degree during 2014-15 as UK research funders' OA policies are implemented in HEIs, in part using their repositories. Such support should be fully aware of, and integrated with, related work such as that outlined above and existing services such as the Sharing Hybrid Environment for Research Preservation and Access (SHERPA) suite, also bearing in mind existing projects such as Repository Junction Broker and CORE. This support initiative should focus on and encourage:

    a. More widespread use of consistent identifiers
    b. Improved and consistent practice in search engine optimisation (SEO)
    c. More consistency of institutional mandates for OA and better awareness and facilitated monitoring of funder policy compliance

6. If undertaken alongside other service options, this would complement those options by raising awareness and addressing issues such as trust and duplication of effort.

Given the scale of the task, much has been achieved in the ten years since the Berlin Declaration on Open Access. It is clear that shared service infrastructure has been developed and that content is being aggregated. We will need to demonstrate value by not duplicating services available elsewhere, but by building on them instead, and helping to improve their interoperation and usability and by extending their potential use to all parts of the UK community and economy. Consultation will be a key aspect of any further work.

[5] Key players and initiatives are mentioned in the report but it is worth mentioning Gateway to Research, the British Library, particularly with their expertise in preservation and aggregation activities, CrossRef, Google Scholar and commercial CRIS systems.

[6] Users in this context may be library professionals, repository and research managers and staff from funders as well as researchers.

Summary of footnotes

1. Our working definition is: any Open Access publication with an author based in the UK. (Page 3)
2. Research managers were also involved in the review, but unfortunately could not attend the workshop. (Page 8)
3. The RIOXX profile is a simple schema designed to meet funder requirements, and will have a supported roll-out in the UK through 2014-15. (Page 8)
4. The Repository Junction Broker is now called "Publications Router". (Page 8)
5. Key players and initiatives are mentioned in the report but it is worth mentioning Gateway to Research, the British Library, particularly with their expertise in preservation and aggregation activities, CrossRef, Google Scholar and commercial CRIS systems. (Page 10)
6. Users in this context may be library professionals, repository and research managers and staff from funders as well as researchers. (Page 10)

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