

Coming of Age?

CIBER, March 2012

*Strategic
Directions for
Digital
Repositories.
Charleston
Observatory
Study III*

Acknowledgements

Dr. Ian Rowlands (University of Leicester), who undertook much of the analysis and writing while a member of the CIBER team; Katrina Strauch and colleagues (Charleston Conference) for help in administering the library director survey and the sponsors, Institute of Physics Publishing, who also helped administer the researcher survey, Elsevier and Emerald.

Contents

Executive summary	4
1.0 Introduction	9
1.1 Aims and objectives	9
1.2 Scope and definitions	10
1.3 Methodology	11
1.4 Research context	13
2.0 Results	14
2.1 Library director study	14
2.1.1 Sample population	14
2.1.2 Type of repository	14
2.1.3 Size of repository	15
2.1.4 Budget	15
2.1.5 Type of content deposited	16
2.1.6 Quality of content	17
2.1.7 Versions of articles deposited	17
2.1.8 Subject repositories	17
2.1.9 Mandates	18
2.1.10 Copyrighted material	18
2.1.11 User assistance offered	19
2.1.12 Goals and achievements	19
2.1.13 Impact of repository	21
2.1.14 Critical success factors	21
2.1.15 Advantages, disadvantages and perceptions	23
2.1.16 The future for repositories	27
2.1.17 Gold open access: who pays?	28
2.1.18 Free text comments	29
2.2 Researcher study	30
2.2.1 Depositing to repositories	30
2.2.2 Reasons for deposit	31
2.2.3 Type of repository	31
2.2.4 Type of content deposited	32
2.2.5 Website availability	33
2.2.6 Type of repository preference	34
2.2.7 Repositories as sources of information	34
2.2.8 Relative repository performance	35

2.2.9 Advantages and disadvantages of repositories	36
2.2.10 Free text comments	38
3.0 Discussion	40
Appendix 1 Free text comments from library directors	44
Appendix 2 Free text comments from researchers	47

Executive Summary

Digital repositories (whether institutional, subject or format¹ based) have been with us for more than a decade and the study reported here sets out to find how far they have come in that time. Our aim was to take stock of the achievements of digital repositories; find out what they looked like, what their impacts were, where they were going, what their users thought of them and how successful they were thought to be. We especially wanted to discover how they are dealing with emergent issues, such as data, and the extent to which they were facilitating scholarly communication and research itself.

In general terms we know there is growth in the open access availability of research publications, both of gold (author pays for publication) and green (self-archiving by the researcher). For example, approximately 30% of all articles are claimed to be available in open access, two thirds in green and one third gold. These figures are not definite and various pundits give different estimates but they provide us with some context. This study focuses on the green open access movement, and the archives or digital repositories established mainly by libraries to store and make green publications of all kinds accessible. According to OpenDOAR data (1/3/2012) there are 2,173 repositories worldwide, 82% are institutional and 11% are subject based.

The study surveyed both library directors of digital repositories and individual researchers, and though both reflect the demand side for content arising out of scientific research, the relatively low numbers surveyed means the results need to be treated with a degree of caution. Particularly in the context of the millions of researchers active worldwide and the several thousands of repositories serving their needs. The 100 plus responses received from library directors cover a wide spread of interests, whereas the user survey (1600 plus respondents) was much narrower in scope, focused as it was on physical scientists.

Findings

There is paucity of robust research on digital repositories, especially on their use and performance, and for this reason we make no apology for furnishing our data in full, statistically verified where appropriate and supported by many quotes. The Executive summary provides the quick overview.

Library director survey

Most responses came from directors of institutional repositories so the findings tend to focus on that type of repository.

- **Size of repository.** The resources associated with repositories are relatively modest. Two-thirds of repositories can be categorized as 'small', having only one or two people working on

¹ Based on one type of content, like research data or e-theses)

them. The implications would appear to be that there has not been a major swing in resourcing away from collection development to the institutional repository as being a way forward for libraries.

- **Budget.** Repositories are mainly funded within the library. Recurrent spending on digital repositories is minuscule, averaging only 1.8% per cent of library operational budgets.
- **Type of content deposited.** Research outputs are very much more diverse than just journal articles and it is a wrong for publishers to view repositories as just collections of journal articles. More than two dozen forms of content are collected by repositories in the survey. Journal articles are however the most common content and they were followed by conference proceedings and dissertations.
- **Quality of content.** Librarians are not pro-active in assessing the quality of inputs to repositories and there is a clear presumption that if it comes from staff and students of the institution it must be OK. This is the 'provostial assessment' which elite universities have been able to adopt by capitalizing on the quality of their researchers. Nearly 70% of repositories said that they accepted everything without review.
- **Versions of articles deposited.** Many repositories will take any version of an article but there is a preference for later versions. Repositories are more concerned with getting content than worrying about what version it is.
- **Subject repositories.** A large minority (38%) of repositories are putting materials in their own repositories as well as in subject-based ones, and thus actively supporting both models.
- **Mandates.** Almost half of institutions either have or are planning a mandate. However, mandates take a number of forms, and the severity of a local mandate is tempered by the attitudes of the researchers on campus.
- **Copyrighted material.** A substantial amount of effort is spent obtaining permissions before placing articles in repositories. However, often this was thought to be the responsibility of the author rather than the library. Free text comments received reflect a general unease about whether articles can be included within the institutional repository without having first gained publisher permission, irrespective of the stage the manuscript is in at the time.
- **User assistance offered.** Most repositories provide a range of services to their users, with assistance with deposit and metadata creation and management being the most widely offered, but of course most repositories are run on a shoe string so this limits help.
- **Goals and achievements.** Three goals were thought to be very important – 1) to make the formal literature more openly available; 2) to create a research shop window for the institution; and 3) to ensure the long term preservation of material. The goal of long term preservation and curation is attributed significantly more importance in North America.
- **Impact of the repository.** Claims were relatively modest in this regard, with small gains claiming to be made in raising the profile of the library within the institution.
- **Critical success factors.** To get people to deposit and use the deposited material are the main ones. You can mandate people to deposit publications but you cannot make them use deposited material.

- **Priorities to be addressed by repositories.** Low awareness, not enough content and too little use were the main priorities. You could also add more resources to this list.
- **Advantages of having repositories.** There are many, with providing access to publicly funded research and the long term preservation of institutional materials the most important of them. Contributing to the reform of scholarly communication and publishing, came a not so distant sixth.
- **Disadvantages of having repositories.** Copyright is confirmed as the main headache for directors, followed by lack of awareness and funding. European directors rate funding issues significantly less important than North America.
- **Perceptions and attitudes.** Library directors definitely do not see repositories as impacting negatively on publishers' revenues: rather the opposite, they feel that repositories help to raise the digital visibility of research materials.
- **The future for repositories.** Not unnaturally perhaps – after all institutions have made a considerable emotional investment in IRs – hardly anyone thought that repositories would be less important. The vast majority felt that they would become more important.
- **Gold open access – who pays?** In most institutions, money for author pays open access publishing is unavailable locally and so authors need to turn to their funder or some other source.

Researcher survey

- **Depositing.** Two thirds of researchers said they had deposited material in a repository. The high deposit rates are probably down to the high proportion of physical scientists in the sample. Of those that did not deposit the main reason given was that they were unaware that there were repositories.
- **Reasons for deposit.** Voluntary deposit was the main reason, and this was followed by mandates. Responses suggest that institutional funding mandates and peer pressure are extremely helpful in terms of driving depositing behaviour. There are age-related differences in terms of responding to an institutional mandate, with those researchers under 30 being more than twice as likely to comply as those aged 50 or above. The demographics could therefore work to the advantage of a mandated repository system in future.
- **Type of repository.** More respondents say they have deposited in an institutional rather than a subject repository, despite the iconic status of the arXiv for the physical sciences. There are strong age-related effects here: the Under 30s are 34 per cent more likely than the 50 and older age group to have deposited in an institutional repository and 31 per cent less likely to have deposited in a subject repository. Younger researchers are taking institutional repositories (and the mandates that underpin them) very seriously.
- **Type of content deposited.** Journal articles and e-theses are the types of document most deposited, accounting for 84 per cent of all deposits. Journal articles are most commonly deposited at Stage I (the author's manuscript before peer review). Articles at Stages II (after peer review but before publication) and Stage III (publishers' final version) seem almost equally acceptable. These findings probably confirm the relatively weak importance attaching

to peer review in the physical sciences. There are large and significant age-related differences here. Older (50 plus) researchers are 50 per cent more likely than the youngest researchers to deposit a Stage I manuscript and 73 per cent less likely to deposit a Stage II manuscript. The implication is that the stamp of peer review is much more important for younger researchers who perhaps have less confidence depositing materials that have not been reviewed

- **Websites.** Three-quarters of depositors make their articles available on their own or departmental websites, but only about 27% make all of them available.
- **Repository preference.** The overall preference is to deposit via a more formal route: a subject or institutional repository. Over 60% did so, and the subject repository was the preferred location. Note: the findings are at odds with the findings in the Type of Repository bullet above, but here we are asking about preference and above about practice.
- **Relative repository performance.** Asked to compare repositories with the best information services they knew researchers thought the quality of content, speed of response and ease of use to be very good. Breadth, depth and ease of navigation are not so highly rated.
- **Advantages of having repositories.** Researchers view the opening out of access to their research materials as the greatest single advantage offered by repositories, and this was followed by reducing the time between production and 'publication'. Young researchers were more inclined to the belief that repositories "contribute to the reform of scholarly communication and publishing".
- **Disadvantages of having repositories.** The two main gripes are the variable quality of materials in repositories and insecurity over their long term viability. Younger researchers are much more concerned about the potential for confusion caused by different versions of the same material being in circulation and they are also more likely to feel that repositories add to a growing fragmentation of the literature and that copyright issues create confusion and uncertainty.
- **Perceptions.** Researchers were asked to respond to a series of propositions and the one they most agreed with was that repositories might be beneficial to publishers by virtue of creating greater digital visibility for research materials. However younger researchers were much more likely to agree with this statement than the older age groups. The main issue though that divides the generations is a tendency for younger researchers to agree with the proposition that Gold OA will gradually make digital repositories redundant.
- **Free-text.** When liberated from asking our questions what did the researchers bring up? Most of the comments could be loosely described as pro-repositories, with the French and the developing nations particularly keen on them. A good number were also anti publisher, quite passionately so in a number of cases. Some however saw a role for both. Not surprisingly the topics of peer review came up quite frequently, some arguing that the existing system was no longer fit for purpose. There were a good number of requests for repositories to collect data and, surprisingly perhaps, libraries took a few knocks.

Conclusions

Differences between responses of library directors and researchers

- Library directors, probably inevitably, seem to be looking at digital repositories as a process whereas the researchers are seeing them as a tool. Librarians were operationally focused; concerned about how the repository can provide the library with a valuable shop window, how they can deal with curation, and how they get researchers to submit and use. Researchers focused on the challenges the repositories posed to their research experience: uncertainty over copyright, what stage in the article to deposit, how to deposit, and what impact issues such as gold will have featured highly.
- Researchers gave greater preference to depositing their material in a subject-based repository (which can be explained by the heavy physics focus, which has arXiv as the main repository) and as a result most respondents use repositories to find information. It is unclear how much of this is attributable to arXiv and how much to the state of local institutional repositories, but it does give a lie to many publishers' assumptions that repositories are not used.
- As regards the future there is a stronger feeling among researchers that 'gold' open access will replace 'green', and that within the repository movement itself, subject based repositories are better placed than institutional repositories to meet their needs. By contrast, and inevitably, library directors give a higher rating to institutional repositories.

Implications for publishers

The study contained some important findings for publishers: a) the resources associated with repositories are very modest indeed; b) repositories are not just as a second rate collection of journal articles, collection policies and ambitions are much wider than this; c) repositories are not thought, by library directors, to herald a major reform of scholarly communication and publishing system; d) library directors generally think that the increased digital visibility raised as a consequence of content being accessible within repositories will benefit publishers in terms of usage; e) free text feedback from researchers show that there is a smoldering resentment about the inequity of the existing publisher-controlled system among a minority of the research community.

Coming of age?

Not quite; digital repositories are still in a formative stage – not yet come of age, still aspirational. They are not totally accepted as a cultural feature of the scholarly communication landscape but there are drivers in place which are moving them towards early adulthood. Indeed the big surprise is that they have moved so far on such limited resources

1.0 Introduction

This study of digital repositories is the third in a series of research projects produced by the Charleston Observatory, the research arm of the annual Charleston Library Conference². The Observatory, established in 2009, is a mechanism by which important topics raised at the Charleston Conference can be researched and the results reported back to conference to provide a feedback loop. It is a virtual space where evidence can be collected globally in a robust manner and where all the key information stakeholders (librarians, publishers, subscription agents and academics) can come together and share data for the benefit of all. The Observatory's first project (2009) looked at the impact of the world-wide recession on libraries. Last year, the Observatory considered social media and how they are impacting on research practice. Both studies received widespread press coverage and generated a series of reports and peer-reviewed publications³.

According to the JISC '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. They are commonly used for open access research outputs'.

Digital repositories have been with us for more than a decade and have become an established component in an increasingly complex scholarly communications landscape. The digital and institutional repository community has big and diverse ambitions and sometimes finds itself in dispute with publishers. Anecdote has it that repositories are beginning to bear fruit. So it was thought timely to make them a subject of some robust and independent research and see whether they had become strategic to scholarly communication and scientific research.

Thanks to the Institute of Physics Publishing, Elsevier and Emerald for their sponsorship of the research.

1.1 Aims and objectives

The broad aim of the study was to take the pulse of digital repositories and establish how far they have come, what had they achieved and where were they going next. Also how far were they meeting user needs and expectations? We particularly wanted to see how they were dealing with emerging issues, such as datasets and supplementary non-textual information and the extent to

² <http://www.katina.info/conference/>

³ http://ciber-research.eu/CIBER_news-201112.html

which they were facilitating scholarly communication now and whether, in future, they would be central to its development.

The study was conducted in two parts. The major part of the study was an investigation of the views of library directors on digital repositories; we targeted directors because we wanted to learn about the big picture, ambitions, policies and the strategies. The second part was a study of how scientific researchers used and viewed digital repositories.

The objectives of the library director study were to:

- Understand what library directors saw as the goals of digital repositories;
- Identify the critical success factors behind successful digital repositories;
- Assess the wider impact of digital repositories.

The objective of the researcher survey was to obtain a demand-side response to the key questions that we had asked the library directors. In particular we were interested in: 1) what they saw - in general terms - as the main advantages and disadvantages of digital repositories; 2) whether they agreed or disagreed with a series of provocative statements about digital repositories; 3) whether they thought institutional and subject repositories would be more or less important to researchers in three years' time.

1.2 Scope and definitions

Digital repositories were defined broadly to include:

- Institutional repositories which aim collect widely across a particular university or similar institution, possibly in a wide range of formats.
- Subject repositories based on collecting only within a certain discipline, probably across more than one institution.
- Format repositories whose scope is limited by collecting in a particular format, perhaps student dissertations and theses, or research data.

In addition, researchers were also asked about their use of personal web pages to store and make openly available their own publications, a function similar in some respects to that of institutional repositories.

While the library directors had no problems with understanding the scope of the study a few problems emerged with users. Despite the fact that during the survey we regularly reminded users of the scope and the definitions there were the inevitable confusions that arise from an international and interdisciplinary study of the digital scholarly landscape.

Thus one user commented: "I maintain a digital repository of research results and I am involved with some others maintained locally by professional organizations I am involved with. But the survey appears to be about something else - something I probably have never used, and hence I have no opinion on many of these things". Where we believe such comments have impacted

significantly on the answers to our questions we mention this in the text.

1.3 Methodology

Library director study

To provide the necessary reach and international coverage, as in the two previous Charleston Observatory studies, online questionnaires were the primary means of collecting data. Invitations were sent out to relevant individuals and organisations to participate in the survey and a link provided to the questionnaire via:

- Charleston Conference email lists – direct mail (North America biased)
- Home grown lists compiled by CIBER researchers going through OpenDOAR and providers' website – direct mail (necessarily incomplete because of time pressures and focusing on Europe and the rest of the world)
- Sconul list-sconul list – direct mail (UK-focused)
- LIBER list-serv direct mail (EU-focused)
- UKSG e-resources list – direct mail
- Emerald – direct mail to library directors list
- Special Libraries Association – blog entry
- Canadian Library Association – list serv direct mail
- North American Serials Group – list serv direct mail
- Oberlin Group – direct mail

The OpenDOAR database lists 2,173 repositories using the above definitions. It is difficult to know how comprehensive this source is, but it seems to be the database with the broadest coverage. If we assume that is the population from which we are drawing, and we have completed responses for 106 institutions, then the figures in this report should be interpreted with error bars of plus or minus 9.3% (at the 95% confidence level). Therefore the data needs to be treated with caution and any interpretation must be broad stroke. Later tables split out the findings by size of repository and by region. Of these, size is the more reliable.

Calculating a response rate is very difficult because there must have been extensive cross-posting, and we did not know in advance which libraries have a repository (those that have not are very likely not to take part). On the plus side, the sheer diversity of lists used means that any bias (geographical, subjects, size) has been minimized. If we had to make a guess, we would say the response rate was probably of the order of 2 per cent. Given that the survey was largely distributed through creditable third parties (e.g. Sconul) one can only assume that the survey either failed to ignite the imagination, or that directors see repositories as only a minor part of their business (borne out by our findings on staffing numbers) and not the big beast that some publishers fear. Or perhaps they are simply embarrassed at the relative unsophistication of their efforts (again something borne out by our findings).

We have indicated any 'statistically significant' differences by size of repository or by region in

some of the tables that follow. This should be taken with a pinch of salt and it is probably only worth commenting on those with the largest differences (ANOVA F of 4 or greater⁴) given the limitations of the sample size. Although significances less than this are worth following up in focus groups or interviews.

The vast majority (91%) of respondents came from universities or colleges. North America was the best represented region of the world (47%); Europe accounted for 42% of respondents and the Rest of the World 11%. The USA was the best represented country (43%) and was followed in rank by the UK (18%), Canada (3%), Netherlands (3%) and France (3%).

Researcher study

This part of the survey was distributed across lists owned by the Institute of Physics Publishing to 85,000 email addresses between 19 and 31 December 2011. Fully useable completions totaled 1,685, a response rate of just below 2 per cent. This is low by industry standards and the results should be interpreted with this in mind: self-selection bias is a major issue in surveys of this kind. Given the size of the sample, the figures in the report should be interpreted with error bars of plus or minus 2.4 per percentage points at the 95 % confidence level: so a value of 50 per cent should be interpreted as lying in the range 47.6-52.4.

Responding researchers were quite widely dispersed geographically; nearly a 100 countries featured. The US was best represented with 16% of respondents; it was followed by China (10%); Italy, India and Germany (all 6%); and Russia and the UK (both 5%). Fifty-five percent of respondents were academic faculty, 32% salaried researchers and the rest students (mainly PhD students). In regards to age 15% were in their twenties; 36% in their thirties; 22% in their forties; 15% in their fifties and 12% were older than that. Eighty percent worked in the academic sector and another 10% for Government. Although the sampling frame was that of a physics membership institution in fact our respondents came from a surprisingly wide range of subject fields, albeit heavily scientific:

	n	Column%
Physical sciences	1,012	59.7
Engineering and technology	382	22.5
Mathematical and computer sciences	187	11.0
Biomedical and life sciences	93	5.5
Education and other subjects	21	1.2
Total	1,695	100.0

⁴ ANOVA = Two-way Analysis of Variance

We cannot claim, of course, that a sample dominated so much by physicists constitutes a representative group of researchers, as physicists have gone the furthest in terms of embracing subject repositories (notably the arXiv but also the Astrophysics Data Service). For this reason they are likely to be very pro open access but possibly more anti local institutional repository solutions than most others. Unlike biological scientists, they probably care less about formal peer review, so again this predisposes them to green open access. However, what they do give is a very knowledgeable user group against which we can measure the claims and statements of library directors, and this triangulation was important for the purposes of this study.

1.4 Research context

There is a large literature on the mechanics and practices of institutional repositories and a healthy literature on researchers and their academic reading habits, but nothing really substantial or robust on how they are doing and used. As far as the published views of library directors or repository managers– and there are plenty – they are generally anecdotal and usually plugging a particular (often local) agenda. There is nothing that we are aware of which compares with the triangulation or the global or the independent approach of this Charleston Observatory study.

2.0 Results

2.1 Library director study

The preliminary results of the library director study were presented at the 2011 Charleston Conference⁵ and the PowerPoint is available on the CIBER site⁶. Here we provide a more comprehensive and detailed analysis of the study.

2.1.1 Sample population

One hundred and fifty two organisations responded to the survey and with nearly 70% of respondents saying that they had implemented a digital repository and another 23% saying they are planning one, repositories are clearly very much a feature of the modern day library. Clearly we need to treat this finding with caution as it is quite likely that institutions with, or planning, a repository were more likely to respond to the questionnaire on the basis they had something to say or prove. Of the 7% (11) of institutions, mainly small US liberal arts colleges, that provided an explanation of why they were not planning to have a repository nearly half said it was down to limited resources and just over a quarter gave not being a highly research intensive institution as the reason. The other main reasons were that they were unconvinced of the benefits or lack of management support. For one institution there were several, quite complex, reasons why they did not have a repository:

“A combination of reasons really - 1) the institution operates a tight research performance management system for academics linked to publication in an official list of peer-reviewed journals - we would need to be able to demonstrate robustly that an open access repository would not undermine this system. 2) we have not had the staff capacity to deliver a repository in Library or IT in the past 2.5 years. 3) we have not yet tested the water internally as to whether there would be acceptance that open access could meet some of the School's strategic objectives” [UK university].

2.1.2 Type of repository

The analysis of type of repository should be treated with caution. Many directors ticked ‘other’ on the questionnaire and the data contained in Table 1 are our interpretation of their answers, modified as appropriate. The categories ‘institutional’ and ‘format-based’ are difficult to isolate clearly and it is possibly best not to make too much of the results. It is certainly not a useful categorization for splitting out the data. For the record three quarters of the repositories were classed as being institutional, very few were subject based. This is in line with global practice, where there are far more institutional repositories than there are subject-based ones.

⁵ <http://www.katina.info/conference/>

⁶ http://ciber-research.eu/CIBER_news-201112.html

Table 1: Which of the following best describes the main collection focus of your repository?*(n=106)**Numbers and column percentages (tick one only)*

	n	%
Institutional repository	81	76.4
Format-based repository	21	19.8
Subject-based repository	4	3.8
Total	106	100.0

2.1.3 Size of repository

The resources associated with repositories are relatively modest (something publishers will find very encouraging). For the sake of breaking out subsequent findings and tables/figures in a more interesting way, we have created a dummy variable (small or medium to large) on the basis of the data below (Table 2). Two-thirds of repositories could be categorized as ‘small’, having only one or two people working on them. At the other end of the scale just 7.5% had five or more people working on them. Of course it is difficult to know how fulltime these staff really are on repository issues, but the implications are that there has not been a major swing in resourcing away from collection development into the institutional repository as being a way forward for libraries.

Table 2: How many staff work on your repository? (n=106)*Numbers and column percentages (tick one only)*

	n	%
<i>Small repositories</i>		
Less than 1 FTE	28	26.4
1 to 2 FTE	39	36.8
<i>Medium to large repositories</i>		
3 to 4 FTE	31	29.2
5 or more FTE	8	7.5
Total	106	100.0

2.1.4 Budget

Library directors were asked questions about the size of their overall library budget and specifically how much was devoted to the repository. Answers were sparse in terms of detail and some of it clearly unreliable because many of the respondents misinterpreted the question. However, we did ask people to give a rough indication of how the repository was funded. This data appears reasonably unproblematic and the percentage breakdown is provided in Table 3.

This is a collective average of course and does not mean that all repositories are funded in the same way. The main message is that they are mainly funded within the library. Recurrent spending on digital repositories is relatively low, averaging only 1.8% per cent of library operational budgets.

Table 3: How is your repository funded? Overall analysis (n=98)*Budget breakdown as column percentages*

	%
Regular budget line for your institution’s library	71.0
Special initiative supported by your institution	15.3
Departmental budgets	6.6
Grant from an external source	4.3

Contributory or membership basis	2.5
Publication budget	0.4
Total	100.0

2.1.5 Type of content deposited

Research outputs are very much more diverse than just journal articles and it is a fallacy for publishers in particular to just see repositories just as collection of journal articles. As Table 4 demonstrates repositories are Hoovering up a lot of different kinds of materials, reflecting a more object-orientated approach to research outputs on the part of many academics. In fact coverage is even wider than that shown in the Table with blog posts, interview notes, student magazines and scanned herbarium materials also being mentioned by respondents. Having said that, however, journal articles are very important, as are theses and special collections (Table 5). Table 4 is shorter than 5 because many of the content forms above were considered 'the most important' by the respondent, whereas others were just ephemeral or specialized to meet local circumstances.

Table 4: Which types of content are in your repository? (n=98)
Numbers and column percentages (tick as many as apply)

	n	%
Journal articles	85	86.7
Conference papers	83	84.7
Student dissertations and theses	80	81.6
Book chapters	70	71.4
Working papers	66	67.3
Books or monographs	59	60.2
Technical reports	57	58.2
Images or photographs	55	56.1
Special collections (e.g. digitized copies of rare books, archives, etc.)	41	41.8
Metadata-only records	38	38.8
Sound recordings	36	36.7
Video recordings	30	30.6
Annual reports	26	26.5
Research datasets	23	23.5
Administrative / corporate records	21	21.4
Learning objects	17	17.3
News or press coverage	14	14.3
Patents	12	12.2
Computer software	10	10.2
Total	98	100.0

The numbers sum to more than 100% since people could choose more than one format.

Table 5: Which do you consider to be the *single most important* content type for your repository? (n=97)
Numbers and column percentages (tick one only)

	n	%
Journal articles	42	43.3
Student dissertations and theses	32	33.0
Special collections (e.g. digitized copies of rare books, archives, etc.)	10	10.3
Working papers	4	4.1
Research datasets	3	3.1
Books or monographs	2	2.1
Images or photographs	2	2.1
Learning objects	1	1.0
Metadata-only records	1	1.0
Total	97	100.0

2.1.6 Quality of content

Directors were asked how they maintained the quality of content in the repository, clearly something of a concern to both depositors and users. The answer appears to be not much; librarians are not pro-active in this area and there is a clear presumption that if it comes from staff and students of the institution it must be OK. This is the ‘provostial assessment’ which some of the elite universities have been able to adopt by capitalizing on the quality of their researchers. . Thus nearly 70% said that they accepted everything without review. Of the rest nearly 16% said there was a post acceptance review or rating and nearly 15% said they used peers to review quality In addition, some academic communities establish their own criteria for what is acceptable to deposit.

The following analyses (2.1.7 – 2.1.92) are related to repositories containing journal articles only

2.1.7 Versions of articles deposited

This analysis is restricted to those respondents with a repository that ingests journal articles. Table 6 shows that many repositories will take any version of an article but there is a preference for later versions. Repositories do not seem to care too much about the stage in the article’s progression to article of record.

Table 6: Which of the following versions of journal articles do you allow to be deposited? (n=85)
Numbers and column percentages (tick as many as apply - only repositories that manage journal articles)

	n	%
Author MS before peer review	24	28.2
Author MS after peer review	21	24.7
Accepted MS before publication	38	44.7
Final published version	36	42.4
Total	85	100.0

2.1.8 Subject repositories

Table 7 is interesting because it suggests a reasonably large minority (38%) of repositories are actively supporting submission to remote subject-based repositories in addition to their own repositories.

Table 7: Does your institution actively support submission to remote subject-based repositories (such as PubMed Central, arXiv, RePEc, etc.)? (n=85)

Numbers and column percentages (tick one - only repositories that manage journal articles)

	n	%
Yes	32	37.6
No	34	40.0
I’m not sure	19	22.4
Total	85	100.0

2.1.9 Mandates

Whatever their current effectiveness, it is not enough because around half of institutions either have or are planning a mandate (Table 8). Mandates take a number of forms, but the severity of a local mandate to deposit in the institutional repository is tempered by the attitudes of the researchers on campus. Many mandates are therefore watered down in terms of the sanctions given for non-compliance, but may still be classed as a ‘mandate’. In fact there are probably hardly any (there is said to be one in Australia but we know of no others) really binding mandates. Almost all of them have a proviso relating to whether the researcher can deposit under the terms of the agreement you have with your publisher. Agreements, like that at Harvard, often say that you can opt out if you write in and say you are doing so, and in no case that we know of is there any attempt to force researchers to comply. One university librarian said he sees mandates as being educational. He thinks that scholars will get used to depositing gradually but that if he tried to force them he would come up against academic freedom questions.

Table 8: Does your institution operate a mandate requiring research publications from your institution to be deposited in your repository? (n=85)

Numbers and column percentages (tick one - only repositories that manage journal articles)

	n	%
Yes	28	32.9
No, but there are plans to introduce one	13	15.3
No	44	51.8
I’m not sure	0	0.0
Total	85	100.0

2.1.10 Copyrighted material

This is another analysis for which we had to simplify a lot of complicated responses. The basic message coming out of Table 9 is that a substantial amount of effort is spent obtaining permissions before placing articles in repositories. This needs qualification however, as a lot of the comments suggested that this was the responsibility of the author rather than the library, or that deposits were checked against (for instance) the Sherpa-ROMEO lists.

Table 9: What is your policy for including copyright-protected materials? (n=85)

Numbers and column percentages (tick as many as apply - only repositories that manage journal articles)

	n	%
Obtain permission	68	80.0
Accept preprints only	26	30.6
Include but access is password controlled	24	28.2
No items of this type are added to the repository	13	15.3
Total		100.0

Free text responses to the ‘other’ category has been allocated to the broad categories above, but here we give a few quotes:

“Ask depositors for assurances that they have obtained copyright permission before deposit.”

“To date most content is public domain or deposited with permission to provide access; in future we will need to move into deposit of content which will require more complex rights management”

“We have different sets of policies for different sets of materials; however, for the most part, onus is placed on the depositor to get permissions if they are depositing material for which they do not hold copyright.”

“Obtain permission, but I not sure”

“We check Publisher's policies regarding OA”

“We add a citation and link to the full version (if held in a licensed database)”

“Use Sherpa Romeo to determine rights and work with authors to include what we can.”

These comments reflect a general unease about whether articles can be included within the institutional repository without having first gained publisher permission, irrespective of the stage the manuscript is in at the time. They probably reflect the natural wish of librarians to observe the law (a result of professional training).

2.1.11 User assistance offered

Most repositories offer a range of services from assistance with deposit to personalized webpage (Table 10).

Table 10: What services do you offer to your users? (n=106)
Numbers and column percentages (tick as many as apply)

	n	%
Assistance with deposit	93	87.7
Assistance with metadata creation and management	88	83.0
Assistance with copyright clearance	66	62.3
Download statistics	59	55.7
Personalised web page or cv	33	31.1
Total	106	100.0

It does appear that library directors see a role for themselves in providing assistance with the deposition of articles, though this still seems still to be undertaken on a shoestring (see staffing allocation above). It is almost as if library directors are waiting to see at what speed digital repositories take off before committing themselves wholeheartedly to what is still an unclear phenomenon.

Our analysis now moves away from operational to strategic concerns.

2.1.12 Goals and achievements

These are responses from those institutions that have a repository already and show what they think are the key purposes and the extent to which they feel that repositories have delivered on those purposes. Respondents rated three goals as very important – 1) to make the formal literature more openly available; 2) to create a research shop window; and 3) to ensure long term preservation of material (Table 11). Interestingly, they also felt that they were most successful in achieving another goal, rated fourth in importance terms, helping change the culture of the library.

Table 11: To what extent do you judge the following goals important, and to what extent have they been achieved for your repository? Overall findings (n=106)

Mean ratings on a scale where 0=Not at all important (or successful) and 3=Very important (or successful)

	IMPORTANCE	SUCCESS
To make the formal literature more openly available	2.38	1.43
To provide a research shop window for your institution	2.34	1.53
Long term preservation and curation	2.33	1.45
To help change the culture of the library	2.03	1.58
To manage informal literature better	1.49	1.14
To provide management information on research productivity	1.48	1.06
To provide access to rare materials	1.38	1.07
First steps to becoming a digital publisher	1.01	0.94

Breaking this analysis down, by size of repository makes no differences to any of the answers provides above.

There are however some regional differences (Table 12) and the main ones are:

- The goal of long term preservation and curation is attributed significantly more importance in North America (ANOVA $F=4.38$, $p<5\%$), although, interestingly, there were no difference in reported success by region. We would imagine that the concern about archiving and preservation stands alone and is not attached to a particular type of content because there is no need to worry about it in the case of journal articles – publishers do a very good job.
- The goal of providing a research shop window is rated as significantly less important in North America (ANOVA $F=4.89$, $p<1\%$) and North American library directors also reported significantly less success in achieving this goal (ANOVA $F=10.76$, $p<1\%$).
- The goal of making the formal literature more openly available is rated as significantly less important in North America (ANOVA $F=8.29$, $p<1\%$). However, interestingly there are no significant differences in reported success by region.
- The goal of providing access to rare materials, which is generally rated fairly lowly, is rated as significantly more important in Europe and North American than the rest of the world (ANOVA $F=9.11$, $p<1\%$). There is no difference in reported success by region.
- The generally very lowly rated goal of providing management information on research productivity is rated as significantly more important in North America (ANOVA $F=9.96$, $p<1\%$) and North American library directors also reported significantly less success in achieving this goal(ANOVA $F=11.56$, $p<1\%$).

The result needs to be treated with care because ‘significant’ just means statistically unlikely purely by chance and this does not imply that the difference is real or meaningful, especially when the absolute numbers are small. There are also the sampling issues, noted earlier, that mean any regional differences should be treated carefully.

Table 12: To what extent do you judge the following goals important, and to what extent have they been achieved for your repository? Analysis by region (n=106)

Mean ratings on a scale where 0=Not at all important (or successful) and 3=Very important (or successful)

	IMPORTANCE			SUCCESS		
	North America	Europe	Rest of the world	North America	Europe	Rest of the world
Long term preservation and curation	2.59	2.14	2.18	1.43	1.43	1.64
To help change the culture of the library	2.14	1.92	2.09	1.50	1.65	1.64
To provide a research shop window for your institution	2.02	2.55	2.64	1.09	1.76	2.18
To make the formal literature more openly available	2.02	2.65	2.55	1.23	1.55	1.73
To provide access to rare materials	1.93	0.96	1.09	1.25	1.00	0.64
To manage informal literature better	1.64	1.39	1.36	0.95	1.25	1.36
First steps to becoming a digital publisher	1.20	0.84	1.00	1.09	0.86	0.73
To provide management information on research productivity	0.95	1.82	2.00	0.61	1.31	1.64

2.1.13 Impact of repository

We asked respondents to self-evaluate what level of impact they feel they have achieved. The consensus, given the scale used is somewhere between 1 and 2, between something ‘small’ and something ‘significant’. So no one is getting carried away with the success of repositories.

Table 13: To what extent do you agree or disagree with the following statements regarding the impact of your repository? Overall findings (n=106)

Mean ratings, where 0=No impact and 3=Highly significant impact

	Overall rating
It has raised the profile of the library within my institution	1.67
It has raised the profile of the library outside my institution	1.47

No difference by region, but size makes a significant (ANOVA F=3.91, p <1% within, and ANOVA F=6.42, p<1% outside) difference.

However, as Table 14 shows, the larger the repository, the greater the perceived impact, with larger institutions claiming a greater impact.

Table 14: To what extent do you agree or disagree with the following statements regarding the impact of your repository? Analysis by size of repository (n=106)

Mean ratings, where 0=No impact and 3=Highly significant impact

	Small	Medium to large
It has raised the profile of the library within my institution	1.55	1.87
It has raised the profile of the library outside my institution	1.31	1.74

Size makes a significant difference (ANOVA F=3.91, p <1% within, and ANOVA F=6.42, p<1% outside).

2.1.14 Critical success factors

The next question asked people to rank (rather than rate) six critical success factors for a digital repository. There are clearly problems attracting material and users, which are clearly fundamental. You can mandate people to deposit publications but you cannot make them use deposited material. Shades of taking a horse to water! The size of the depository is not a significant factor.

Table 15: What do you consider are the most important critical success factors for a digital repository? Overall findings (n=106)

Mean rankings, where 0=Least important and 5=Most important

	Overall ranking
People should want to put material in	4.19
People should want to get material out	3.15
Its scope and extent should be clear	2.75
It should have a strong sense of ownership	1.77
It should have a clear, single purpose	1.73
It should serve a clear community	1.41

There really is not much in the way of regional differentiation either, the rankings are roughly the same and only show anything substantial on one item. The only significant difference is that for 'people should want to get material out' (ANOVA F=3.17, p<5%) where the rest of the world is very much out of line.

Table 16: What do you consider are the most important critical success factors for a digital repository? Analysis by region (n=106)

Mean rankings, where 0=Least important and 5=Most important

	North America	Europe	Rest of the world
People should want to put material in	4.07	4.24	4.45
People should want to get material out	3.18	3.39	2.00
Its scope and extent should be clear	2.77	2.65	3.18
It should have a strong sense of ownership	1.89	1.76	1.36
It should have a clear, single purpose	1.61	1.71	2.27
It should serve a clear community	1.48	1.27	1.73

The next Table (17), which focuses on the priorities for repository managers, strikes a rather desperate note: low awareness, not enough content and too little use. You could also add more resources needed.

Table 17: What are your priorities for the development of your repository? Overall findings (n=106)

Mean ratings, where 0=Not on the agenda and 3=High priority

	Mean rating
Raising awareness of repository	2.65
Attracting more use	2.63
Getting more stuff in	2.53
Integration with other institutional services and systems	2.33
Greater interoperability	2.14
Develop institutional mandate	2.06
Preservation and curation	2.05
Better usage statistics	2.02
Managing research datasets	1.85
Raising awareness of preservation issues	1.81
Extend scope in terms of content types	1.74
Consolidating existing policies into a more coherent framework	1.72

There is little difference by size and what there is might have been expected, for instance the stronger prioritization of greater interoperability and managing research datasets in the case of medium to larger repositories (Table 18).

Table 18: What are your priorities for the development of your repository? Analysis by size of repository (n=106)

Mean ratings, where 0=Not on the agenda and 3=High priority

	Small	Medium to large
Raising awareness of repository	2.63	2.69
Attracting more use	2.60	2.69
Getting more stuff in	2.57	2.47
Integration with other institutional services and systems	2.24	2.49
Develop institutional mandate	2.06	2.05
Preservation and curation	2.01	2.11
Greater interoperability	2.00	2.38
Better usage statistics	1.96	2.13
Raising awareness of preservation issues	1.70	2.00
Extend scope in terms of content types	1.67	1.87
Managing research datasets	1.67	2.15
Consolidating existing policies into a more coherent framework	1.64	1.87

There are really quite large differences between Europe and America and this suggest a difference in outlook (Table 19). There might be a structural explanation: universities in Europe are public, most US institutions private.

- Europe rates preservation and curation a significantly lower priority than North America (ANOVA F=6.80, p<1%).
- Europe rates raising awareness of preservation issues as a significantly lower priority than North America (ANOVA F=3.40, p<5%).
- Europe rates better usage statistics as a significantly higher priority than North America (ANOVA F=7.07, p<1%).

Table 19: What are your priorities for the development of your repository? Analysis by region (n=106)

Mean ratings, where 0=Not on the agenda and 3=High priority

	North America	Europe	Rest of the world
Attracting more use	2.68	2.61	2.55
Raising awareness of repository	2.68	2.61	2.73
Getting more stuff in	2.56	2.55	2.36
Preservation and curation	2.37	1.78	2.00
Integration with other institutional services and systems	2.23	2.48	2.09
Raising awareness of preservation issues	2.14	1.55	1.73
Develop institutional mandate	2.07	2.04	2.11
Greater interoperability	1.93	2.31	2.18
Managing research datasets	1.93	1.73	2.09
Better usage statistics	1.77	2.22	2.09
Extend scope in terms of content types	1.74	1.73	1.82
Consolidating existing policies into a more coherent framework	1.60	1.75	2.09

2.1.15 Advantages, disadvantages and perceptions

Advantages

The numbers jump up here because we asked those who were planning a repository to rejoin the survey because their views on perceived advantages and disadvantages are well worth including in the analysis.

First, Table 20 provides the advantages, and clearly there are many of them. Providing access to publicly funded research is thought to be the biggest advantage, although long term preservation of institutional materials is a very close second. There are probably no surprises here but it is interesting that ‘outward facing’ factors (i.e. aimed at specific groups people outside their institution) generally fare less well. Of most interest to publishers, contributing to the reform of scholarly communication and publishing, came sixth.

Table 20: Thinking about digital repositories in general, how important or unimportant do you consider the following possible advantages? Overall findings (n=141)

Mean ratings, where 0=Not at all important, 3=Very important

	Mean rating
Providing maximal access to the results of publicly funded research	2.50
Long-term preservation of your institution’s digital materials	2.48
Enhancing the external prestige of your institution	2.41
Reducing the time between discovery and dissemination	2.35
Better services to students inside your institution	2.34
Contributing to the reform of scholarly communication and publishing	2.33
Contributing to the changing library culture, more digital	2.21
Maintaining control over your institution’s intellectual capital	2.20
Better services to researchers outside your institution	2.17
Provision of identifiers for easier citability of digital materials	2.04
Better services to learning communities outside your institution	2.01
Registration of new ideas	1.53

One respondent pointed to a, possibly, growing advantage of repository deposit:

Tenure and promotion have become a very important reason that our faculty deposit ... and this will become even more of a focus as junior faculty (who are increasingly committed to sharing their outputs) become established and as what counts for tenure and promotion shifts to away from the published literature to a wider range of object content.

There were a few minor differences according to size of repository but none really significant (Table 21).

Table 21: Thinking about digital repositories in general, how important or unimportant do you consider the following possible advantages? Analysis by size of repository (n=106)

Mean ratings, where 0=Not at all important, 3=Very important

	Small	Medium to large
Providing maximal access to the results of publicly funded research	2.69	2.51
Enhancing the external prestige of your institution	2.43	2.46
Better services to students inside your institution	2.39	2.21
Long-term preservation of your institution’s digital materials	2.34	2.41
Contributing to the reform of scholarly communication and publishing	2.28	2.46
Reducing the time between discovery and dissemination	2.24	2.51
Contributing to the changing library culture, more digital	2.12	2.38
Better services to researchers outside your institution	2.10	2.28
Maintaining control over your institution’s intellectual capital	2.07	2.13
Better services to learning communities outside your institution	1.93	2.05
Provision of identifiers for easier citability of digital materials	1.88	2.15
Registration of new ideas	1.54	1.51

However, there are some regional differences, but these are relatively small in absolute terms (Table 22). To determine whether they are real and meaningful we would need to turn to qualitative methods. These differences are:

- (a) Europe rated long-term preservation of your institution’s digital materials significantly **lower** as an advantage than North America and the rest of the world (ANOVA $F=3.70$, $p<5\%$).
- (b) North America rated maintaining control over your institution’s intellectual capital significantly **higher** than Europe but not the rest of the world (ANOVA $F=3.14$, $p<5\%$).
- (c) Europe rated enhancing the external prestige of your institution significantly **higher** than North America and the rest of the world (ANOVA $F=3.68$, $p<5\%$).
- (d) Europe rated better services to researchers outside your institution significantly **higher** than North America and the rest of the world (ANOVA $F=3.10$, $p<5\%$).

Table 22: Thinking about digital repositories in general, how important or unimportant do you consider the following possible advantages? Analysis by region (n=106)
Mean ratings, where 0=Not at all important, 3=Very important

	North America	Europe	Rest of the world
Long-term preservation of your institution’s digital materials	2.65	2.30	2.47
Better services to students inside your institution	2.43	2.21	2.47
Maintaining control over your institution’s intellectual capital	2.40	2.00	2.18
Providing maximal access to the results of publicly funded research	2.35	2.77	2.06
Reducing the time between discovery and dissemination	2.32	2.36	2.41
Contributing to the reform of scholarly communication and publishing	2.25	2.38	2.41
Enhancing the external prestige of your institution	2.22	2.59	2.47
Contributing to the changing library culture, more digital	2.17	2.18	2.47
Better services to researchers outside your institution(d)	2.03	2.34	2.06
Better services to learning communities outside your institution	1.95	2.02	2.24
Provision of identifiers for easier citability of digital materials	1.89	2.16	2.12
Registration of new ideas	1.52	1.61	1.29

Disadvantages

Table 23 provides the disadvantages of having repositories. Copyright is confirmed as a big headache and we return again to the issues of lack of awareness and funding. Library directors are very clear that they do not see repositories as threatening to publishers’ interests, which rings true given other answers in this survey (and not just those that relate to effectiveness).

Table 23: Thinking about digital repositories in general, how important or unimportant do you consider the following possible disadvantages? Overall findings (n=141)
Mean ratings, where 0=Not at all important, 3=Very important

	Mean rating
Confusion and uncertainty over copyright issues	2.20
Lack of awareness by users	2.14
Long term funding and support for repositories uncertain	1.91
Lack of interoperability between repositories	1.88
Costs of long-term preservation and digital curation	1.82
Not comprehensive: lack scale and critical mass	1.57
Fragmentation of access points to the literature	1.43
Confusion caused by different versions of the same material	1.33
Variable quality of material	1.33

Software not comparable with the best that publishers offer	1.22
Fear of plagiarism	1.08
Threatens the business models of publishers	0.78

Smaller repositories differed from larger repositories in two significant regards (Table 24):

- They rate not being comprehensive: lack scale and critical mass as significantly less a disadvantage (ANOVA $F=4.33$, $p < 5\%$).
- They rate fragmentation of access points to the literature as significantly less a disadvantage (ANOVA $F=4.04$, $p < 5\%$).

We might have thought that they would have answered differently but perhaps small repositories just do not see themselves being part of a larger scholarly communication picture – they perhaps have very local ambitions?

Table 24: Thinking about digital repositories in general, how important or unimportant do you consider the following possible disadvantages? Analysis by size of repository (n=106)

Mean ratings, where 0=Not at all important, 3=Very important

	Small	Medium to large
Confusion and uncertainty over copyright issues	2.24	2.13
Lack of awareness by users	2.24	1.97
Long term funding and support for repositories uncertain	1.88	1.95
Lack of interoperability between repositories	1.75	2.10
Costs of long-term preservation and digital curation	1.72	2.00
Not comprehensive: lack scale and critical mass	1.42	1.82
Variable quality of material	1.36	1.28
Confusion caused by different versions of the same material	1.28	1.41
Fragmentation of access points to the literature	1.28	1.69
Software not comparable with the best that publishers offer	1.21	1.23
Fear of plagiarism	1.12	1.03
Threatens the business models of publishers	0.79	0.77

Of the regional differences the largest concerned cost (Table 25). Europe rates this factor as significantly less important than North America (ANOVA $F=5.89$, $p < 1\%$), which suggests very different levels of concern about long term funding for repositories. Perhaps there is just a lot more public money swirling around (still) in Europe as opposed to North America. The second major difference was in regard to the variable quality of material; North American and Europe rated this factor significantly less important than the rest of the world (ANOVA $F=4.46$, $p < 1\%$). Finally, North American and Europe did not really believe that repositories threatened the business models of publishers at all but the rest of the world thought there was some risk (ANOVA $F=3.47$, $p < 5\%$).

Table 25: Thinking about digital repositories in general, how important or unimportant do you consider the following possible disadvantages? Analysis by region (n=106)*Mean ratings, where 0=Not at all important, 3=Very important*

	North America	Europe	Rest of the world
Costs of long-term preservation and digital curation(a)	2.24	1.62	1.88
Confusion and uncertainty over copyright issues	2.22	2.18	2.59
Lack of awareness by users	2.22	2.00	2.29
Long term funding and support for repositories uncertain	2.06	1.82	2.18
Lack of interoperability between repositories	1.84	1.90	1.94
Fragmentation of access points to the literature	1.54	1.33	1.65
Not comprehensive: lack scale and critical mass	1.52	1.62	1.47
Confusion caused by different versions of the same material	1.40	1.41	1.53
Software not comparable with the best that publishers offer	1.37	1.15	1.76
Variable quality of material	1.30	1.46	2.12
Fear of plagiarism	0.95	1.20	1.65
Threatens the business models of publishers(d)	0.63	0.85	1.29

Perceptions and attitudes

Repository managers were asked whether they agreed or disagreed with a number of, possibly contentious, statements about repositories (Table 26). Probably, the most interesting finding is that library directors definitely do not see repositories as impacting negatively on publishers' revenues: rather the opposite they tend to agree reasonably strongly that repositories help to raise the digital visibility of research materials. They think the two systems are complementary rather than antagonistic. There were no significant differences in responses by size of repository or geographical region

Table 26: To what extent do you agree or disagree with the following statements about digital repositories? Overall findings (n=106)*Mean ratings, where -2=Strongly disagree and +2=Strongly agree*

	Overall rating
Researchers should not link from their own web pages, but from a repository	+0.59
They are the publishers' friends - they raise the visibility of research material	+0.58
They should be organized by region, country or subject to achieve critical mass	+0.26
They are the first step towards universities becoming digital presses	+0.20
They will gradually become redundant as more material becomes open access	-0.64
They impact negatively on publishers' revenues	-0.72

2.1.16 The future for repositories

Library directors were asked whether they thought that institutional repositories would be more or less important in three years' time. Not unnaturally perhaps – after all they had made a considerable emotional investment in IRs – hardly anyone thought they would be less important. The vast majority felt that they would become more important (Table 27). Indeed, one fifth thought they would be much more important:

They are the future and libraries should lead the way - it's what we've always done but with digital resources not print. [Small institutional repository, Higher education, Ireland].

Tenure and promotion have become a very important reason that our faculty deposit in our repository -- and this will become even more of a focus as junior faculty, who are increasingly committed to sharing their output, become tenured and as what counts for tenure and promotion

shifts from solely the published literature to a wider variety of content. Also an important shift is the object-based way research is now done, so the broader acceptance for variable quality of materials within a single repository. PLoS ONE, if considered a repository of sorts, has really changed the playing field in this regard, and repositories more generally are embracing this approach. And, finally, the question of who will take care of the data needs of the university, especially as journals move away from allowing supplemental materials of any and all kinds, as well as requirements from funders for data sharing, will be the major driver for dataset content in repositories of all kinds. The wave is coming; for some it is here now. [Medium to large institutional repository, Higher education, United States]

There was no difference in response by size of repository or geographical region.

Table 27: Compared with today, in three years' time, do you think that *institutional* repositories will be more important or less important to researchers? Overall findings (n=106)
Column percentages and mean ratings, where 1=Much less important and 5=Much more important

	All respondents
Much less important	0.0%
Less important	3.8%
About the same	20.8%
More important	54.7%
Much more important	20.8%
Total	100.0%
Mean rating	3.92

When asked just about subject-based repositories (which is what the publishers are probably most worried about) we find a very similar profile of response and again there was a consensus regardless of size of repository or where it was located.

Table 28: Compared with today, in three years' time, do you think that *subject-based* repositories will be more important or less important to researchers? Overall findings (n=106)
Column percentages and mean ratings, where 1=Much less important and 5=Much more important

	All respondents
Much less important	0.0%
Less important	2.8%
About the same	23.6%
More important	57.5%
Much more important	16.0%
Total	100.0%
Mean rating	3.87

2.1.17 Gold open access: who pays?

This was an additional question which sought some contextual information about the gold route to open access. Table 25 shows that in most places, money for author pays open access publishing is unavailable locally and so authors will need to turn to their funder or some other source (perhaps, departmental?).

Table 25: Does your institution provide financial support for journal processing fees for author to publish in `Gold' open access journals? (n=85)

Numbers and column percentages (tick one - only repositories that manage journal articles)

	n	%
Yes	23	27.1
No	48	56.5
I'm not sure	14	16.5
Total	85	100.0

2.1.18 Free-text comments

Large numbers of comments were made including: the provision by publishers of author friendly rights management; the need for far greater co-ordination among repositories; the important role of repositories in research assessment exercise; a more nuanced role for repositories; the strategic role repositories play in the shift to digital and we are only witnessing the beginning of the shift; and merits of subject repositories over institutional ones and vice versa. The detailed comments can be found in Appendix 1.

2.2 Researcher study

As mentioned in the introduction to this report we wanted a demand-side response to the questions we asked library directors in order to triangulate the data, to see whether the two sectors were singing from the same hymn sheet. Researchers were obtained from an IoPP email list. The subject distribution was as follows: Physical scientists (60%), engineers (23%) and mathematicians (11%) accounted for the large majority of respondents were grouped into three broad age bands: Under 30, 30 to 49, and 50 or older.

2.2.1 Depositing to repositories

A definition of a digital repository was provided and nearly two thirds of researchers said they had deposited in such repository (Table 1). However approaching a quarter was not sure whether they had or not. On this question, there is a small but statistically significant age effect: 30-49 year olds are the most likely to deposit research materials (about 7 per cent more likely than the Under 30s or 50 and older age groups). There are no great surprises why some respondents have not deposited (Table 2), the modal response being that they are not aware of the existence of non-publisher repositories. These responses are uniformly distributed across the three age groups.

Table 1: Have you ever deposited any of your own research outputs in a digital repository (see definition below) or given permission for someone to do this on your behalf?

ALL RESPONDENTS

	<i>n</i>	Column%
Yes	1,079	63.7
No	240	14.2
I'm not sure	376	22.2
Total	1,695	100.0

Digital repository

Many libraries are actively involved in building digital repositories of their institution's books, papers, theses, and other works that can be digitized or were 'born digital'. Many of these repositories are made available to the general public with few restrictions, in accordance with the goals of open access, in contrast to the publication of research in commercial journals, where the publishers often limit access rights (adapted from Wikipedia).

Table 2: What is the main reason why you have not deposited your research outputs in a digital repository?

ONLY RESPONDENTS WHO SAID THEY HAD NEVER DEPOSITED

	<i>n</i>	Column%
Unaware of their existence	65	26.3
Lack of knowledge of how to deposit material	51	20.6
Lack of time and inclination	43	17.4
Fear of copyright and publishers' policies	35	14.2
Unconvinced there is any personal benefit	26	10.5
Perception that repositories contain second-rate material	18	7.3
Fear of plagiarism	9	3.6

Other	7	2.8
Total	247	100.0

2.2.2 Reasons for deposit

The categories in Table 3 are overlapping to an extent (so for example, someone whose institution had a mandate in place but would have deposited in any case, could have answered either of the options at the top of the Table). With this caveat in mind, the responses suggest that institutional or funding mandates and peer pressure are extremely helpful in terms of driving depositing behaviour. There was no age-related difference with respect to responses except in terms of responding to an institutional mandate. This is statistically very significant, with those researchers Under 30 being more than twice as likely to comply as those aged 50 or above. This has important policy implications: institutional mandates may be very effective over time as the academic workforce becomes refreshed. No such age-related effect was evident in relation to funders' mandates.

Table 3: Thinking about the last item you deposited in a digital repository, what was your main reason for doing so?

ONLY RESPONDENTS WHO HAVE ACTUALLY DEPOSITED

	n	Column%
I did so voluntarily	515	47.3
I was responding to a mandate from my institution	235	21.6
I was invited to do so by my publisher	121	11.1
I was invited to do so by the repository	56	5.1
A co-author asked me to do it	51	4.7
A colleague suggested it	44	4.0
I was responding to a mandate from my funder	38	3.5
A student suggested it	2	.2
Other	26	2.4
Total	1,088	100.0

2.2.3 Type of repository

It is very interesting that more respondents say they have deposited in an institutional rather than a subject repository, given the iconic status of the arXiv in the physical sciences (Table 4). Again, there are strong age-related effects here. The Under 30s are 34 per cent more likely than the 50 and older age group to have deposited in an institutional repository and 31 per cent less likely to have deposited in a subject repository. These findings suggest that younger researchers are taking institutional repositories (and the mandates that underpin them) very seriously.

Table 4: What type of repository was it? (See definitions below)

ONLY RESPONDENTS WHO HAVE ACTUALLY DEPOSITED

	n	Column%
Institutional repository	478	44.1
Subject repository	430	39.7
Format repository	81	7.5
I don't remember	48	4.4
Other	46	4.2
Total	1,083	100.0

Institutional repository. Institutional repositories are digital collections of the outputs (and possibly metadata about such outputs) created within a university or research institution. They may contain a wide range of materials in various subjects and formats, from journal articles to research datasets.

Subject repository. A subject repository is a specialised digital collection of research outputs confined to a single subject area, such as physics or economics. Material may be deposited from many different institutions.

Format repository. A format repository is a specialised digital collection of outputs confined to a particular form, such as electronic dissertations and theses, or research datasets. Material may be deposited from many different institutions.

2.2.4 Type of content deposited

As might be expected, journal articles and e-theses are the types of document most deposited, accounting for 84 per cent of all deposits (Table 5), with no significant differences between the age groups.

Table 5: Again, thinking about the last research output you deposited, what was it?

ONLY RESPONDENTS WHO HAVE ACTUALLY DEPOSITED

	n	Column%
Journal article	694	64.1
PhD or Masters' thesis	154	14.2
Conference paper	64	5.9
Technical report	31	2.9
Working paper	31	2.9
Research dataset	27	2.5
Book chapter	22	2.0
Book or monograph	17	1.6
Computer software	10	.9
Video recording	4	.4
Patent	4	.4
Image or photograph	3	.3
Metadata-only record	1	.1
Other	21	1.9
Total	1,083	100.0

With regard to journal articles, these are most commonly deposited at Stage I (the author’s manuscript before peer review). Articles at Stages II (after peer review but before publication) and Stage III (publishers’ final version) seem almost equally acceptable. These findings probably confirm the relatively weak importance attaching to peer review in the physical sciences compared with other disciplines.

There are large and significant age-related differences here. Older (50 plus) researchers are 50 per cent more likely than the youngest researchers to deposit a Stage I manuscript and 73 per cent less likely to deposit a Stage II manuscript. This result probably needs probing through interview or focus group discussion at a later date but the implication is that the stamp of peer review is much more important for younger researchers who perhaps have less confidence depositing materials that have not been reviewed.

Table 6: What stage was it at in the publication cycle? (Tick one only)

ONLY RESPONDENTS WHOSE LAST DEPOSIT WAS A JOURNAL ARTICLE

	n	Column%
Author manuscript, before peer review	269	38.9
The final published version (e.g. publisher PDF)	206	29.8
Accepted manuscript, but before publication	168	24.3
Author manuscript responding to the peer reviewers’ comments	49	7.1
Total	694	100.0

2.2.5 Website availability

Around three-quarters of depositors also make additional arrangements to provide access via their personal or institutional web site (Table 7). It would be interesting to find out more about this through qualitative research: why bother to post material from a web site *and* a repository, as well as via the publisher - a case of touching all bases? Surprisingly, there are no age-related differences in response to this particular question. This is an interesting result but unfortunately we do not know how much use is being made by researchers of personal/departmental websites in their search for relevant articles. It may be worth following the ‘web site’ trail a bit further in future as it has all the speed and comprehensiveness aspects of the digital repositories without the problems over converting to a standard IR system, meeting mandate conditions or even copyright concerns. In the debate about the respective values of gold or green open access, the personal web site approach (grey) is always forgotten.

Table 7: Do you provide access to your research outputs from your personal or departmental web site?

ONLY RESPONDENTS WHO HAVE ACTUALLY DEPOSITED

	n	Column%
Yes, all of my outputs	290	26.8
Yes, the majority of my outputs	300	27.8
Yes, a minority of my outputs	104	9.6
Yes, a very selective few	131	12.1
No, never	256	23.7
Total	1,081	100.0

2.2.6 Type of repository preference

Nevertheless, despite the popularity of websites researchers, overall preference is to deposit via a more formal route: a subject or institutional repository (Table 8). Over 60% did so, although, somewhat at odds with the finding in section 2.2 which showed that for most researchers the institutional repository was the place in which they deposited their most recent publication, the subject repository was the preferred location. Just over a fifth (22 per cent) expressed no preference which again is interesting and needs to be followed up. Are they floating voters? Younger researchers expressed a slightly stronger preference for institutional repositories and a slightly weaker preference for subject repositories than the older age groups, but this is not statistically significant.

Table 8: When considering in which kind of repository to deposit your outputs, do you generally have a preference?

ONLY RESPONDENTS WHO HAVE ACTUALLY DEPOSITED

	n	Column%
I prefer to deposit in a subject repository	399	36.9
I prefer to deposit in an institutional repository	257	23.8
I have no preference	244	22.6
I prefer to my personal or departmental web site	125	11.6
I prefer to deposit in a format repository	36	3.3
I generally prefer not to do this at all	19	1.8
Total	1,080	100.0

2.2.7 Repositories as sources of information

In another filter question, we asked whether respondents use non-publisher repositories to *find* information as the basis for two subsequent critical incident questions about their experience. A large majority had (Table 9) and most of the others were not sure they had.

Table 9: Do you use digital repositories to find information? (Tick one only)

ALL RESPONDENTS

	n	Column%
Yes	1,418	83.7
I'm not sure	275	16.2
No	2	.1
Total	1,695	100.0

Table 10 shows to which type of repository researchers went for information – the answer being subject repositories. However, given the tentative answers to some of the previous questions, we should perhaps not assume that respondents fully understand the difference between these repository types, although the definitions even though the questions again reminded them of the differences.

Table 10: Which of the following types of repository have you used to find information (tick as many as apply)

ONLY RESPONDENTS WHO USE REPOSITORIES TO FIND INFORMATION

	n	Row%
Subject repositories	1,053	63.7
Institutional repositories	742	52.3
Format repositories	74	5.2

2.2.8 Relative repository performance

Table 11 shows clearly that respondents associate the repository experience primarily with quality of content, speed of response and ease of use. The other aspects (breadth and depth and ease of navigation) are significantly less valued. The explicit comparator in the phrasing of the question was “the best information services you have used” and so the responses need to be seen in that light. They are “very good”. No age-related or subject differences were found.

Table 11: Thinking of the last time you used a repository to find information, how would you rate your experience (compared with best services used) in respect of the following?

Mean ratings on a scale where 0=Poor and 3=Excellent

ONLY RESPONDENTS WHO USE REPOSITORIES TO FIND INFORMATION

	Mean	Confidence intervals (95%)
Quality of content	2.10	2.05-2.14
Speed of response	2.08	2.03-2.13
Ease of use	2.03	1.98-2.08
Breadth of content	1.92	1.87-1.97
Depth of coverage	1.90	1.85-1.95
Ease of navigation	1.89	1.84-1.94

2.2.9 Advantages and disadvantages of repositories

Respondents view the opening out of access to their research materials as the greatest single advantage offered by repositories, followed by reducing the time between production and ‘publication’ and there is no difference by age group here.

There are some interesting and statistically age differences in response to this question however:

- younger researchers (mean=2.13) rate long-term preservation more highly than older - over 50 – researchers (mean=1.95)
- they also are more inclined to the belief (mean=1.90) that repositories ‘contribute to the reform of scholarly communication and publishing’ (over 50 mean=1.74)

Table 12: Thinking about digital repositories in general, how important or unimportant do you consider the following possible advantages?

Mean ratings on a scale where 0=Not at all important and 3=Very important

ALL RESPONDENTS

	Mean	Confidence intervals (95%)
Providing wider access to the results of publicly funded research	2.36	2.32-2.41
Reducing the time between discovery and dissemination	2.23	2.19-2.28
Better services to researchers outside your institution	2.20	2.15-2.24
Better services to learning communities outside your institution	2.10	2.05-2.14
Long term preservation of your institution’s digital materials	2.04	1.99-2.09
Better services to students inside your institution	2.02	1.97-2.07
Registration of new ideas	1.96	1.90-2.01
Provision of identifiers for easier citability of digital materials	1.91	1.87-1.96
Contributing to the reform of scholarly communication and publishing	1.84	1.79-1.89
Changing library culture, more digital	1.84	1.79-1.89
Enhancing the external prestige of your institution	1.71	1.65-1.76
Maintaining control over your institution’s intellectual capital	1.49	1.44-1.55

In regard to disadvantages the two main gripes that stand out from the rest are the variable quality of materials in repositories and insecurity over their long term viability (Table 13). Comparing the youngest group of researchers (Under 30) and the oldest group (50 and over), we find major differences in respect of the importance attached to the following disadvantages. Younger researchers (mean=1.71) are much more concerned about the potential for confusion caused by different versions of the same material being in circulation (older researchers’ mean=1.42). They are also more likely to feel that repositories add to a growing fragmentation of the literature (mean=1.53 as against 1.29 for older researchers) and that copyright issues create confusion and uncertainty (mean=1.73 as against 1.34 for older researchers).

Table 13: Thinking about digital repositories in general, how important or unimportant do you consider the following possible disadvantages?

Mean ratings on a scale where 0=Not at all important and 3=Very important

ALL RESPONDENTS

	Mean	Confidence intervals (95%)
Variable quality of material: no consistent peer review standards	1.88	1.83-1.94
Long term funding and support for repositories uncertain	1.63	1.58-1.68
Confusion caused by different versions of the same material	1.57	1.52-1.62
Fear of plagiarism	1.54	1.48-1.60
Lack of awareness by users	1.52	1.47-1.57
Confusion and uncertainty over copyright issues	1.51	1.45-1.56
Lack of interoperability between repositories	1.51	1.45-1.56
Fragmentation of access points to the literature	1.42	1.37-1.47
Costs of long term preservation and digital curation likely to be high	1.38	1.32-1.43
Not comprehensive: lack critical scale and critical mass	1.37	1.32-1.42
Software difficult to use	1.29	1.26-1.50
Threatens the business models of society and commercial publishers	0.99	0.93-1.04

Researchers were asked to respond to a series of fairly loaded statements (Table 14). These issues really need to be explored qualitatively rather than simply enumerated, because they are rich in meaning. The statement that really seemed to catch their eye was the idea that repositories might be beneficial to publishers by virtue of creating greater digital visibility for research materials, but there is an age split here with younger researchers (mean=+0.68) much more likely to agree with this statement than the older age group (mean=0.51). However, the main issue that divides the generations is a tendency for younger researchers (mean=+0.36) to agree with the proposition that Gold OA will gradually make digital repositories redundant (older group mean=+0.06). On the other issues, there is no evidence of attitudes attaching to particular age groups.

Table 14: To what extent do you agree or disagree with the following statements?

Mean ratings on a scale where -2=Strongly disagree and +2=Strongly agree

ALL RESPONDENTS

	Mean	Confidence intervals (95%)
Digital repositories are the publishers' friends because they raise the visibility of research material	+0.60	+0.55 to +0.65
Digital repositories are the first step towards universities becoming digital presses	+0.44	+0.38 to +0.49
Digital repositories should be organized at higher levels of aggregation: by region, country or subject to achieve critical mass	+0.38	+0.32 to +0.44
Digital repositories will gradually become redundant as more material becomes open access (e.g. Gold OA)	+0.15	+0.09 to +0.20
Digital repositories impact negatively on publishers' revenues	-0.03	-0.08 to +0.03
Researchers should be discouraged from linking materials from their own web pages, it should be in a repository	-0.26	-0.33 to -0.20

In a parallel question (Section 2.1.16), we asked library directors for their views on the likely direction of travel for institutional and subject repositories over the next three years. There is a clear divergence of views with regard to institutional repositories, where 49.8 per cent of researchers believe they will become more, or much more important, compared with 76.6 per cent of library directors (Table 15). This is of course not an entirely fair comparison since library directors have to consider all subject areas, not just the physical sciences.

Table 15: Compared with today, in three years' time, do you think that *institutional* repositories will be more important or less important to researchers?

Row percentages

ALL RESPONDENTS

	Much less important	Less important	About the same	More important	Much more important	Total row
Researchers in this survey	2.5	12.3	35.3	44.0	5.8	100.0
Library directors	0.0	2.8	20.6	53.2	23.4	100.0

Researchers are much more bullish about the future prospects for subject-based repositories, with 70.2 per cent (as against with 49.8 per cent for institutional repositories) predicting that they will become more, or much more, important (Table 16). Library directors seem to feel that both institutional and subject repositories will equally become more important (76.6 and 75.9 per cent respectively).

Table 16: Compared with today, in three years' time, do you think that *subject-based* repositories will be more important or less important to researchers?

Row percentages

ALL RESPONDENTS

	Much less important	Less important	About the same	More important	Much more important	Total row
Researchers in this survey	0.6	2.7	26.5	59.2	11.0	100.0
Library directors	0.0	2.8	21.3	58.2	17.7	100.0

For neither question could we detect any age-related differences among researchers.

2.2.10 Free text comments

The questionnaire ended with a question which asked researchers to add anything they liked in connection with repositories. And they did, in droves. It was certainly not the stolid topic we thought it would be, instead it seemed the kind of question guaranteed to draw out every platitude, worry, concern and complement that could possibly be made in regard to the functioning of the scholarly communication system. Because of the general quality of the remarks and their off the cuff quality we have reproduced a large number of them here. Most of the comments could be loosely described as pro-repositories, with the French and the developing nations particularly keen. A good number were also anti publisher, quite passionately so in a number of cases. Not surprisingly the topics of peer review, data inclusion, the digital fog, copyright, personal

web pages, intellectual freedom, and cost featured highly in the comments. Surprisingly perhaps libraries took a few knocks. Detailed comments can be found in Appendix 2

3.0 Discussion

Publisher relations

Both topical and operational issues were covered, and despite some vocal free-text complaints about how publishers were exploiting their position within the delivery of research reports, there was overall a muted response on this emotional issue within the questionnaire returns. By the same token, publishers have been expressing concern that the growth of digital repositories potentially destroyed the existing journal publication system with its in-built quality control mechanism. However, on this the library directors felt there was some complementarity between the less formal repositories and the more formal journals – that increased digital visibility raised as a consequence of content being accessible within repositories is an advantage to publishers. Only in some of the detailed supplementary commentaries sent in by the researchers show that there is a smoldering resentment about the inequity of the existing publisher-controlled system.

This finding represents a starting point to a quite separate discussion. Thus it is pretty obvious that if all the papers or the great majority were deposited and easily accessible (harvesting of papers that works)) why should anyone buy a journal? However it looks as if the green approach has just been too difficult to realize with harvesting never working properly. As a consequence several senior librarians have come to the conclusion that green does not work and are moving towards the gold OA option.

Disparities

The disparate nature of institutional repositories – with differing software, formats, interfaces – and how this meets a need currently being met by Google and other worldwide search engines in achieving one-stop comprehensive coverage –emerged more strongly, in the few free-text comments of the researchers. However, given the prominence of arXiv as a global subject based repository of choice for the researchers contacted, the issue of IR dispersion is not really a problem we would not have expected too much comment in this direction.

National differences

The results do suggest that there are differences in attitudes of library directors between the main geographical regions on some fundamental strategic issues. Many of these differences are embedded in the different traditional library cultures – for example, there being more open/public library approaches in Europe compared with the United States. However the differences are not that significant.

Copyright

One burning issue which confronts researchers is the copyright issue – what they think they are entitled to do with their article post-publication. Only 35 out of 247 non-contributors to repositories specifically highlighted the copyright issue in this study, although a number of free

comments also highlighted copyright. A separate report from Publishing Research Consortia (PRC) demonstrated the confusion among authors about the rights over their articles⁷. This confusion has a dampening effect on creating an extensive and comprehensive and 'open' system.

Multi-media

On the other hand there is support for the use of digital repositories in which to store non-article formats, particularly datasets, video clips, and other non-textual supplementary material. Though journal articles are seen by library directors to be the main content, followed by conference papers, there is a long list of other format types which are seen as being relevant for inclusion. E-thesis and book chapters are of obvious interest, but working papers, research datasets, which were raised by many researchers in their free text comments, and learning objects are also there – all formats which are not included as part of the current mainstream journal publishing system.

The problem of course is that handling datasets require training that librarians do not have and this costs, so many libraries who would like to ingest datasets and support the research needs of faculty cannot do what they would like to do.

Promotion

There is clearly a marketing/promotional issue which has yet to be dealt with in bringing institutional repositories in particular to the forefront of the communication system. As far as researchers are concerned 47% of those not currently depositing in repositories (and don't forget this is a group which largely has a high awareness of a repository system because of arXiv) don't know about repositories' existence. Sixty-five of the 247 who admitted to be non-users were unaware of its availability and a further 51 lacked knowledge on how to deposit their manuscripts. A further 43 lacked time and inclination.

Low budget enterprises

Surely the biggest shock to come out of the project was the realization of how small enterprises digital repositories were, especially when compared with publishing houses. Their bark would appear to be bigger than their bite. The financial commitment to institutional repositories can be reflected in the manpower allocated to running the repository within the library and only 8 of the 106 respondents claimed that the staff used to run the repository exceeded five people. Most run on miniscule budgets and in these circumstances you cannot help but wonder how repositories, in their present form, can deliver their ambitions.

Submission

Many library directors have established a local mandate (involving varying degrees of sanction) – one-third of the 85 directors who responded operated a mandate, and a further 15% were

⁷Morris, S., (2009), Journal Authors' Rights: perception and reality, PRC Summary Paper 5

planning to do so. Thus more than half (51.8%) had no mandate, and one suspects that most of those who had not responded to this question (a further 21) also had no repository. There are obviously local administrative hurdles to overcome in many institutions before compulsory inclusion of all locally-produced publications can be implemented. On the flip side most (47%) authors/researchers willingly submit their manuscripts to the local repository and this can be partly put down to the culture of physical science. Only 22% did as a direct result of an institutional mandate, although of course mandates might have been in force in institutions where researchers voluntarily contribute their content. It is the younger age group which seems to respect a mandated approach to submissions than the more elderly age groups. The demographics could therefore work to the advantage of a mandated repository system in future. There was no strong support for a third type of repository – that of individual researchers maintaining their own or departmental web sites of their own publications. Only 11.6% were supportive of this form of repository, which was four times greater than the support given to format repositories.

Scholarly communication

In general this study highlights some important issues about the acceptance of the digital repository as a mainstream function within the current research information system. It also dispels some widely held assumptions about the impact repositories are having on the main stakeholders. More work still needs to be done to monitor an emerging development, but this attempt to bring together the views and opinions of administrators and users provides tangible evidence in what is a sector still fraught with emotion and misunderstandings.

Differences between views of library directors and researchers

- Library directors, probably inevitably, seem to be looking at digital repositories from a process point of view, whereas the researchers were seeing them as a tool. The distinction meant that the librarians were operationally focused, having concerns about how: the repository can help with providing the library (rather than the institution?) with a valuable shop window; they can best deal with curation; they get authors to submit and use the information; more generally, they can they can provide more support for open access. The researchers seemed to be looking at the challenges the repositories created for their research experience – uncertainty over copyright, what stage in the article to deposit, how to deposit, and what impact issues such as gold will have.
- The impression given by the library director is that repositories are there to project the research of their institutions and to act as custodians for it. The researchers gave greater preference to depositing their material in a subject-based repository (which can be explained by the heavy physics focus, which has arXiv as the main repository) and as a result most respondents (almost 84%) use repositories to find information. It is unclear how much of this is attributable to arXiv and how much to the state of local institutional repositories, but it does give a lie to many publishers' assumptions that repositories are not used.
- As far as the future is concerned there seems to be a stronger feeling by the researchers that 'gold' open access will replace 'green', and that within the repository movement itself, subject

based repositories are better placed than institutional repositories to meet their needs. This differs from the perceptions given by library directors which give a higher rating to institutional repositories.

Coming of age?

Finally, what is the answer to the research question we posed – have digital repositories come of age? Well, it would seem on the evidence of the survey – and we need to bear in mind the physical science slant, that digital repositories are still in a formative stage – not yet come of age, still aspirational. They are not totally accepted as a cultural feature of the scholarly communication landscape but there are drivers in place which are moving them towards early adulthood. And they do seem to have made significant inroads despite their limited resources. However, attitudes still have to change among the main section of the research community for repositories to be seen as having reached any degree of maturity.

Appendix 1: Free text comments from library directors

Author friendly rights management.

Publishers should support institutional repositories with author friendly rights management. I imagined that the hard part of launching this was going to be adoption. We have adoption! We can't get the rights to put the full content up - even the pre-publication version. [Small format-based repository, Higher education, United States]

Need for greater co-

ordination/organisation. *A more structural, long lasting European strategy and framework would be desirable. There are currently so many (overlapping) projects ongoing, as well as ended ones.* [Small institutional repository, Higher education, Italy]. *Collaboration among smaller institutions will become more important for financial and staffing issues; working together although each have own identity to preserve and make accessible.* [Higher education, United States]. *Interoperability is the most important topic of the next years. We must work with the metadata of the repositories to make the content visible.* [Medium to large institutional repository, Higher education, Germany]. *In the UK at least, I would have thought that institutional repositories represent an ideal opportunity for a shared national service, as we are all reinventing the wheel at a time when we can't really afford to do so.* [Small institutional repository, Higher education, United Kingdom]

At the forefront of paradigm shift from print to e-only. *Sloppy digital repositories can be harmful, as they lower standards of scholarly communication. In fact, they are the battlefield for better metadata and improved recognition of granularity of scientific discourse. Moreover, they have to redefine that granularity and control it, otherwise we are just lost, one speaks to one and scientific*

endeavour loses its social dimension. Also, digital repositories are the natural outcome of the paradigm shift: from print to e-only. If they are not cultivated, they are just practiced, this way or another. [Higher education, Poland]

Small can be beautiful too. *I believe undergraduate liberal arts colleges actually contribute extensive amount of peer-reviewed journal articles and excellent student work. Our undergraduates often produce works that equal or excel works produced by graduate students. Our faculty are just as productive in their disciplinary journals; however, more interest is focused on ARL level institutions and their experiences with repositories. I will take this opportunity to emphasize that it is not the SIZE of the institution, but the quality of published materials, and that non-ARL institutions are contributing quality materials to OA scholarship.* [Medium to large institutional repository, Higher education, United States]

Role in research assessment. *The repository is a unique asset to become involved in research assessment. E.g. have a look at my presentation for liber 2011* <http://www.slideshare.net/Wowter/liber2011> *1.* [Medium to large institutional repository, Higher education, Netherlands]

Nuanced role for repositories. *The last set of questions here seemed to have the underlying assumption that the focus of repositories are on providing OA to the published literature; in my experience, many repositories -while still focused on that set of material - are more concerned with grey literature, research datasets, and student work (both graduate and undergraduate). While some on both sides of the open access debates do tend to see repositories only as tools for OA, I think that most in the*

repository communities, see a much more nuanced role for institutional and subject repositories as part of a much larger system that includes publishers, societies, etc.

[Medium to large institutional repository, Higher education, United States]

Need a better platform. *Much could be done to improve the finding and viewing of digital materials -- particularly where large or multi-volume documents are concerned.* [Medium to large subject-based repository, Private sector, United States]

Subject v institutional.

It does not matter: I don't see repositories as either / or e.g. institutional or subject. If we find ways to enable our researchers to submit articles once to interoperating repositories why cannot their work appear in more than one place? We already use our repository to create personal web pages publication lists for each researcher and why not? Why should this stop researchers showing their work in a subject repository? [Small institutional repository, Higher education, United Kingdom].

Against IRs: I am personally less convinced of the value of institutional repositories and effectiveness of institutional OA mandates. I think academics are more comfortable with subject repositories such as SSRN. I think publishers will move the sector to an author-pays model rather than subscriptions model in order to preserve their business. [Higher education, United Kingdom].

For IRs: IR has an important role to play both in the institution and external to the institution. Aggregation of content by topic/type is useful to users and can be accomplished by portals, but doesn't replace the need for a comprehensive record for the institution. Content creators can (and will) learn to rely on institutional IR as the stable, permanent location of their materials and source of links to be displayed on their own

websites, included in their CV, and used in other publications. [Medium to large institutional repository, Higher education, United States]

University libraries until now have promised to collect or gain access to the research outputs of all other universities and research institutions, a task that is both impossible to accomplish and costly to attempt. With the widespread introduction of institutional repositories, however, it is now feasible for each university or research institution to collect all the research outputs of its own scholars, and make them available to all other universities. This task, by contrast, is finite and achievable; the costs are commensurate with the research standing, and income of the academic institution. [Small institutional repository, Higher education, United Kingdom]

Which software? *Is there a particular software emerging as the most chosen?* [Higher education, United States]. *Digital repositories of evidence based nursing research that is unpublished, existing at an institutional level could benefit from repository software, but typically healthcare institutions are not aware of the resources academia uses to organize this valuable information.* [Private sector, United States]

There is a lot more stuff in the pipeline. *We are actually beginning an in tandem project of repository and publishing of research. We will unveil our first faculty proceedings, which will include presented but not necessarily published papers, abstracts of published materials, and links to online copies of research conducted by our faculty. We have published this as a bibliography annually - we wanted to be able to share the research that wasn't published in the traditional manner. We just started a master's program and will be publishing master's research as well as their theses or final projects. In addition we have a more and more active undergraduate research component and we want to give*

those students an outlet to publish. When publishing we will be making this available as open access material at the same time as saving it. As a smaller institution this is something we would not have been able to

do even a few years ago, but we have a willing faculty and we also are better set in the library to offer these services and make them available to the wider academic world.
[Higher education, United States]

Appendix 2: Free text comments from the researchers

UPSIDES OF HAVING REPOSITORIES

They are good for science

All initiatives that may help to disseminate scientific knowledge, including digital repositories, must be supported. Scientific knowledge belongs to all of us; we must preserve it and pass it on to future generation of researchers. (Academic faculty, Physical sciences, 55-59, Spain).

There's a huge range of value in the available repositories I'm familiar with. In astronomy, researchers cannot survive for more than a day without referencing the ADS literature database or the arXiv preprint server. Those have superseded all journal-based and library-based literature repositories in our field and are quite well optimized for use by active scientists. There are separate data repositories in astronomy, e.g. CDS and the Space Telescope Science Institute, holding a variety of published and unpublished datasets. Again, these are more useful and better organized than repositories associated with publishers. (Academic faculty, Physical sciences, 65-69, United States).

They offer digital visibility

Almost everybody prefers to access material digitally over the internet; if it's not there, and not accessible, my research simply will not be read, will thus not have impact and will not be cited. Therefore I can't afford to publish in venues that don't allow me to make my material accessible digitally. In my experience, if it's out there anywhere, scholar.google.com will find it, so it doesn't matter where it is or what the interface is or whether it's part of a coherent collection or distributed randomly across the web; Google serves

as a decent default interface even in the absence of one provided by the repository. What's important is that it remains in the repository and accessible over the long term. Once open and perpetual repositories for digital materials exist (e.g., University libraries), there is no fundamental reason why the functions of quality-validation/peer-review and providing access to accepted material need be inextricably coupled. A "virtual journal" supplying the peer-review and editorial process could in principle "publish" simply by posting a permanent table of contents and abstracts that merely links to the accepted articles in their permanent repositories, say distributed across the various libraries at the various Universities hosting the scholars, or which have committed to host that journal on behalf of unaffiliated authors. This only requires confidence in the continued existence and commitment of the institutions supporting the libraries ... but in any event the continued existence of a traditional publisher's material depends on the continued existence of that publisher or libraries that archive the paper or electronic material, so fundamentally the "new" model is no more or less permanent and secure in the long term than the current one. (Academic faculty, Engineering and technology, 50-54, United States).

They offer unrestricted publishing

An issue concerning the classical peer reviewed press. Is that they keep vying to increase their citation index. To be able to do this they are currently asking to shorten the articles (4000 - 5000 words), have a single table and few figures. It becomes harder and harder to provide a publication with some scientific depth.

Furthermore, many times I have been asked to provide more citations from the journal I am submitting to. These restrictions do not exist in most of the digital publications. Also when colour print is necessary, the digital repositories do not charge you, same for exceeding the 10 page limit. (Academic faculty, Physical sciences, 50-54, Belgium).

They are good for developing nations

Lack of specialized information is kind of poverty. Venezuela is getting right now closer and closer. Please, keep going to an open access world. (Academic faculty, Physical sciences, 45-49, Venezuela).

The world has tend to be a little village so more integratable tasks should be maintained so as to link scientists everywhere and hence to make easy thinking over a most challenging problems facing us and shortening our planet life time. (Academic faculty, Physical sciences, 45-49, Sudan).

Researchers in many countries do not have enough economic strength to access good quality research. Can repositories think on the line of GLOBAL UNIFICATION to impart free access to any kind of knowledge to one and all? Why knowledge (that is not patented one!) should be seen in terms of making money???. These are assets of human existence in the universe. (Academic faculty, Physical sciences, 45-49, India).

They remove publishers from the scholarly chain

ArXiv has essentially become the standard method of publication in high energy physics; many don't even bother with journals any more. (Doctoral student, Physical sciences, 20-24, Canada).

As we have less and less access to electronic journal due to their cost, electronic depositaries like arXiv.org will soon in my opinion completely replace electronic journal. The peer review process does not directly exist as known in official journal but still exist as only some of the work is quoted and therefore more fair. It is certainly less dependent to the friendship you (or the co-author) have with the reviewer. (Salaried researcher, Physical sciences, 25-29, Switzerland).

Digital repositories, if not anything else, is a counter balancing initiative to the commercial publishing which, to my experience, has become slaved by so called big research groups and funding agents to a great extent which is a very bad news for science. There are time tested methods of calling something science or not but that, these days, are being ignored under the pressure of big nexus or networks of scientists who are trying to define science in their way and it simply takes arranging for enhancing some indexes which in many case is completely artificial. Open access publications, so long as they charge the author for her/his publication, cannot take the place of a repository because of the simple reason - why would I pay for the hard work that I have already done and the result being important? Repositories provide the safety and the visibility to the works of those scientists who do not want to fall in the above mentioned nexus or whose results are going against some cooked up ones of the big nexuses. So, I think, it has huge potential and one has to think about better organization and management of repositories. (Academic faculty, Physical sciences, 40-44, India).

I welcome the day that publishers no longer derive profit from academic publishing. This research, paid for by

public funds, should be in the public domain. (Academic faculty, Physical sciences, 55-59, France).

I hope that publisher will not find in the future a workaround to take back ownership on researcher material. Indeed, they tighten the ropes of knowledge for many decades. Now it's releasing, they will surely find a way to survive. Good for them, but how will they do? To be continued... (Salaried researcher, Engineering and technology, 30-34, France).

In my field, high energy physics, we use the arXiv now for more than 25 years and I think everyone in the field agrees that this had extremely beneficial impact on our science and research. I think there is no space for for-profit publishers in our field anymore and completely digital journals run by scientists for scientists will be the future. (Academic faculty, Physical sciences, 35-39, United States).

In my field, many colleagues read arXiv on daily basis and journals only rarely. (Salaried researcher, Mathematical and computer sciences, 40-44, Czech Republic).

In terms of scientific publications, there is no need for publishers anymore, really. Right now they are just sucking the blood off libraries and universities while all the work involved is done by the researchers. (Salaried researcher, Physical sciences, 30-34, Germany).

My main concern is the outrageous price required by publishing companies for institutions to subscribe to scientific literature. I view digital repositories as a fundamental tool to provide free access to publications. I even start to feel that the quality of peer-reviewing is become so low

that it does not make a big difference anymore. (Academic faculty, Physical sciences, 30-34, Belgium).

Publishers have for a long time been in the business of extracting research money from universities. Many of the services they provide are negative value. The main value provided is a prestigious name to having your work published in journal of important results. However, once a referee process is overlaid on top of the physics arXiv, and this is accepted as being equally prestigious, the journals will go out of business. This is a good thing. (Academic faculty, Physical sciences, 40-44, United Kingdom).

Subscription to peer-reviewed journals is more and more expensive and tends to be unsustainable to "small" university or college / this increases the difference between small and big university. People will find alternatives to avoid expensive subscriptions: this is the main problem with the music and film (YouTube, dailymotionb). (Academic faculty, Physical sciences, 35-39, France).

They provide a date stamp

For me, the most important reason for publishing in a digital repository is to get a date stamp on my work, ahead of peer-review. I publish in peer-reviewed journals because they provide my work with more prestige than a digital repository (no quality control there), but I use the repository to make my work visible to the community as soon as possible, with a date stamp on it to prove that I was "faster than the competition". I'm not worried about plagiarism. If people want to steal my work, they will do so no matter where I publish. But in such cases, I could even use the date stamp from the digital repository to prove who was first to publish and who was second. For my

purposes, it is best to have recourse to independent, subject-based repositories like the arXiv preprint server at Cornell University. That way nobody can ever accuse me of manipulating the date stamp to my own ends. (Salaried researcher, Physical sciences, 35-39, United Kingdom).

They help unblock the blockage

Digital repositories seem to be interesting especially from the development point of view. It expands the way a researcher can access articles/information related to his/her field which is limited right now due to membership issues for journals. (Doctoral student, Physical sciences, 25-29, India).

They create a level playing field

I think the importance of digital repositories in 'democratizing' science has not been fully appreciated. It is no longer necessary to work in a major well-funded centre with a large research library to access the leading research as it happens. Anyone with access to the internet, whether in Stanford or in a minor African university, has, in principle, equal access. This is an enormous benefit to scholarship. (Academic faculty, Physical sciences, 55-59, Ireland {Republic}).

They save money

I strongly value the peer-review process and value traditional journals for this service. I understand that there are some publication costs; however I really do not understand how traditional journals continue to charge such high publication costs when the distribution costs for digital media is much less than that of physical media. If subject repositories such as astroPh and NASA ADS are able to operate without charging users anything, I am amazed that traditional journals still manage to charge so much. (Salaried

researcher, Physical sciences, 30-34, United Kingdom).

I am increasingly concerned by the high cost and price associated with current scientific publishers. Ideally, I would like to see a compromise situation in which scientific publishing presses can remain profitable, since their editorial and oversight is important, but that access is not restricted to their databases due to budgetary problems. Perhaps generalized repositories can alleviate this growing problem. (Salaried researcher, Physical sciences, 60-64, United States).

And it is still early days early days

Digital repositories are still fairly new and the quality of them should be gradually improved to make better contribution to academy and community. (Academic faculty, Engineering and technology, 30-34, Australia).

DOWNSIDES OF REPOSITORIES

They are redundant - websites and Google a much better bet

Given the fact that most academics maintain web pages, local digital repositories are redundant. Moreover, most libraries at research 1 institutions maintain web access to the best, most highly rated journals and hence their content. If one is publishing, then there is access to it internationally and probably with better hit rates through search engines like Google. (Academic faculty, Engineering and technology, 55-59, United States).

Publishing self-produced material through the researcher's own web pages is the least bureaucratic way of publishing. A researcher hardly wants to engage himself into a process of spending time with

contracts and filling forms and accepting all sorts of limitations and rules related to the publishing through third parties. (Academic faculty, Mathematical and computer sciences, 50-54, Finland).

Putting the paper title into Google works so well that I don't currently see the need to care whether the PDF is on a personal web site or an institutional repository. My institution's repository has a terrible interface, and even trying to find the URL for it is difficult because the first hit on Google is for a now defunct earlier institutional repository. I put all my papers on my own web site once they are accepted. I prefer not to put them in a repository so that I can easily update the PDF with a DOI, page numbers, etc as the paper progresses through the publication process. (Academic faculty, Physical sciences, 35-39, United Kingdom).

They create plagiarism worries

I am more concerned about the copyright and the plagiarism of the materials issues. (Academic faculty, Physical sciences, 35-39, Malaysia).

They lack peer review

Journals provide a very useful service in peer reviewing the best work and 'validating' it for non-experts. I fear that this may be lost as people make their work available in open-access formats. I distrust open access journals, not because I am opposed to the principle of freely sharing information, but because if you pay \$1000 to get an article published the journal is much less likely to reject it, i.e. I fear that market forces discourage maintenance of academic standards. I would always choose to publish my best work in a traditional peer reviewed journal, if I additionally put a pre-edited copy in an institutional repository then I undermine the journal (it

has to make money somehow) ... I don't know much about this - these are just my feelings. (Academic faculty, Biomedical and life sciences, 35-39, United Kingdom).

There are problems in regard to trust

A very important point is accountability. Digital repositories need to be run by publishers or independent organisations like arXiv. The reason is that one has to be sure WHEN an article appeared first and that one can be sure that no tempering with the electronic version has been done. This is important for historians for example, in case a discovery has been done at the same time by different groups. This is why I believe that publishers of journals do not have to fear the future. I primarily trust published results, and then preprint archives which are run by independent organizations. I trust repositories from institutions less because it is not clear how reliable they are. Also, in the past, these repositories have become neglected. After a while, nobody even knows that such a repository exists. (Academic faculty, Mathematical and computer sciences, 40-44, United States).

There are problems of longevity

Longevity is critical to acceptance. My university has had several digital repositories come and go. Hence I no longer contribute to them. In contrast the arXiv looks solid, even though its interface sucks. (Academic faculty, Physical sciences, 50-54, Australia).

They just add to the digital fog

With the current surge of new free-access journals, the way researchers publish is bound to change. How? I do not know. Digital repository will certainly add more confusion. (Academic faculty, Physical sciences, 60-64, Japan).

There are too many repositories. It has become a challenge to search for materials through all these academic silos. It is certainly a time of transition. The NIH repository is an interesting case of an unnecessary service - thematic repository. (Academic faculty, Education and other subjects, 40-44, United States).

PEER REVIEW AND QUALITY INSURANCE

[This topic is dealt with separately because it does not necessarily fall into one camp or another]

At first, one could put papers on the web without the need for any kind of referee to endorse them. This is no longer always the case and I wish the old system would come back. Digital repositories should be completely open so that controversial but ultimately sound ideas are not prevented from being aired. (Academic faculty, Physical sciences, 50-54, Zambia).

It is important to preserve international peer review and therefore create a kind of international label for digital repositories following a clear reviewing process. (Salaried researcher, Engineering and technology, 55-59, France).

Although strict peer review not needed, the digital repositories still need to go through some rating mechanisms so that new readers can have a rough adjustment about the material at the first glance. (Doctoral student, Engineering and technology, 25-29, Singapore).

Digital repositories should not be submitted to "peer revision". It is my personal experience that sometimes when one reaches a result that contradicts the mainstream, the real merit of the work is not evaluated. Powerful research groups tend to use their influence to prevent this

result from being published. I once came upon a "pearl of wisdom" from an anonymous referee: "I did not check the calculations, but as they are contradicting what everybody is doing, they can only be wrong". And the editor forwarded that "evaluation", instead of assigning the work to another more serious referee. Referees sometimes behave quite indecently, shielded by being anonymous. (Salaried researcher, Physical sciences, 55-59, Brazil).

I really like the arXiv model (see arxiv.org) because researchers are vetted the first time they submit a paper. This seems to keep the quality of submissions sufficiently high, without the expense and annoyances associated with the peer review provided by the "real" journals. Note that I send my arXiv submissions to those "real" journals as well (for the peer review badge of approval of course, but mainly for the desired audience). (Salaried researcher, Physical sciences, 55-59, United States).

The primary role of publishing houses in the future will be for quality assurance. Publication ability will vanish but quality control and assurance will become a central issue. (Academic faculty, Mathematical and computer sciences, 45-49, United States).

I think the articles or contributions should be archived after some sort of internal review. This will ensure uniform level of the materials and increase their credibility. (Salaried researcher, Physical sciences, 30-34, India).

There may be a problem with "moderators", namely, with the definition of the rules they have to follow and on the way they apply these rules. In digital repositories, there are no referees,

but also (usually) no editor... (Salaried researcher, Biomedical and life sciences, 50-54, France).

Just leave them alone: repositories like arxiv.org are fantastic. Researchers are keen on depositing their quality work there, and are disinclined to upload rubbish for the obvious reasons that it is visible to everybody. (Salaried researcher, Mathematical and computer sciences, 25-29, Portugal).

Such repositories need to have a consistent and unbiased process for moderating content. Ideally content would be effectively moderated after submission and publication. A prime example is known as vixra.org. On this site if a paper is bad it will accumulate negative comments. Peer review itself becomes an open process. Quality can be preserved. Beware the attractiveness of exclusivity in publishing. That which is hard to obtain is often thought to be precious, whether it has true value or not. (Masters or undergraduate student, Physical sciences, 30-34, United States).

THINGS REPOSITORIES COULD DO TO IMPROVE THEMSELVES

Improve performance

At present the user interfaces are terribly bad. (Academic faculty, Physical sciences, 55-59, Germany).

Till today I could not find a single repository which is robust, gives accurate return of contents and do not have multiple entries for same journal. I do not understand why publishers are unable to make robust software and powerful database in this 21st century. I would like to request all concerned officials to look into this matter so that repositories will be

more helpful and user friendly. (Academic faculty, Physical sciences, 30-34, India).

Pass the Google test

Can web search engine (e.g. Google) automatically reach the materials inside of a repository? (Doctoral student, Biomedical and life sciences, 25-29, Belgium).

Better content coverage

Conference proceedings are an anachronism (Salaried researcher, Physical sciences, 60-64, United States).

Digital repositories should publish more e-textbooks with multimedia additions and animations. (Academic faculty, Physical sciences, 65-69, Ukraine).

Conservation and access to classical work; Work that has been produced before the electronic age came (Academic faculty, Biomedical and life sciences, 70 years or older, Germany).

Gathering of some kind of data as a research data set is very difficult and needs a large amount of work/time of a researcher to prepare the data for using (e.g. EEG recording from special illnesses). So, I think if a digital repository support the researcher in the process of preparing data, the interest or co-operation of the researchers would be increased. (Academic faculty, Mathematical and computer sciences, 25-29, Iran).

I am a huge fan of data being easily and consistently made public. There is nothing more annoying and wasteful than to email the author of a paper based on data that you want to re-analyse and have them tell you, ""I can't send you the data I published because my hard drive crashed between now and then and the data no longer exist."" (Happened to me, with

data only a few years old.) If you have published results based on data, the data should be stored somewhere accessible to future checks. (Salaried researcher, Physical sciences, 25-29, United States).

In Australia there is an increasingly large push to store publicly funded data on a publicly searchable meta-data repository. Increased visibility and access to research outputs seems to be the norm and there are a number of large government funded projects to make this so. I'm heavily involved with e-research at our university and it is clear to me that there is an increasing need to adequately address data storage and publication curation issues. I think such initiatives are a good idea long-term and my main hesitation in making my publications freely available has been due to concerns related to copyright. For this reason I have preferred to upload such publications on the university library's e-repository - it has become mandatory now in any case. (Salaried researcher, Mathematical and computer sciences, 30-34, Australia).

Repositories need to set up a universally recognised dataset identification - that includes versioning, backward and forward citation and citability. Doi is good for articles, but not sufficient for data. Incorporation of datasets directly into analysis software is key - including automatic updates if needed, or fixed versions if not. All code deposits could be automatically entered into a version control system so that any users/readers can contribute extensions/corrections in a transparent and useful manner. Data synthesis is not the same as data archiving, yet it is the former that is really the challenge. (Salaried researcher, Physical sciences, 40-44, United States).

Become a one-stop shop (like arXiv)

I think one of the reasons arxiv.org is successful is that people post there, no matter where they eventually publish, so it serves as an aggregation point. I only need to check one place for papers, or rather, two places since I also check inspires, which also aggregates physics papers. (Salaried researcher, Physical sciences, 40-44, United States).

I want everything ever published by anyone searchable through Google, then available at one or mouse clicks. Some digital repositories are a step in the right direction. (Academic faculty, Physical sciences, 35-39, United States).

In astrophysics the arXiv repository for preprints & reprints is universally used; and in combination with the NASA Astrophysics Data System (ADS), which provides a database of essentially all published papers and indeed links to arXiv, means that one never has use for any other portal (i.e. all the journal sites are linked through ADS). (Academic faculty, Physical sciences, 45-49, United Kingdom).

There are too many different and in part competing methods dealing with the problem of data overload. May the best win soon. (Academic faculty, Physical sciences, 70 years or older, Australia).

Searching 100 journals is a pain, searching 1000 repositories nonsense. Only large-scale enterprises like arXiv bring an advantage to the community. (Academic faculty, Mathematical and computer sciences, 40-44, Germany).

Obtain entry to web of science

If digital repositories like arXiv.org are listed in web of science and similar compendia, then it becomes a viable digital press. Many scientists already use arXiv as such and with some limited peer

review support, it could be an excellent model for all digital repositories (Academic faculty, Engineering and technology, 30-34, Canada).

GENERAL AND MISCELLANEOUS COMMENTS

Scholarly communication too complex

Different Institutions have different rules and systems for access. Different publishers have different restrictions as well. In my multi-campus institution our libraries even have different access rules. (Academic faculty, Biomedical and life sciences, 70 years or older, United States).

Stop publisher bullying

It is important that Universities do not restrict access to academic material funded via the public, e.g. in preventing researchers distributing on their website or placing in institutional, or subject depositories. I know some universities in the US have reacted to fears of commercially driven law suits by publishers, by demanding that researchers do not distribute their publications on university websites. I expect the same will happen in Australia eventually. This is outrageous as far as I am concerned. I favour personal websites as the most practical but would be happy to see subject area depositories have a greater profile so that there is an alternative to personal websites. (Academic faculty, Physical sciences, 45-49, Australia).

Publishers need not worry

I believe the astronomy community has a very good example of a well-established repository (originally meant as pre-print server, now much wider than that), the arXiv. This is universally used by researchers in astronomy and physics, and has clearly achieved critical mass many years ago. I believe it has not damaged the

traditional journals financially. (Salaried researcher, Physical sciences, 30-34, United Kingdom).

I like to make my accepted paper available as a preprint into a subject-based repository, until the paper is published. Then I delete the link to the preprint. I do like to use the published version and do not like to submit accepted papers to my institution-based repository, where the accepted papers would stay in the public domain even after the published version is out. (Salaried researcher, Physical sciences, 35-39, United Kingdom).

I would guess that Most of the material on the Physics arXiv winds up published in journals (Academic faculty, Physical sciences, 70 years or older, United States).

Librarians keep out

Librarians and others research outsiders are in love with digitization and the quality and availability of the information accessible to researchers has declined in direct proportion to their forays into expensive, clumsy software designed for librarians, not information users and creators. Outsiders should not be determining the policies for information dissemination or access. (For example, here, the librarians are teaching engineering students that the best primary sources of information are the patent literature and dissertations, with peer-reviewed publications being ""gray literature"" and not primary sources, and that anything older than three years is obsolete. Repeated attempts by researchers to correct these misconceptions fall on deaf ears. Any business which failed to assess and consider the customer needs so blatantly would and should rightfully fail. Yet the libraries keep getting more and more money to digitize and to try to circumvent

copyrights and author wishes in an effort to make all work done by others under their control for profit for their institutions.) (Academic faculty, Physical sciences, 50-54, United States).

Many of the issues raised in this survey are of greater importance to librarians and archivists than they are to working scientists. For example, whether a repository is peer-reviewed or not is immaterial, SO LONG AS one knows for sure in which category the repository falls. One then knows how to assess what is retrieved (in Physics, compare arXiv with Physical Review for example). On the other hand, librarians and archivists tend to miss the importance of disseminating one's work from one's own web site. The program officer who oversees my grant needs to be able to look at the work he is funding with one click - on my website - not chase down a potentially incomplete record on multiple repositories over which I, as Principal Investigator, have no control. Plagiarism, particularly from English to Asian languages, is believed to be a huge issue. But it is difficult to get any reliable measures to assess anecdotes. Repositories tend to worsen this problem but their existence is not the cause and they cannot be blamed. I see little hope of solving this problem, especially in view of the proliferation of low-quality online journals. (Academic faculty, Physical sciences, 70 years or older, United States).

Fragile model

Looking at my present library, I have the strongest doubts about the continuity of digital libraries. Digital papers of 20 years ago are for the most part unreadable today, unless some special care has been taken to translate them from obsolete to recent formats. I have no reason to believe that more of 10% of the present digital literature will be accessible in 20

years; while all of a 19th century library is still directly accessible now on paper. (Academic faculty, Mathematical and computer sciences, 50-54, France).

Publishers create their own repository

Publishers should put all articles older than some time (say, 5 years) in free-to-access public digital repositories. (Doctoral student, Physical sciences, 25-29, Bulgaria).

Rock and a hard place

Several publishing companies have unclear policies for the demands from funders that their funded results should be disseminate in an open access database. Researchers are squeezed by publishers that they want the sole version of the paper, and the funding agency demand that the results should be open access. (Academic faculty, Engineering and technology, 40-44, Sweden).

Subject repositories

Subject oriented repositories are much more useful than institution or region based repositories in my academic field, for which the community is worldwide and based across many institutions. (Salaried researcher, Physical sciences, 35-39, United States).

Subject-based repositories are critically important. They allow the aggregation of relevant material in a central location that all members of the field can access. Published journal articles can be difficult to reach, even for University members whose institutions have digital subscriptions---sometimes the library proxies just don't work. This can be a problem when I am travelling, and is definitely a problem for some of the undergraduate students I work with. (Salaried researcher, Physical sciences, 30-34, United States).