Digital Repositories Roadmap Review: towards a vision for research and learning in 2013

Document details

Author:	Rachel Heery
Date:	2009-05-04
Version:	4
Notes:	Final version revised

Acknowledgements

Many thanks to all those who contributed to the review by completing the email questionnaire, participating in the IdeaScale Web forum, or attending the workshop. The author takes responsibility for interpreting input and for any change of emphasis that comes with collating the viewpoints of the various contributors.

Contents

1 Executive Summary

The Repositories Roadmap was published in 2006 and presented a vision for the role of repositories in 2010. Given the many changes in practice, policy and technology since 2006 it is timely to review the Roadmap. This Review seeks to extend the horizon to 2013, to clarify the relationship of repositories to the broader environment and to steer the future work of JISC and others interested in furthering repository implementation and enhanced scholarly communication. The review is based on consultation with experts using a Web forum (IdeaScale), a questionnaire based survey, and a workshop.

The Review is structured into two parts. Firstly it makes a number of recommendations targeted at the JISC Executive to inform further funding of repository related activity. The Review then goes on to identify a number of milestones of relevance to the wider community that might act as a measure of progress towards the wider vision of enhanced scholarly communication. Achievement of these milestones would be assisted by JISC through its community work and funding programmes.

Recommendations to JISC

Recommendation 1: Chart progress of the implementation and usage of repositories by measurable indicators, for example

- Measure cultural and organisational change by whether relevant repository related deliverables are included in institutional information strategies (in particular open access, integration of institutional systems, digital preservation and curation, well managed digital collections).

- Produce baseline analysis of current repository content and analysis of potential repository content.

- Enhance repository statistics to measure the availability of open access full text items (e.g. using ROAR or OpenDOAR or Intute).

- Analyse patterns of deposit in those institutions with an institutional mandate compared to other institutions. Who is self-depositing and why?

- Scope metrics, qualitative as well as quantitative, to demonstrate the value of repositories to institutions and their members.

Recommendation 2: Analyse current communication behaviour of researchers and teachers, and involve them in development of future scholarly communication services.

Recommendation 3: Given the environmental changes in the education, Web and scholarly environment we need to articulate an updated vision of richer scholarly communication. The vision should be based on the scholarly life-cycle from experiment through to research, teaching and learning. The various roles repositories might play should be mapped out, particularly as regards management of digital resources, open access and re-use.

Recommendation 4: Communications about repositories should emphasise higher level objectives. It would be helpful for JISC to differentiate high level objectives and subsume repository activity under those objectives. There needs to be a shift in emphasis from the 'repository' to the objective. JISC should ensure calls and funded activity relate to particular objectives rather than to 'repositories'.

Recommendation 5: Tools and project outputs should be presented in terms of solutions to local institutional repository requirements.

Recommendation 6: Repository activity needs to be joined up more closely with other forward looking work such as the JISC Virtual Research Environment activity, open science, data sharing and preservation. At the same time at the local level there needs to be support for integration of repositories with institutional systems (Current Research Information Systems, Research Excellence Framework, Virtual Learning Environment, author identity systems).

Recommendation 7: Harness existing JISC funded expertise to take the lead on exploiting the Web for teaching and research. Ensure existing JISC services have responsibility to advise and demonstrate use of new Web developments and services for scholarly communication to include aspects where repositories have a role. This may involve UKOLN, CETIS and/or the Repositories Support Project. Extend the remit of CRIG (or a similar group) to address interfaces with Web based systems and Web 2.0 initiatives.

Recommendation 8: Ensure clear classification of the outputs of JISC funded activity into categories useful to repository managers such as relevance in short term or long term; ease of implementation; relevance to different repository types.

Recommendation 9: Explore options for moving to a more Web based architecture for repositories, taking into account the requirement to move forward existing 'legacy repositories'.

Recommendation 10: Explore the concentration of the collection of different types of content at different levels - small group (lab, research group, student cohorts), discipline and global levels - and how these different levels might facilitate social networking effects.

Recommendation 11: Explore deployment of a cut down version of SWAP, possibly at the copy level, retaining the cataloguing rules to ensure a consistent approach to linking to full text. Evaluate whether use of SWAP is consistent with a Web architecture approach to repositories.

Recommendation 12: Explore use of OAI-ORE to enable applications to handle complex objects. Demonstrate how OAI-ORE facilitates the re-use of research outputs and research data. Clarify different roles of OAI-ORE and SWAP.

Recommendation 13: Target UK contribution to ePrints and DSpace in terms of development effort and funding to ensure strategic deliverables are prioritised.

Recommendation 14: Explore use of cloud computing to support repository storage and services. Consider what repository infrastructure is best located at the local institutional level and what is better outsourced to help alleviate cost implications.

Recommendation 15: Follow SWORD development pattern for other repository related applications. Demonstrate use of SWORD to deliver deposit to multiple repositories.

Milestones by repository content type

Research Outputs
Milestone 1: Repositories need to clarify their roles in relation to other repositories, whilst acknowledging they exist within a 'mixed economy'.
Milestone 2: Support population of institutional repositories by advocacy, case studies, guidelines on best practice, encouraging institutional mandates, encouraging inclusion of open access and management of the life-cycle of digital content in institutional strategies.
Milestone 3: Explore and address integration between institutional repositories and: other institutional systems, other types of repository, funders' systems.
Milestone 4: Develop added value services layered on repository content (e.g. tools for deposit, search, re-use, linking data, metadata enhancement, citation metrics, publication lists).
Milestone 5: Establish an agency/lead body within the JISC HE community to take a lead on legal issues connected to copyright and publisher policies.

Research Data
Milestone 6: Clarify responsibility for feeding existing repository implementation experience into current planning activity for research data centres. Potential candidates for this role are DCC, UKOLN and JISC. Taking experience of previous JISC programmes, and existing IRs on board, the interaction between different types of research data centre should be defined at an early stage. Work in this area needs to be undertaken in collaboration with the UK Research Councils. There is a need to establish metrics for populating research data centres and measure impact.
Milestone 7: Ensure the multiple objectives that management and re-use of research data supports (e.g. discovery, access, re-use and preservation of data) are taken into account in proposed solutions. Ensure the requirements of different research data types (big science, small science, different disciplines) are taken into account in proposed solutions.
Milestone 8: Formulate national strategy to take account of differing roles for different types of HE institution.
Milestone 9: Target outcomes of repository activity at appropriate stakeholder groups.
Milestone 10: Incorporate research data management training into repository projects.

Learning Materials
Milestone 11:

- Identify best practice for the management (including discovery, access and re-use) of learning materials at disciplinary, regional and national levels.
- Establish rewards and incentives for sharing.
- Set up measures for progress and impact.
- Integrate repositories with VLEs and e-portfolios.
- Work with QAA towards adherence to guidelines for good management of learning materials.
- Facilitate deposit of learning materials in repositories.
- Explore how different learning materials repositories might interact and in particular how Open Educational Resource initiatives fit with repository initiatives.

- Explore how different learning materials repositories might interact and in particular how Open Educational Resource initiatives fit with repository initiatives

2 Introduction

Aim

Since the Repositories Roadmap[1] was published in 2006 a lot has changed – on the Web, within the higher education (HE) research and teaching environment, and within the repository landscape itself. On the Web there has been a growth in social networking, user generated data and network level services[2]. Within UK HE institutions the RAE 2008 is in its final stages and there is now discussion of its replacement (the Research Excellence Framework). Taking a wider perspective, there has been an increase in cross-disciplinary and data-centric research, and growing awareness of the need for digital preservation. Significant targeted funding in the UK has led to many repositories being set up by both institutions and research funders. In particular JISC funding programmes have generated large numbers of repository based projects. Internationally there have been notable repository related initiatives in North America, Australia, Europe and Japan. In the light of these environmental changes it is timely to review the original Repositories Roadmap and consider the impact of these changes on the way forward.

The original Roadmap presented a vision for the role of repositories in 2010. It compared this vision to the landscape current in 2006 and indicated milestones that would need to be achieved to reach that vision. Given the many changes in practice, policy and technology since 2006 it is timely to review the Roadmap. This Review seeks to extend the horizon to 2013, to clarify the relationship of repositories to the broader environment and to steer the future work of JISC and others interested in furthering repository implementation and enhanced scholarly communication.

This Review is based on a consultation exercise with experts in the field by means of a Web based forum (IdeaScale), a questionnaire and a workshop. The experts' views have been used to inform the author's presentation of a set of priorities for action. Of necessity the contributions have been interpreted and emphases may have been changed. (Note that within the Review direct quotes from participants are indicated by text in italics.) The Review is a light-weight exercise to highlight priorities, it does not intend to give a comprehensive picture of what is now a highly complex landscape. JISC has funded several studies to inform and influence decisions on the funding of future activity. It is not the role of this Review to offer a synthesis of this large body of work, rather to recommend direction and milestones based on high level analysis of the results of the consultation process.

The Review is structured into two parts. Firstly it makes a number of recommendations targeted at the JISC Executive to inform further funding of repository related activity. The Review then goes on to identify a number of milestones of relevance to the wider community that might act as a measure of progress towards the wider vision of enhanced

---

[1]Rachel Heery and Andy Powell. Digital repositories roadmap: looking forward. 2006 www.jisc.ac.uk/uploaded_documents/rep-roadmap-v15.doc

[2] Examples of network level services in this context are Amazon S3, Google Docs, authentication services, identifier services.

scholarly communication. These milestones are for the near term to be addressed as a means to reach the vision for scholarly communication in 2013. The milestones are targeted separately at those with an interest in specific types of repository content (research papers, research data, learning materials). Achievement of these milestones will best be assisted by JISC through its community work and funding programmes. There is an additional section focusing on enabling technology. Whist the consultation exercise has informed both the recommendations and milestones, the latter are more closely based on direct input from consultation.

Audience
The principal audiences for this the Roadmap Review are:

- the JISC Executive,
- the Repositories, Preservation and Asset Management Advisory Group (RPAG),
- the wider repository community

There is potential for a further version (or versions) of the Review to be targeted at particular audiences whether JISC project participants, institutional repository managers or the wider digital information community. Options for taking this Review forward were considered at the October 2008 JISC RPAG meeting and are discussed in the conclusion. A possible outcome will be further integration of the Roadmap into JISC's evaluation process for funded repository activity.

Scope
Just as with the original Roadmap, the Review will focus on repositories within the context of UK higher education. The Review follows the pattern of the original Roadmap by considering milestones for different content types separately: research outputs[3] (including text, images and multimedia), learning materials, and research data. In the Review geodata is subsumed under research data rather than being given separate treatment.

The original Roadmap was concerned with objects created, owned and shared by members of the HE community not those made available on a commercial basis. Also institutional administrative records were deemed out of scope. Whilst this overall focus remains, given the increase in repository activity it is now appropriate for a repository roadmap to take a wider view and consider the potential for interaction between repositories and other institutional systems that might be considered 'administrative', particularly research information systems, content management systems and author identity systems. Similarly as institutional and national repositories become established then their role in relation to commercial repositories becomes interesting, and this Review suggests development of a stronger relationship between repositories and journal publishers, A&I service publishers, and global service providers. This broadening in scope has been done only to a limited extent in the Review due to constraints on time and effort.

[3] Within the Roadmap and Roadmap Review the phrase 'research outputs' refers to the textual material that makes up a high percentage of the content of existing repositories i.e. research papers, journal articles and dissertations. In future research outputs might also include textual work in progress, and more complex items such as multimedia pieces and images.

3 Recommendations to JISC Executive

3.1 Charting progress

Over the last two years the level of repository activity has increased within the UK and beyond. Progress against the original Roadmap has been made on several fronts. A majority of UK HE institutions have implemented repositories. Institutional repositories (IRs) have been populated with content of various types, albeit much of the attention and discussion has focused on open access to peer-reviewed journal articles. At the national level six out of the seven RCUK funders are encouraging open access to research outputs by means of mandates (for details see JULIET[4]). Perhaps encouraged by such mandates, researchers are now contributing at least some of their output to institutional repositories, national funders repositories (notably UK PubMed Central), and subject repositories (such as arXiv). Intermediaries are also populating repositories on behalf of authors.

One of the motivating drivers for establishing repositories has been to support open access in order to disseminate research outputs more effectively. The open access agenda is influencing publishers who are to a limited extent accommodating the ambition to make research outputs open access, while attempting to cover their costs, for example by payments from Wellcome's Value in People[5] awards. Open access journals are establishing a business model that is beginning to be absorbed by the more innovative publishers, for example the recent acquisition of the Biomed Central Group by Springer.

It seems likely that repository activity has introduced a level of organisational and cultural change within implementing institutions, and that technology capacity has grown with the spread of skill sets among institutional repository staff. IRs were used to support the 2008 RAE, and several institutions are establishing repositories motivated by the possibility of metric based indicators within the forthcoming Research Excellence Framework (REF) assessment.

Wider interest in the strategic importance of research data management has emerged partly driven by the DCC. There has been a growing interest both in the curation and preservation of data. Initiatives are underway to provide a UK Research Data Service[6]. Within HE institutions, where research is increasingly data-centric, there is growing awareness of the need to preserve and re-use data.

Though various statistics are available from both ROAR and OpenDOAR, the figures to measure this progress are less than precise. Whilst these statistical services show that approximately 97 UK HE institutions now have active repositories (whilst 72 do not) there are gaps in the information available about content, in particular how many full text items are accessible. Recent estimates from repository managers suggest that their IRs contain at least 35K full text items available on open access, but these figures are not very useful even as an estimate as they are based on responses from only 14 institutional repository managers. The figures result from a one-off request made by a JISC Programme Manager to IR managers in mid-2008.

---

[4] Sherpa JULIET http://www.sherpa.ac.uk/juliet/

[5] Wellcome Trust open access funding http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/Guides-and-FAQ/WTX036803.htm

[6] UKRDS. The UK research data service feasibility study. July 2008. http://www.ukrds.ac.uk/

Also, there is little information available about level of deposit and usage of repository content. Detailed analysis of trends by type of content being deposited and level of usage would be welcome. Similarly more understanding of who is depositing content, their motivation and behaviour patterns, would assist fruitful approaches to encouraging deposit.

At present there is no indication as to the overall target for IR content by which success criteria could be set. In particular, the lack of information regarding learning material was highlighted during the consultation exercise. It may be possible to estimate publication numbers for scholarly works in the UK (for example using existing figures from Evidence Ltd derived from Thomson Scientific data[7]). If so these could be used to set targets for deposit of content. As more attention is given to opening access to research data then some baseline analysis of targets for data content, by type and size, would be useful. Whilst simple totals of the number of items in repositories could be misleading, particularly when legacy content is being gathered, some measures need to be in place to understand the patterns of growth in repository content e.g. by content type, whether content includes full text, etc. JISC may want to consider working with the community to set targets for a certain percentage of content to be made available. This could serve as a focus for funding, advocacy and support. JISC would be unable to set such targets unilaterally and would need to design a process for producing targets in collaboration with the HE community.

To justify the effort involved, any performance metrics would need to provide value to the institution as well as to funders. Metrics need to be tied to the benefits provided by repositories, and therefore should be associated with the objectives that JISC funding and institutional investment is aiming to achieve. As well as metrics for open access material, value needs to be placed on collaboration within subject groups or institutions (sometimes in closed ways). In addition the benefits of content being well managed needs to be factored into any metrics.

As well as statistical metrics, JISC should consider softer evidence such as testimony from senior managers and researchers as to how good content management has improved their working lives.

Recommendation 1: Chart progress of the implementation and usage of repositories by measurable indicators, for example

- **Measure cultural and organisational change by whether relevant repository related deliverables are included in institutional information strategies (in particular open access, integration of institutional systems, digital preservation and curation, well managed digital collections).**

- **Produce baseline analysis of current repository content and analysis of potential repository content.**

[7] *PSA target metrics for the UK research base*. Evidence Ltd. for Department of Trade and Industry Office of Science and Innovation, March 2007. http://www.berr.gov.uk/files/file38817.pdf

- **Enhance repository statistics to measure the availability of open access full text items (e.g. using ROAR[8] or OpenDOAR[9] or Intute[10]).**

- **Analyse patterns of deposit in those institutions with an institutional mandate compared to other institutions. Who is self-depositing and why?**

- **Scope metrics, qualitative as well as quantitative, to demonstrate the value of repositories to institutions and their members.**

3.2 Working towards user involvement

It is notoriously difficult to involve end-users in development projects. Nevertheless their involvement is vital in discussions of scholarly communication. Although JISC projects already have input from experts in digital information management, such experience is very different from that of many researchers and teachers. IR managers are well placed to remedy this situation by developing contacts with research and teaching staff to enable their involvement. If we could learn more about what researchers and teachers are doing now, particularly around Web 2.0 applications, we could build on this. David de Roure's presentation[11] at the Repository Fringe in Edinburgh 2008 exhorted "Don't think roll-out of services, think roll-in of researchers".

Recommendation 2: Analyse current communication behaviour of researchers and teachers, and involve them in development of future scholarly communication services.

3.3 Guidance through a complex landscape

Progress has been made since the original Roadmap. However the plethora of repository related activity means that higher level objectives can easily get lost in detail. This is an issue at the policy level but perhaps more so for the individual institutional repository manager trying to reconcile the needs of their particular institution with wider activity.

The repositories landscape is complex. A network of repositories is emerging made up of various repository types and multiple stakeholder groups are involved in this repository activity. Achieving good communication between these various groups becomes increasingly important. Various actions might be taken to achieve more clarity in communicating repository roles and objectives.

---

[8] Registry of Open Access Repositries: http://roar.eprints.org/

[9] The directory of open access repositories: http://www.opendoar.org/

[10] Intute Repository Search: http://www.intute.ac.uk/irs/

[11] Closing Plenary by Dave de Roure. Repository Fringe. Sep 2 2008, The University of Edinburgh. http://video.google.com/videoplay?docid=3946866546394622361&hl=en

3.3.1 Terminology

The first step of the Review addressed ongoing concerns expressed within the repository programmes about use of the word 'repository'. The IdeaScale Web tool[12] was used to initiate discussion and this was followed up by further consideration of the definition at the Roadmap workshop.

Within the JISC repository development community many have been critical of over-use of the 'repository word'. Such on-going debate over terminology is not unusual within the context of development programmes, however a number of themes emerged in these discussions revealing that uneasiness about use of terminology around 'repositories' is symptomatic of uncertainties about the role of repositories, and how repositories fit with higher level objectives.

It is worth noting that terminology has been an issue since the early stages of the 'open archive movement'. The differences between repositories, archives and digital libraries are unclear. The underlying technology is similar, the purposes overlap. As 'digital stores' repositories can fulfil a number of diverse roles, sharing the same technological characteristics as archives and digital libraries, leading to ambiguity as to what is the primary aim of implementing repositories. 'Repository' as a term in itself has not carried much distinctive meaning about purpose nor motivation.

Discussion about terminology went wider than mere definition into questioning the role of repository in relation to high level objectives[13]. Various themes emerged:

- use of the word itself is unhelpful to most audiences as it has the wrong implications, is imprecise, and is disconnected from higher level objectives
- the term always needs qualification

Use of the term 'repository' to audiences of researchers and teachers is unhelpful. The term is unfamiliar and has negative implications of inactive, inert storage. A repository is intended to support higher level objectives (open access, asset management, sharing and re-use), so the emphasis needs to be on these aims not on the technical means to achieve them. Repositories do not stand alone in being criticised in this way. At the recent iPres 2008 conference the Chief Executive of the British Library, Lynne Brindley, argued for a new strap line for digital preservation to emphasise its purpose: 'Preservation for Access'[14]. It is important for communication with stakeholders to stress the objective a repository fulfils e.g. making content available on the Web; sharing research data; improving digital information management.

Often assumptions are made about repositories that lead to misunderstanding. There needs to be acknowledgement of the diversity of repositories e.g. it cannot be assumed that repositories are open access as a repository can be used to share resources amongst a

---

[12] Repositories: communicating the idea. IdeaScale web discussion available at http://jiscrepository.ideascale.com/

[13] Chris Rusbridge, Digital Curation blog. Summary of IdeaScale discussion. http://digitalcuration.blogspot.com/2008/08/comments-on-negative-click-research.html

[14] IPres 2008. JISC Information Environment team blog.

http://infteam.jiscinvolve.org/2008/10/02/ipres-2008/

closed group; repository content is frequently a hybrid of metadata only items and full text items rather than completely full text; content is varied and does not consist only of peer-reviewed journal articles; depositors are often intermediaries rather than authors.

Discussion within the Web forum concluded that while a focus on objectives rather than on repositories would be helpful to wider audiences, within the development community use of the term is inevitable, providing a focus for widespread existing activity as well as providing some indication of shared experience. Though even amongst repository managers and developers the word 'repository' needs qualification as there are so many different types of repository.

On a pragmatic basis, the definition of 'repository' used by the Repository Support Project[15] was recommended, with minor modification. This definition refers to the objectives of a repository, it indicates the various organisational types of repository, and acknowledges diverse content:

A digital repository is a managed, persistent way of making research, learning and teaching content with continuing value discoverable and accessible. Repositories can be subject or institutional in their focus. Putting content into an institutional repository enables staff and institutions to manage and preserve it, and therefore derive maximum value from it. A repository can support research, learning, and administrative processes.

3.4 Multiple roles of repositories

Discussion of terminology broadened to consider the multiple roles of a repository and how these fit with the vision of enhanced scholarly communication. Following two years of intense repository activity, the role of repositories seems diffuse, and the lack of clarity about objectives is having a real impact on those working in this area.

Themes that emerged included

- institutional repositories need to meet the local requirements of their users (as content creators) as well as the wider objectives of other stakeholders (searchers, institutions, funders)
- the 'boundaries' of what constitutes a repository are unclear
- repositories, particularly institutional repositories, need to interact with other systems to achieve the objectives associated with them

The original Roadmap vision for 2010 was articulated as

"a richer scholarly communication environment, based on open access to, and re-use of, scholarly materials. The phrase 'scholarly communication' is used here in its richest sense to include the life-cycle of information and knowledge from research to learning."

In general terms this vision is still considered valid. However the vision is being interpreted in different ways from a fairly narrow focus on open access to a much more wide-ranging change in the publication process and the research environment.

[15] RSP Web site. http://www.rsp.ac.uk/repos/definitions. Accessed June 2008.

Indeed open access itself is understood differently, varying from concentration on peer reviewed journal articles, to providing access to previously hidden grey literature[16], to making accessible versions of work in progress. Similarly there are different emphases on what richer scholarly communication incorporates, from a narrow interpretation of an increase in electronic journals, to a revolution in scholarly publishing patterns using a more Web based environment (Open Science, scholarly social networks and less emphasis on journal articles, centralised stores of shared learning materials).

Repositories might play various roles in future visions of scholarly communication, whether repositories as we know them now or more innovative types. Existing work in Open Science, Grid and Virtual Research Environments provides pointers to change in the scientific research process (see Carole Goble's presentation to the British Library Board[17] and David de Roure's presentation at Repository Fringe 2008 as mentioned in section 3.2 above).

This Review was not intended to look at this wider picture of the future of scholarly communication in detail. However a very brief preliminary outline will be given which might suggest how further scenarios might be developed

Vision for scholarly communication 2013

Scholarly content is created, stored and shared on the Web. Despite the increase in less formal communication, for researchers the peer reviewed paper is still the key communication unit of record and status, and reward systems are based on peer reviewed outputs.
The amount of informal Web based communication is increasing and where it is deemed worthwhile is captured in repositories containing wikis, blog entries etc
OpenScience is the norm, with data produced in labs typically being stored automatically and made open for re-use.
The primary level of social interaction for researchers, teachers and learners is at the group level (cross institutional project, institutional research, lab or teaching group). A typical workflow is for a small group to produce content which is then shared with a few trusted individuals in a wider group for informal peer review. As confidence in the content grows it is shared with a wider audience and undergoes formal peer review. Repositories support these various levels of sharing resulting in an open access content item.
The group also requires access to existing resources for 're-use' in support of the creation of new output[18]. A group repository enables access to research papers and data for re-use, and for the teacher access to existing learning materials for re-use. Students can build their own curriculum from open educational resources repositories.
Repository content would reside in various types and 'levels' of repository, with content gravitating to the appropriate level for preservation depending on the value of the content.

---

[16] Talis talks with Herbert van de Sompel about SFX, OAI, and Repositories. http://blogs.talis.com/xiphos/2008/09/06/talis-talks-with-herbert-van-de-sompel-about-sfx-oai-and-repositories/

[17] Carole Goble, The Future of Research (Science and Technology), presentation at British Library Board Awayday, 23 September 2008, http://www.slideshare.net/dullhunk/the-future-of-research-science-and-technology-presentation

[18] The term re-use refers to manipulation and adding of value to existing research outputs and research data. For example journal article re-use might include citation and text mining. Research data re-use might include re-analysis of original data, combination or comparison of original and additional data, data mining.

Some repositories would reference content stored elsewhere, others would have a preservation function storing complete objects.
The storage of large data sets may be 'in the cloud' as an alternative to storage in HE institutions or data centres.

Recommendation 3: Given the environmental changes in the education, Web and scholarly environment we need to articulate an updated vision of richer scholarly communication. The vision should be based on the scholarly life-cycle from experiment through to research, teaching and learning. The various roles repositories might play should be mapped out, particularly as regards management of digital resources, open access and re-use.

The 2006 vision has been associated with several different high level objectives. From the start the JISC repository funding programmes objectives were broad to include improved asset management and to address needs for preservation, as well as promoting open access.

Repositories are now seen to support several high level strategic objectives:

Making more digital content available
Supporting the open access agenda for peer-reviewed journal articles
Curating and sharing research data
Improving institutional management of learning material
Improving institutional management of research outputs
Preservation of digital resources
Supporting new types of research environment

Whilst repositories have a role in supporting each of these objectives there are many other systems as well as policy and organisational issues involved. There is a danger that by not sufficiently differentiating these various objectives, the role of a repository in relation to any particular objective is unclear. Further distinguishing and prioritising these high level objectives and related activity would assist individual institutional repositories to set their own priorities and relate to wider activity.

Recommendation 4: Communications about repositories should emphasise higher level objectives. It would be helpful for JISC to differentiate high level objectives and subsume repository activity under those objectives. There needs to be a shift in emphasis from the 'repository' to the objective. JISC should ensure calls and funded activity relate to particular objectives rather than solely to 'repositories'.

In the context of these multiple high level objectives, repositories are being established within particular institutions, each institution with its own community and requirements. In these institutions the repository will need to meet the requirements of that institutional community. The stakeholders may be researchers as they act as experimenters/data users/collaborators/authors; or teachers as creators of learning materials. Other stakeholders may be information managers. High level objectives need to be related and aligned with the local requirements of the institution.

It is often the case that a repository is set up as a pilot on the basis of a high level global objective, but with unknown detailed local requirements. Pretty quickly the local repository

will need to meet local requirements in order to achieve short term success. There are a range of requirements that the repository might fulfil. The Repositories Support Project site in its pages on 'making a case for a repository' [19] lists nine benefits for researchers, ten for institutions, two for funders, thus illustrating the variety of local requirements there might be for establishing an IR. The varied motivation for implementation will result in different repository functionality. In addition there is likely to be further diversity in requirements for functionality for different types of content. Recognition of the complexity inherent in this mix of objectives, and the challenge presented to repository managers, is well documented by Dorothea Salo[20]. Progress depends on institutions integrating IRs into their wider strategy for digital collections and digital information management.

In order to engage content creators within the institution, objectives need to be presented to users in terms of *familiar processes and tools*. Ideally objectives will be met by identifying *common points in a large number of workflows that repositories could hook into*. This acknowledges that content creators are likely to be motivated by systems that streamline their own familiar tasks rather than by higher level objectives. This is supported by a case study from the California Digital Library[21] and a recent report by Carole Palmer[22] showing creators are motivated by their own requirements for higher impact and increased citation rather than open access as such. In the IdeaScale discussion this is expressed as *a repository needs to be defined in terms of researcher workflow*, supporting the work of the user and creating repository content as part of this process.

The available software tools and JISC funded activity needs to support these diverse local institutional requirements. There needs to be work done to provide flexible solutions which at the moment tend to be divided between distinct platforms (eprints repository, learning materials repository, digital library, data store, preservation archive).

Recommendation 5: Tools and project outputs should be presented in terms of solutions to local institutional repository requirements.

Differing views were expressed in the consultation exercise on the richness in functionality of a repository leading to questioning the boundaries of a repository. Repositories were variously characterised as rich multi-functional environments (see Chris Rusbridge's Research Repository Systems) to simple back-end data stores with added value services located elsewhere[23]. Whilst the overall ambition is similar i.e. to support the researcher's workflow in an integrated way, this variance in views leads to misunderstanding as to what a repository should be achieving.

Once again it would be helpful for repository activity to be subsumed under higher level objectives e.g. developing Virtual Research Environments, managing content for Virtual

---

[19] http://rsp.ac.uk/repos/justification

[20] Salo, Dorothea. *Innkeeper at the Roach Motel*. Library Trends 57:2 (Fall 2008). http://digital.library.wisc.edu/1793/22088

[21] RSP Web site. eScholarship repository case study. http://www.rsp.ac.uk/repos/casestudies/california.php

[22] Carole Palmer *et al.* Identifying factors of success in CIC institutional repository development. Andrew Mellon Foundation, August 2008. https://www.ideals.uiuc.edu/handle/2142/8981

[23] Chris Rusbridge has summarised much of this discussion on the Digital Curation blog. http://digitalcuration.blogspot.com/2008/08/comments-on-negative-click-research.html

Learning Environments. JISC's late 2008 funding call sought to address this issue by encouraging integration of the e-Research and IE themes in project bids.

Recommendation 6: Repository activity needs to be joined up more closely with other forward looking work such as the JISC Virtual Research Environment activity, open science, data sharing, preservation. At the same time at the local level there needs to be support for integration of repositories with institutional systems (Current Research Information Systems, Research Excellence Framework, Virtual Learning Environment, author identity systems).

3.5 Exploiting the Web

Note that aspects of the debate relating to adherence to Web architecture are covered in section 4 of this Review.

Some frustration has been expressed within the JISC programmes and in the Review consultation that Web 2.0 applications might be better exploited to support objectives associated with repositories. Web 2.0 is being used here as a catch-all for recent developments on the Web, not just wikis and blogs but also social networking, user generated content, cloud computing etc.

This is an area where JISC has a leadership role, however in practical terms this can be difficult to achieve. The rate of change on the Web means that as well as funding exploratory projects within the context of JISC programmes, it would be useful to have existing development oriented groups exploring new technology without the need for a long lead time. In order to provide leadership there has to be the ability for more immediate engagement in the rapidly changing Web environment. For example a recent DLib article asks for collaboration on djakota[24], an open source image server that allows for Web 2.0 style re-use. Such collaboration could be achieved quickly and effectively through an existing group of funded experts.

In addressing how aspects of Web 2.0 might be used in conjunction with repositories it is important to take account of the previous recommendations to focus on objectives rather than on the repository in isolation. The aim is to use Web 2.0 to achieve high level objectives for scholarly communication and information management, the interest is wider than in the repository alone.
It is worth briefly considering various approaches to how Web 2.0 might be exploited in order to clarify what would be best next steps. Whilst some of these approaches are already being explored, the work in this area is patchy at best.

There are a variety of ways in which Web 2.0 might support research and teaching. So for example various applications for social networking might support research team collaboration or interaction between learning groups e.g. use of wikis, Ning. Specific tools have been developed to encourage social networking amongst researchers and to enable sharing of research data such as MyExperiment and OpenWetWare. Exchanging information between groups is facilitated by bookmarking and tagging aggregations like FriendFeed, and Technorati. Some of these applications require access to collections of

[24] Ryan Chute and Herbert Van de Sompel. Introducing djatoka: a reuse friendly, open source JPEG 2000 image server . DLib Magazine Sept/Oct 2008. http://www.dlib.org/dlib/september08/chute/09chute.html

content (whether data or research reports) and it is here repositories may have a role. These might be existing repositories or more innovative Web based transitory repositories.

There has been a growth in large collections of 'concentrated' user generated data (e.g. Flickr, Slideshare) which exhibit a network effect whereby others are encouraged to deposit, rate and tag content. Could aggregations of scholarly content be gathered with the same beneficial network effects? How would such collections interact with existing or more innovative repositories, and other collections of content? What are the licensing and copyright issues? Lorcan Dempsey's ideas are relevant here on issues of diffusion and concentration[25] on the Web.

There needs to be some reflection on how well Web 2.0 applications fit with scholarly communication. There may be particular characteristics of scholarly communication which mean special tools need to be developed to provide a 'well-managed' solution allowing for re-use and preservation. There are issues of IPR and ownership that might discourage (or even make illegal) deposit of peer-reviewed articles in global collections on the Web. There may well be a danger in creating silos of content that are neither interoperable nor accessible for migration. Outsourcing data storage to commercial third party suppliers has implications for sustainability, as has been shown by Google's withdrawal of its beta scientific data storage facility in late 2008[26].

There are further issues on how best to achieve search engine optimisation (SEO) on existing legacy scholarly content, and how best to achieve SEO on future content.

To address these questions a mix of policy and technical expertise is required. In order to lead in this area JISC need to exploit existing Web expertise, and make available Web development skills. In response to the need to address technical issues of interoperability between repositories JISC set up the CRIG (JISC Common Repository Interfaces Working Group). This group has involved developers to identify and address priorities for enhancing inter-working between the emerging base of UK repositories. Here we are thinking about exploiting existing Web 2.0 based tools and services, creating applications that interface with existing tools and services, adapting existing Web 2.0 tools and services.

Recommendation 7: Harness existing JISC funded expertise to take the lead on exploiting the Web for teaching and research. Ensure existing JISC services have responsibility to advise and demonstrate use of new Web developments and services for scholarly communication to include aspects where repositories have a role. This may involve UKOLN, CETIS, and/or the Repositories Support Project. Extend the remit of CRIG (or a similar group) to address interfaces with Web based systems and Web 2.0 initiatives.

As with much JISC activity, within the repository programmes projects range from those undertaken to support short term change to those with more experimental, explorative or long term aims. JISC are looking to the future, supporting change now. Institutional repositories need pragmatic short term enhancements to ensure success at the local level. Others are looking at more complex long term developments such as integration of

[25] Dempsey,L. *The two ways of Web 2.0.* March 2008. http://orweblog.oclc.org/archives/001556.html

[26] Savvas, A. *Google closes data storage service for scientists.* ComputerWeekly.com, December 19 2008. http://www.computerweekly.com/Articles/2008/12/19/234014/google-closes-data-storage-service-for-scientists.htm

repositories with VREs or long term digital data preservation. In this environment, it is difficult for any individual repository to keep track of the relevance of much activity and relate it to their own requirements.

Recommendation 8: Ensure clear classification of the outputs of JISC funded activity into categories useful to repository managers such as relevance for adoption or development; in short term or long term; ease of implementation; relevance to different repository types.

4 Enabling technical infrastructure

Note that details of 'where we are now' are not included for the technical infrastructure section as this information was not gathered explicitly as part of the consultation. As it is, much background to the current situation is implicit in the following sections and section 3.5 on exploiting the Web.

4.1 Technical infrastructure: where we want to be in 2013

As previously discussed a renewed vision for scholarly communication in 2013 is yet to be developed. More work needs to be done to develop scenarios of how research and learning will be carried out in the future. The technical infrastructure required to support such scenarios could then be elaborated with more confidence. In the meantime what is self-evident is that scholarly communication in the future will be more Web based than is now the case.

If repositories are to become more aligned with the Web, the processes of deposit, discovery, access, curation and preservation of content will need to be better integrated with Web based services and tools. Repository software and added value services will need to take a RESTful approach and should as far as possible use the existing Web architecture i.e. the Web based HTTP protocol, URIs for identifiers and HTML representation. It is only where scholarly communication has 'special requirements' over and above other content on the Web that there might be a need to use specialised non-Web architecture solutions.

In a presentation at a Talis Xiphos Research Day meeting[27] in 2008 Andy Powell (Eduserv Foundation) argues that we have not got repository architecture right, that the architecture needs to be based on Web architecture rather than the current focus on specialised harvesting protocols (OAI-PMH), institutional collections and aggregators. He goes on to consider how Web 2.0 might further influence 'getting scholarly content on the Web' by exploiting social networks associated with content. Paul Walk, UKOLN, has also undertaken work (in progress)[28] to propose an architecture to support repositories.

This Review will not attempt to rehearse the arguments and make judgements on this debate. Instead some questions and considerations will be raised about options for the way forward. In addition the Review will mention some other technical issues at a policy level.

---

[27] Powell, A. *Web 2.0 and repositories - have we got our repository architecture right?* Presentation at Talis Xiphos Research Day, June 2008. http://www.slideshare.net/eduservfoundation/repositories-and-web-20-have-we-got-our-repository-architecture-right

[28] Walk,P. *Repository architecture #83*. Presentation at JISC Repositories Architecture meeting, July 2008. http://www.slideshare.net/paulwalk/repositories-architecture-83

4.2 Technical infrastructure: how do we get from here to there?

Although we may wish to move to a Web based architecture for repositories there is already within UK institutions a significant deployed base of 'legacy repositories'. Repository software developers need to consider the questions that face many other existing deployed information systems. Is there a way to layer a Web based approach onto this legacy deployment? Would such a Web based approach require development of new software platforms? Should a Web based approach be applied at the level of aggregator (e.g. Intute) or at the local level (the institutional repository)?

Recommendation 9: Explore options for moving to a more Web based architecture for repositories, taking into account the requirement to move forward existing 'legacy repositories'.

Concentrated 'global' collections of content such as Flickr and Slideshare are held up by Powell as examples of successful repositories that "promote the social activity that takes place around content as well as content management and disclosure activity" (see slide 17 of Powell's presentation). However, if we consider current practice in research and teaching it would seem that most social activity takes place at a group level. Typically a research group within an institution or cross-institution works together under a funding grant, sharing knowledge and work in progress. Similarly social interaction connected with teaching and learning typically is based on a cohort of students working together as a group. It would be useful to explore whether repositories associated with research groups and learning groups would be more likely to promote social activity that would encourage tagging, embedded comment, re-use. Note that the JISC Faroes project[29] is doing this with learning materials for language teachers.

On the other hand, 'group facilities' could be layered onto global collections of content, in a similar way to groups forming on Facebook and MySpace, though such an approach might constrain the functionality that a product designed specifically for a small group might deliver.

There are characteristics of some types of scholarly communication that differ from other content on the Web. It may be that such characteristics are incompatible with establishing concentrated collections of scholarly content by means of simple deposit mechanisms. Most obvious is the issue of copyright surrounding deposit of journal articles and learning materials. Another consideration is how concentrated content stored by a third party could be integrated with data in local institutional systems such as a CRIS or REF system. The prospect for centralised collections of datasets seems less contentious than for journal articles.

Recommendation 10: Explore the concentration of the collection of different types of content at different levels - small group (lab, research group, student cohorts), discipline and global levels - and how these different levels might facilitate social networking effects.

Other technical issues include modelling the relatively complex objects that will increasingly be found in repositories. Already there is a significant level of complexity in content such as

[29] http://www.elanguages.ac.uk/researchcommunity/projects/faroes.html

journal articles with several versions with related conference proceedings and associated data; or images with associated rich metadata. Complexity will increase with the collection and re-use of datasets. Two initiatives since the original Roadmap have addressed this problem area, the Scholarly Works Application Profile (SWAP[30]) and Open Archives Initiative Object Reuse and Exchange (OAI-ORE[31]).

As outlined in the project wiki [32], SWAP was intended to address issues that arose from the ePrints UK project, in particular the inconsistent linking to full text within Dublin Core metadata created for repository items. This inconsistent linking makes automated harvesting full text unpredictable, and is currently hampering the work of Intute Repository Search. The resulting SWAP was based on the Functional Requirements for Bibliographic Records (FRBR). It could be argued that the development process for SWAP focused too closely on the modelling of complex objects rather than on the other functional requirements, resulting in an over complex solution for most institutional repository requirements. However if SWAP could be implemented in a user friendly interface the resulting rich metadata would enable rich added value services. Does SWAP deliver sufficient gain for the pain?

Currently the SWAP has not been implemented and the community acceptance and take up plan has not been progressed. However other application profiles based on the FRBR/SWAP model are being formulated for other content types. One option would be to move forward a simple deployment of SWAP at the item or copy level, retaining the cataloguing rules to ensure a consistent approach to linking to full text. This may be an interim solution whilst exploring the benefits of deploying the full SWAP description set of work, expression, manifestation and agent.

Recommendation 11: Explore deployment of a cut down version of SWAP, possibly at the copy level, retaining the cataloguing rules to ensure a consistent approach to linking to full text. Evaluate whether use of SWAP is consistent with a Web architecture approach to repositories.

The Open Archives Initiative Object Reuse and Exchange[33] (OAI-ORE) defines standards for the description and exchange of aggregations of Web resources (otherwise known as complex digital objects). Such aggregations of Web resources might include text, images, multimedia, and increasingly complex data-sets. The intention of OAI-ORE is to describe such aggregations in a predictable way to enable applications to manipulate them to support, for example, authoring, deposit, exchange, visualization, reuse, and preservation. A primary motivation for ORE is to enable re-use of data. Whilst ORE focuses on the relations within aggregations of resources, and SWAP is focused on research papers, there is a potential overlap between functionality enabled by SWAP and ORE.

Recommendation 12: Explore use of OAI-ORE to enable applications to handle complex objects. Demonstrate how OAI-ORE facilitates the re-use of research outputs and research data. Clarify different roles of OAI-ORE and SWAP.

---

[30] http://www.ukoln.ac.uk/repositories/digirep/index/Eprints_Application_Profile

[31] http://www.openarchives.org/ore/

[32] http://www.ukoln.ac.uk/repositories/digirep/index/Functional_Requirements#Conclusions_from_Eprints_UK

[33] http://www.openarchives.org/ore/

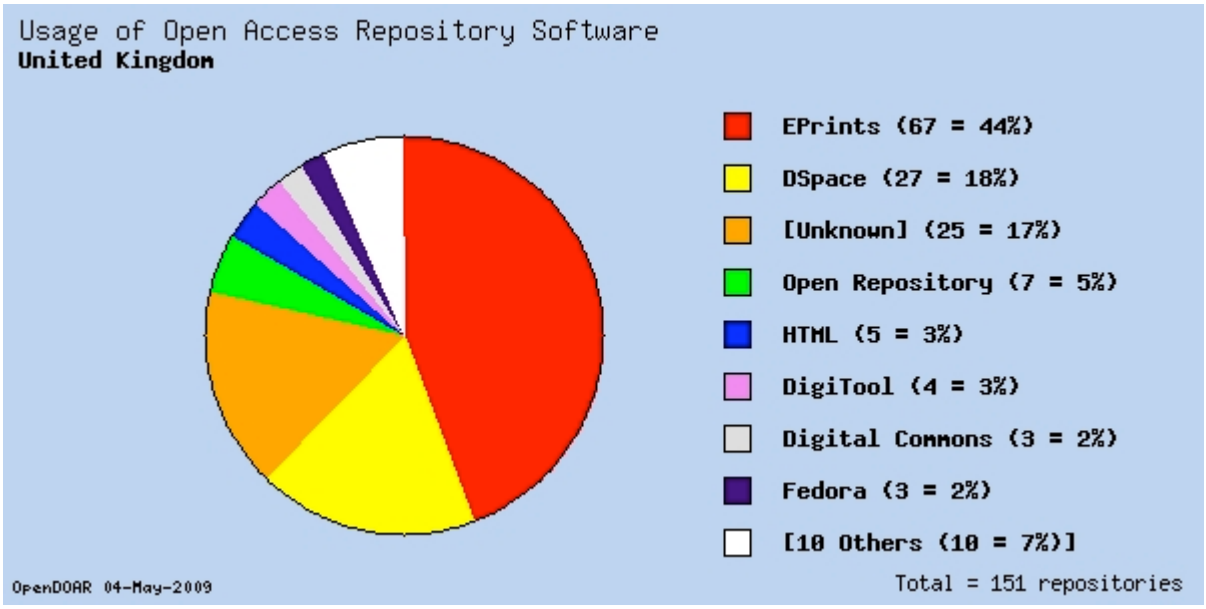

Figure 1. Usage of Open Access Repository Software from OpenDOAR

Turning to software platforms, the majority of UK repositories are dependent on two main software platforms ePrints and DSpace. Both are open source although there are different development models with DSpace looking to a community of developers to contribute code, whereas ePrints development is led by University of Southampton. Extensions to the two platforms will be required to deliver additional functionality to UK IRs such as compliance with SWAP, integration with existing institutional systems, management of datasets, Web oriented architecture. Over time there may be more involvement of commercial suppliers in provision of repository software, and more integration with existing library management systems, but for the foreseeable future there is a significant reliance on the two platforms. Whilst these products will have their own product development plans, contribution from the UK in terms of development effort and funding needs to be targeted to ensure strategic deliverables are prioritised.

Recommendation 13: Target UK contribution to ePrints and DSpace in terms of development effort and funding to ensure strategic deliverables are prioritised.

In reconsidering the repository architecture in the light of recent Web initiatives, issues of outsourcing and scalability arise. What repository functions can be outsourced at the network level, or to put it in Web 2.0 terminology, what can be done 'in the cloud'? Repositories need to grow to accommodate multiple types of materials and the inclusion of data will require much more storage for repositories to work with. It seems likely that data from 'big science' will be stored at some stage in its life-cycle in network services such as Amazon S3.

Other parts of the infrastructure also might be outsourced (preservation, metadata creation, disclosure to search engines, linking data) with national procurement of infrastructure. There could be some local badging of services so that the institutional brand can be preserved. *This would immediately free up a lot of local attention and resource, which currently is trying to solve problems, develop solution and build services in a massively redundant way.*

Institutions need ongoing advice on the use of identifiers, both for repository content and for researchers and institutions. This might be part of the infrastructure offered at a national level.

Recommendation 14: Explore use of cloud computing to support repository storage and services. Consider what repository infrastructure is best located at the local institutional level and what is better outsourced to help alleviate cost implications.

The development of SWORD, an ATOM profile for depositing items in repositories, has proved successful. Developers from various repository platforms came together to specify and develop a simple protocol. This might prove a valuable pattern of development for other repository related applications. SWORD now needs to be taken forward to demonstrate agreement on metadata passed by the protocol for deposit in multiple repositories e.g. deposit into ESRC and a local institutional repository.

Recommendation 15: Follow SWORD development pattern for other repository related applications. Demonstrate use of SWORD to deliver deposit to multiple repositories.

5 Milestones

Milestones for different repository content types were drawn up with the help of input from the questionnaire survey and workshop. Twelve people attended the workshop, excluding JISC staff, and eleven (different) people responded to the questionnaire from a total of nineteen contacted (58%).

Participants were asked to respond to the following questions from the viewpoint of the content type with which they were most familiar:

- Where do we want to be in 2013? What are the main business functions that repositories should fulfil in five years time (from the perspective of the domains you represent)?

- Where are we now as regards fulfilling these functions?

- How do we get to where we want to be? What barriers need to be overcome and how?

Responses to these questions differed significantly depending on the repository content type under consideration. For example responses relating to learning materials indicated the need for a fundamental re-thinking of the higher level objectives, whilst discussion of repositories for data, in particular scientific datasets, acknowledged that the field was immature and that those in the ‘repository community’ could usefully contribute to the initial strategic planning work underway.

Overall the responses showed that it is no longer helpful to concentrate on the future of the ‘repository’ in isolation. As stated above in Recommendation 3 we need to consider a wider picture of where we want to be in 2013 in terms of a renewed vision of scholarly communication. The future of repositories depends on how they can contribute to associated higher level objectives (open access, well managed digital collections, Web

based research environments etc). Rather than concentrating on the technology, such 'business functions' would be a better starting point for discussion of the future. Such an approach may be taken up in any further Roadmap activity, as discussed in the conclusion to this Review.

5.1 Research outputs[34]

5.1.1 Research outputs: where do we want to be in 2013?

Repositories will be established as a collection and delivery channel for content. *In five years time we will have a respectable collection of content at each institution.* Open access will be available to the full text of a variety of research outputs including peer reviewed research papers, work in progress, grey literature, and theses. Whilst some of the content of digital library collections will be closed due to licence agreements, much will be open access. Images and multimedia objects will be included. Content of institutional repositories and institutional digital libraries will be integrated using joint software platforms and a single interface for deposit.

Deposit of research outputs in institutional open access repositories will be mandated by the institution, some funders will also establish subject based repositories. Institutions will showcase (disseminate and share) their research outputs.

Depositing research for open access will be embedded in the research workflow and will be seen by researchers as a positive step towards enabling the wider use of research and achieving high impact. Deposit will be integrated into the research process with minimal effort from the researcher. Deposit of research outputs (including data) will be integrated with the top scholarly research output tools in each subject domain (ranging from word processing to experimental tools).

Institutional repositories will interact with each other to form a network of UK repositories following similar policies and offering interoperable interfaces. UK institutional repositories will further interact with funders repositories and subject repositories by means of a two way flow of content (digital objects) and metadata. There will be an integrated workflow for deposit of journal articles that automatically checks copyright permissions and manages embargoes.

Repositories will support Web based collaboration between institutional and cross-institutional research teams. Collaboration between research groups will be supported by 'transitory' Web based repositories intended to last for the lifetime of the collaboration. These repositories will be used to manage research outputs and data. They will reference content elsewhere, as well as forming the basis for re-use of data and work in progress for the research group. Some of the content of such repositories may well be closed for access by the group until the time comes for wider dissemination. When the group is ready, content will be transferred to institutional or subject repositories for longer term curation as appropriate.

Repositories will facilitate a variety of processes within an institution (publishing, preservation, information management, assessment, teaching, portfolios, marketing). *In five years time we will have more effective processes (e.g. REF) because of the involvement of*

[34] See footnote 2 for how term 'research outputs' is used here.

repositories. Institutions will no longer duplicate organisational systems to manage research process and outputs. Institutions will establish institutional strategies for open access, re-use, and preservation of digital resources as part of their management of the lifecycle of digital content*. Preservation services will be a hybrid of in-house and third party services (national or 'in the cloud').*

Access to content will be integrated with discovery through global and domain search tools. Institutions will ensure search engine optimisation by disclosure of their resources to search engines whether this is done locally or by a third party added value service. *Linked data (links between data, articles, learning materials) will be embedded in portals, search engines and social networking tools.*

The research publishing industry will be influenced by the open access agenda*. A new business relationship will evolve between the Research Communication Support Industry (i.e. publishers) and the Research Industry upon which they depend.*

An emerging e-infrastructure will see publishers and other third parties (commercial and not-for-profits) finding new roles to add value to the process of discovery, access and re-use of research outputs. For example: metadata creation and enhancement, providing XML mark-up versions of texts, text mining, SEO, linking data and journal articles.

Repository access will be integrated with HEI Authorisation and Security provision. This will allow more fine grained access restrictions to content. Whilst this would facilitate restricted group access to work in progress, this is seen as particularly important with regards to access to data.

Automated statistical information will be available on usage and citation of repository content, metrics will be available on the impact of specific items

5.1.2 Research outputs: where are we now?

There is genuine growth in the use of institutional repositories, although deposit of full texts is trailing behind this growth curve. There is much concern that institutional repository content is slow to grow. A recent study[35] by the White Rose Increase project has shown that researchers express willingness to comply with open access mandates, however despite the RCUK and Wellcome mandates, in the absence of awareness raising or enforcement, there is an unwillingness amongst researchers to spend the additional effort required to deposit in open access repositories. In addition there is a lack of awareness in the research community about the benefits of open access and the repository deposit process. Only a few institutional mandates have been introduced (e.g. at Southampton and Glasgow Universities). There is a *lack of strong institutional leadership. The latter is starting to appear, with 21 institutional mandates* (worldwide), *24 funder mandates* (worldwide) *and especially Harvard's statements.*

Within the UK, the RAE/REF have driven social and technological engagement with repositories. Other agendas (e.g. open access and preservation) are trying to take advantage of this government mandated activity. Typically management of digital resources

[35] Increase. Increasing repository content through automation and services: questionnaire key findings. http://eprints.whiterose.ac.uk/increase/Increase_Questionnaire_Findings.pdf

and open access are not embedded in institutional strategy. Overall there is a lack of policy in this area. A change of culture is needed to realise the wider benefits, this depends on development of services layered on repository content and an increase in content types being managed. *The idea of a common managed source of content for different organisational systems does crop up but is not widely implemented. Organisational system management is often too geared around each system, building barriers between them and preventing common solutions from being sought/found.*

Uncertainty about copyright issues is widespread amongst the research community because of lack of clarity regarding publisher and institutional policies. C*onflicting statements of policy from publishers and even delayed and embargoed rights to use are blocking the effective deployment of repositories.* As interest grows in curating and preserving data, publishing companies are seeking to increase their control over the primary data and secondary tools that the research industry depends on. *The UK has a very informed network of repository managers who have effective mechanisms of influence within their institutions, but they are largely being frustrated by the lack of leadership from the research industry over the demands of its own service industry i.e. the publishing industry.*

More weight needs to be given to scholars' requirements. Some studies are underway as to current research practice in different disciplines e.g. JISC's SCARP project[36], RIN disciplinary case studies[37]. There has been too much focus on the repository as a system rather than on the 'resource' (content).

Given the level of activity around repositories it is not surprising that different models of repository interaction are emerging or are being advocated. Little has been said about how institutional, funder and subject repositories can or should interact. A concentrated Web based solution, advocated by some respondents, would be another model. In addition the role of aggregators is unclear, e.g. Intute in the UK, DRIVER at the European level, and OAIster at the international level. There is tension between these different repository content gathering solutions, although they are not necessarily incompatible.

Currently the model for deposit differs between institutional deposit in contrast to deposit in funder repositories (UK PubMed Central or the ESRC Awards and Outputs Database). Wellcome mandates deposit of the publisher's version of journal articles in UK PubMed Central at the time of publication, with a fee to the publisher paid by the institution as part of the indirect costs of the research grant. ESRC requires funded researchers to deposit in their own ESRC repository. The SWORD deposit protocol might assist in enabling a unified deposit to multiple repositories. The White Rose Increase project is investigating the use of SWORD[38] to deposit in a local IR and the ESRC Awards and Outputs Database.

Various areas require more work. There is uncertainty about the best way to disclose the content of repositories to search engines (Search Engine Optimisation). There is limited work progressing the potential for repositories to contribute to linking data, whether linking different versions of journal articles, articles with associated data, articles and presentations.

---

[36] http://www.dcc.ac.uk/scarp/

[37] Research Information Network. Disciplinary case studies. http://www.rin.ac.uk/case-studies

[38] Increase. ESRC scenarios. http://eprints.whiterose.ac.uk/increase/esrc.html

Digital preservation is not an organisational priority. Current repository software platforms are not particularly geared up to preservation, rather the focus has been on access. It is unrealistic to expect small institutional repositories to undertake costly preservation audit exercises. Further exploration and practical use of solutions is required.

Currently many institutional repositories are at project status and are managed as pilots. Institutions will need to address moving repositories from pilot to sustainable service status. As part of this process institutions need to decide who owns the repository: the library or the research support division? Institutions will need to decide whether to establish departmental repositories or a single institutional repository.

5.1.3 Research outputs: how do we get from here to there?

Institutional repositories providing open access to journal articles are the most mature of repository types in the UK, and have been the focus of much of the funded activity over the last few years. These repositories co-exist with activity in new areas of interest (e.g. re-use and preservation of data, REF, Web 2.0, linked data, national repositories and publishers' initiatives). In effect there is a 'mixed-economy' where the role of different types of repository is still evolving and there can be tension between different approaches. In this situation lack of clarity needs to be addressed wherever possible whilst accepting that future change is not always predictable.

Milestone 1: Repositories need to clarify their roles in relation to other repositories, whilst acknowledging they exist within a 'mixed economy'.

Regarding research outputs (in particular academic papers), the proposed way forward for institutional repositories suggested by the consultation can be summarised as: population, integration, adding value.

Population of IRs is a priority. This will involve ongoing advocacy. Case studies are considered to be a valuable approach. There needs to be greater evidence gathered of the benefits of disseminating and sharing open access digital content using the IR as a well-managed source of content supporting preservation and life cycle management for digital content. Case studies need to demonstrate how open access content available for text mining and re-use enables new discoveries to be made. The emphasis needs to be on how to manage digital content rather than on the repositories themselves. *Demonstrate benefits repositories can contribute to other processes, demonstrate novel ways in which content can be used.*

Clear guidelines should be drawn up covering the full range of methods for populating repositories including deposit by intermediaries, harvesting from other repositories, searching and downloading from other existing online services such as the Web of Science. For one account of how to pre-populate a repository see Mark Leggott's description of the RIB project[39], University of Prince Edward Island, that aims to present faculty with a 90% populated IR.

[39] Mark Leggott, Repository in a Box. http://loomware.typepad.com/loomware/2008/09/rib---repositor.html

Encourage further mandates for deposit, particularly institutional mandates and encourage enforcement of these mandates. JISC project funding for repositories might be made conditional on institutions establishing such mandates. Engage with EPSRC in their discussions on mandates at the forthcoming EPSRC Council meetings.

These various approaches to populating IRs need to be underpinned by inclusion of open access in institutional strategies. Institutions also need to consider strategic issues associated with good management of the life-cycle of digital content (preservation policies will flow out of this).

Milestone 2: Support population of institutional repositories by advocacy, case studies, guidelines on best practice, encouraging institutional mandates, encouraging inclusion of open access and management of the life-cycle of digital content in institutional strategies.

There are a number of areas where there is a need for **increased integration:**

Between subject, institutional and funder repositories
Between systems and processes within institutions
With REF systems
Between repositories and Web based services
Between aggregations of content and OpenURL resolvers
With preservation services

Develop demonstrations of embedding repositories in the scholarly workflow e.g. making deposit easy from content creation tools; enabling more automated means of grant reporting and building personal portfolios; assisting funder organizations with metrics and administration.

The emerging model of distributed preservation services needs to be explored further and uncertainties resolved about interaction between different types of repository.

Milestone 3: Explore and address integration between institutional repositories and: other institutional systems, other types of repository, funders' systems.

Provide added value services layered on repositories: tools for deposit, search, re-use, metadata enhancement, linked data, citation metrics, publication lists. These services might be provided at the 'network level' by the university sector or not-for-profits, by publishers or others in the commercial sector; alternatively software solutions might be provided as integrated add-ons to existing repository software platforms.

Customer relationship management (CRM) is a vital and local institutional activity that may be supported by centrally produced services. CRM involves outreach and the understanding of particular research and learning needs at a local level, there may be specialist clusters that can be called on. Local requirements may be met by centralised services, or by enhancing existing repository software platforms.

The role of publishers in supporting the research process needs to be addressed. Publishers add value to the research process incurring costs. Funders must be explicit on how their grant holders can pay these costs. Open access advocates need to develop an

agenda for the UK's research community's activities allowing them to define the scope of the discussion with publishers, *something like a researchers' "Brussels Declaration" remembering that Open Access is just the foundation for a new scholarship*.

Milestone 4: Develop added value services layered on repository content (e.g. tools for deposit, search, re-use, linking data, metadata enhancement, citation metrics, publication lists).

There is a need for an agency/lead body within the JISC HE community to address and **take a lead on legal issues** connected to copyright and publisher polices. Copyright and contractual issues with publishers need to be standardised. The current disparate approach makes ingesting of papers and data too complex and slow

Milestone 5: Establish an agency/lead body within the JISC HE community to take a lead on legal issues connected to copyright and publisher policies.

5.2 Research Data

5.2.1 Research Data: where do we want to be in 2013?

Providing access, sharing and re-use of data will enable 'new' research*. Physical and earth sciences will be working like the biology/bioinformatics communities, drawing on large datasets managed at community level.* Data will be made accessible to research groups for re-use by means of Web based 'research environments'.

A national strategy needs to be in place to ensure discovery, access, re-use and preservation of data. There are differing requirements and differing solutions for each of these processes that need to be taken into account. Key indicators will be put in place to measure progress in terms of data stored, also to measure the impact of sharing and re-use of data.

There will be a network of data centres for access to and preservation of research data working together at international, national, institutional and departmental levels. There will be considerable differences in the requirements and roles of different sized HE institutions. This network will be linked up with commercial sector and public sector data centres. Some of these data centres may use Web based facilities (e.g. Amazon S3). Transitory data stores will be available to support short term collaboration.

Data deposit will be built into the experimental research workflow. Deposit will be integrated into tools being used within labs. Repositories will be integrated into Open Science initiatives and will be used to make data available within collaborating groups.

There will be improved provision and quality of supporting information (metadata), but metadata will only be created where necessary. The future creation of metadata will take into account the rich versus light-weight metadata debate, resulting in various different levels of richness of metadata being created depending on the requirements for re-use and preservation.

There will be increasing demand for repositories to provide geodata as Web services or in more manageable ways/formats. This will enable users to create mash-ups and combine data more easily.

5.2.2 Research Data: where are we now?

We have only just scratched the surface. Technically goals are achievable but in every other respect there is a lot of ground to cover.

Certainly there is much activity by UK (and international) players as regards the curation and preservation of data. In the UK there are initiatives that include the JISC Digital Curation Centre, the Research Information Network, and the UK Research Data Service; internationally there is involvement from the Australian National Data Service, the US National Digital Information Infrastructure and Preservation Program, the Library of Congress, the National Science Foundation etc. As the infrastructure for research data is still in the planning stage, there are many lessons that can be learnt from the past experience of setting up repositories for research papers. It would be useful to clarify who is taking responsibility for inputting 'lessons learnt' from JISC programmes into the wider discussions on data repositories.

The licensing issues for geodata have not really been simplified in practice. The licensing issues are still the same big issue for geodata, standards still the same issue, although now with KML (Keyhole Markup Language) *becoming more prevalent as a data sharing format that may change.*

5.2.3 Research Data: how do we get from here to there?

There needs to be involvement of the 'repository community' with current research data initiatives to ensure policy issues learnt in establishing research output repositories are carried over to establishing research data repositories. Policies and processes for populating data repositories need to be influenced by the experience of existing institutional repositories. Existing JISC services or projects such as UKOLN or DCC might take a lead role here.

There is likely to be a hybrid of institutional, national and international research data centres as well as subject and disciplinary data centres. Deposit of legacy data needs to be considered; mandates need to be put in place and enforced; a framework of rewards and incentives established for sharing and re-use of 'my data'; *support of new grassroots researchers who are not yet established*, they are more likely to bring about cultural change

Milestone 6: Clarify responsibility for feeding existing repository implementation experience into current planning activity for research data centres. Potential candidates for this role are DCC, UKOLN and JISC. Taking experience of previous JISC programmes, and existing IRs on board, the interaction between different types of research data centre should be defined at an early stage. Work in this area needs to be undertaken in collaboration with the UK Research Councils. There is a need to establish metrics for populating research data centres and measure impact.

Meeting the objectives for discovery, access, re-use and preservation of data may require different solutions. Similarly different types of content will have different requirements (big science/small science; different disciplines).

Milestone 7: Ensure the multiple objectives that management and re-use of research data supports (e.g. discovery, access, re-use and preservation of data) are taken into account in proposed solutions. Ensure the requirements of different research data types (big science, small science, different disciplines) are taken into account in proposed solutions.

Most smaller institutions will not be able to sustain the costs of a digital data repository. Nor will they be able to ensure appropriate finding aids can find deposited material. Ingest of diverse idiosyncratic data requires specialised data archives at national level. There needs to be a national strategy involving the largest HE institutions and the national data archives.

Milestone 8: Formulate national strategy to take account of differing roles for different types of HE institution.

Already there is a lot of activity in this area, and it will increase over the next few years. Outcomes need to be targeted at the appropriate stakeholder groups on an ongoing basis. In particular what is relevant to the existing research repository manager? What is still being debated at the policy level? What can be acted on now?

Milestone 9: Target outcomes of repository activity at appropriate stakeholder groups.

There will be opportunities to incorporate training of personnel in data management into repository projects. *The barriers between domain experts, informaticians / data managers must be broken down with training and career incentives/opportunities put in place for work at this interface to begin.* Different communities need to be brought together to address the problem.

Milestone 10: Incorporate research data management training into repository projects.

ShareGeo will hopefully give us the opportunity to continue to educate the community on sharing and reusing geospatial data within current licensing and security models

5.3 Learning materials

5.3.1 Learning materials: where do we want to be in 2013?

Learning repositories should support better management and sharing of teaching and learning materials by individual lecturers. Feedback from consultation suggests learning material comprises two different sorts of content. Firstly a wealth of ‘day-to-day’ content produced by lecturers and uploaded to local institutional VLEs. Secondly quality assured material that is uploaded to repositories with more highly controlled policies.

Overall, there is agreement that it would be beneficial for closed content to be opened up. The existing national open access learning materials repository (Jorum) needs to be well populated with quality assured content. In parallel a UK network of institutional learning

materials repositories could be established with common interfaces to support sharing and re-use in institutional VLEs and Open Educational Resources initiatives. This would assist discipline based access to aggregated learning material content. Data flow needs to be established between institutional repositories and Jorum.

It would be useful to have QAA good management guidelines for learning materials. This would contribute to *establishing the role of the learning materials curator.* There is an ongoing debate in relation to quality assessment of learning materials, curators of learning materials need to address QA within their policies.

5.3.2 Learning materials: where are we now?

A baseline survey is needed of where we are now. In common with other repository content types, metrics for measuring progress need to be established.

There is a need to articulate the open access agenda to fit with learning and teaching terminology, currently the OA vocabulary does not fit. There is a need to investigate the requirements for different types of learning material repositories. A recent JISC study[40] reported on improving the evidence base in support of sharing learning materials, the intention being to develop business cases for sharing and re use of learning material, and to identify where there are gaps in this evidence base.

The relation between individual institutional repositories, as well as between IRs and national repository services such as Jorum and Intute is unclear. The recent JISC OER programme may address some of the issues around sharing and aggregation. Similarly the relation between learning materials repositories and recently emerging Open Educational Resources (OER) initiatives needs clarification. There has been a lot of OER activity internationally e.g. OpenLearn at the Open University, MIT OpenCourseWare, Rice University Connexions as well as activity in China and Japan. The background to OER and details of these recent initiatives are outlined in a CETIS briefing paper[41]. There needs to be consideration as to how repositories might support OER initiatives, and how 'legacy repositories' such as Jorum might be integrated with OER implementations.

There is little being done about discovery of learning materials. Learning materials repositories would benefit from guidelines on disclosing their content to search engines. Some OERs (but not all?) have open access policy enabling re-use see LabSpace[42] enabling re-use of OpenLearn material at the Open University.

5.3.3 Learning materials: how do we get from here to there?

Thinking around the role of repositories in relation to learning material is less mature than that around research outputs. There is an underlying need to examine the case for curation of learning materials, whether management of learning materials offers benefits at the institutional and national levels. There has been less repository related activity within the

---

[40] McGill, L.., Currier,S., Duncan, C., Douglas,P. *Good intentions: improving the evidence base in support of sharing learning materials*, December 2008. http://ie-repository.jisc.ac.uk/265/1/goodintentionspublic.pdf

[41] Yuan,L., MacNeill,S., Kraan,W. *Open Educational Resources – Opportunities and Challenges for Higher Education.* CETIS, 2008.

[42] http://labspace.open.ac.uk/

learning community. A number of milestones emerged in the Review discussion, all of which could be investigated in more detail.

Milestone 11:

- **Identify best practice for the management (including discovery, access and re-use) of learning materials at disciplinary, regional and national levels.**
- **Establish rewards and incentives for sharing.**
- **Set up measures for progress and impact.**
- **Integrate repositories with VLEs and e-portfolios.**
- **Work with QAA towards adherence to guidelines for good management of learning materials.**
- **Facilitate deposit of learning materials in repositories.**
- **Explore how different learning materials repositories might interact and in particular how Open Educational Resource initiatives fit with repository initiatives.**

6 Conclusion

Whilst progress has been made since the original Roadmap, not surprisingly feedback from the consultation concentrates on what still needs to be done. Various changes in emphasis are indicated: the focus should be on higher level objectives rather than on repositories; the technology approach should be more aligned with Web 2.0; repositories are part of a wider network of digital services, and their role within that network needs to be clarified. The Review highlights a number of challenges facing repository managers, some of which can be addressed with help from JISC as outlined by the recommendations. Others require support from institutional stakeholders.

Taking a step back, the JISC funded repository programmes have introduced a level of cultural change into a large number of HE institutions within the UK. A majority of HE institutions are now implementing repository systems, some with initial funding from JISC, others self-funding. All these institutions are likely to be actively involved with other JISC repository projects and JISC support activities e.g. the Repositories Support Project. This means JISC is connecting to a much wider set of institutions than previously, presenting an opportunity to broaden their area of influence.

Many repositories have been funded initially as projects and will now be maturing to move towards a sustainable service footing. For JISC this will be a test of their ‘project to service’ approach, a topic of much discussion and planning over the years. Much of the guidance and support JISC has developed for shared services and its other projects will be of relevance to institutional repositories as they establish themselves as ‘services’ within institutions.

From a technology perspective, repositories are facing the challenge of keeping up to date with new technology developments. In these days of rapid change, a repository becomes a ‘legacy’ system soon after set-up, relying on a software platform that will need to manage change whilst still supporting existing implementions. The need to align legacy technology with new Web developments is an ongoing challenge that faces many other information services (libraries, library utility systems, publishers) and it is an area that repository software developers will need to address.

Whilst good practice for managing repositories for research outputs is maturing, treatment of data and learning materials is less certain. Planning the management of these content types may well benefit from the experience so far. However it is clear that there are some significant differences in the overall aims and objectives for managing different content types. It is important that future planning is properly informed by previous repository experience.

This Review has led to discussion of the role of a Roadmap in relation to repositories. In an increasingly complex landscape it becomes more difficult to present a coherent roadmap covering the diverse number of repository types. As well, a roadmap needs to be integrated with strategic objectives and evaluation, in this case with the objectives and evaluation of the various JISC repository programmes. Rather than focusing on repositories the Review suggests that a more useful structure for the Roadmap would be based around higher level objectives, otherwise characterised as 'business functions'. In addition it may be that a small number of separate targeted milestones could be introduced based on specific metrics for different repository content types. These would provide evidence of value for money and would benefit both institutions and funders. Within the Review process there was widespread acknowledgement of the need to introduce both quantitative and qualitative metrics to demonstrate value, and that this would require consultation with the wider community. Discussion at the October 2008 JISC Repositories, Preservation and Asset Management Advisory meeting suggested that this Review might be used to move the Roadmap forward in these directions.