Digital repositories ten years on: what do scientific researchers think of them and how do they use them?

David NICHOLAS, Ian ROWLANDS, Anthony WATKINSON, David BROWN and Hamid R. JAMALI

CIBER

ABSTRACT. *Digital repositories have been with us for more than a decade, and despite the considerable media and conference attention they engender, we know very little about their use by academics. This paper sets out to address this by reporting on how well they are used, what they are used for, what researchers' think of them, and where they thought they were going. Nearly 1,700 scientific researchers, mostly physical scientists, responded to an international survey of digital repositories, making it the largest survey of its kind. High deposit rates were found and mandates appear to be working, especially with younger researchers. Repositories have made significant inroads in terms of impact and use despite, in the case of institutional repositories, the very limited resources deployed. Subject repositories, like arXiv and PubMed Central, have certainly come of age but institutional repositories probably have not come of age yet although there are drivers in place which, in theory anyway, are moving them towards early adulthood.*

Introduction

We know there is growth in the open access availability of research publications, both gold (author pays for publication) and green (self-archiving by the researcher). For example, approximately 30% of all articles are thought to be available as open access, two-thirds in green and one-third gold. These figures are not definite, and various pundits give different estimates, but they do provide us with some context.[1] 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, not just journal articles. According to OpenDOAR data (1 March 2012) there are 2,173 repositories worldwide, 82% institutional and 11% subject based.[2]

Digital repositories (whether institutional, subject, or, indeed, format based[3]) have been with us for more than a decade and have become an established, albeit controversial component in an increasingly complex scholarly communications landscape. Many of those journal publishers that are dependent on subscriptions fear that the growth of author self-archiving in their local institutional or subject-based repositories would have an effect on the sale of journals. The assumption is that research libraries would begin to cancel their subscriptions knowing that, as a fallback, they might be able to access almost the same material for free by navigating to the appropriate institutional websites and/or relying on global search engines such as Google Scholar to locate relevant items. This fear led to a group of publishers taking part in a joint research project with the European Commission and a number of European institutional repositories to see whether there are indications of a balkanization of publications (the PEER pro-

David Nicholas

Ian Rowlands

Anthony Watkinson

David Brown

Hamid R. Jamali

ject[4]). However, not all publishers are concerned at the encroachment of self-archiving at the expense of subscriptions – one leading publisher (Springer S&BM) has gone on record as feeling that they are a long way from green open access being commonplace.[1]

Surprisingly, then, given their possible importance, while there is an extremely healthy body of literature on repositories, hardly any of it relates to use and impact on researchers, the very people for whom they are designed. The vast majority of the research literature concerns the mechanics and problems of running them. This paper therefore aims to fill a void and show what impact more than a decade of digital repositories has had on scholarly attitudes and information-seeking behaviour.

The specific objectives of the study were to determine: (1) whether researchers used repositories and, if so, what for; (2) what they saw – in general terms – as the main advantages and disadvantages of digital repositories; (3) whether they agreed or disagreed with a series of provocative statements about digital repositories; (4) whether they thought institutional and subject repositories would be more or less important to researchers in three years' time. For contextual purposes we also wanted to find out about researchers' 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 – 'personal repositories' perhaps.

While the focus of this paper is on users, the Charleston Observatory study upon which it is based was actually broader than that and covered the views and practices of library directors and the findings of this part of the study can be found in the final report of the project.[5] However, we have drawn on this part of the study for the researcher study, to provide context and point to differences in outlook.

Scope and definitions

Digital repositories were defined broadly to include:

- Institutional repositories which aim to collect widely across a particular university or similar institution, possibly covering a wide range of formats.
- Subject repositories based on collecting only within a certain discipline, usually across more than one institution and often international in coverage.
- Format repositories whose scope is limited by collecting in a particular format, e.g. student dissertations and e-theses, research data, digital images.

We were conscious that while library directors would have no problems understanding what was meant by the terms 'digital' or 'institutional repository', researchers might have problems with these, and for that reason during the survey we regularly reminded users of the scope and the definitions. However, the free-text comments we obtained from researchers clearly showed that there were nonetheless the inevitable confusions, and we have pointed these out at appropriate points in the text.

The study was international in scope but the resources were not available to conduct a study of researchers from all subjects. The subject footprint of the study was very much determined by the sampling frame used, which were email lists of the Institute of Physics Publishing (IoPP). This meant that the researchers we are reporting upon in this paper were almost exclusively scientists, and the majority of the scientists were physical scientists. More details about sampling can be found in the following methodology section.

Research context

There is much literature published about repositories. Google Scholar, for example, identifies 62,700 items with institutional repositories as part of the title or within the abstract. But the preponderance of these articles concern themselves with the processes and strategies for implementation; few are evaluative, describe how far the institutional repository (IR) development has been 'user-driven', or even make assessments of use being made of the IR. The few we have found are described here.

Academics are creators of the content of repositories as well as the biggest users of their content. As creators, their low partici-

this paper therefore aims to fill a void and show what impact more than a decade of digital repositories has had

pation rate has been a major problem for repositories, especially institutional ones.[6,7] That is why some institutional repositories have made it mandatory to deposit content.[8] Different reasons have been mentioned for low volunteer participation, including redundancy, fear of plagiarism, learning curve and confusion with copyright.[9] However, academics generally seem to be keener on using subject repositories than institutional ones.[10] Many academics use alternatives to institutional repositories to make their publications available, such as their personal web pages and disciplinary repositories.[9] As a mixed-method study across Europe found out, there are also clear differences between scholars from different disciplinary backgrounds in their understanding of open access repositories and their motivations for depositing articles within them.[11]

We know very little about academics as users of repositories, and what we do know is rather old. In 2007, Dana McKay stated that 'There are no known reports of actual usage of any IR' and that 'virtually nothing is known about IR end-users.'[12] Since then a few surveys and attitudinal studies have been conducted on academics with regard to repositories. A relatively large survey of 1,118 faculty members at the University of California[13] showed that 82% of respondents were 'not aware of' or 'aware of but don't know much about' institutional repositories, and 79% of respondents were 'not aware of' or 'aware of but don't know much about' subject repositories, while 8% had submitted to subject repositories. Davis and Connolly[9] reported that Cornell's institutional repository was largely underpopulated and underused by its faculty as they had little knowledge of and little motivation to use the repository. Another survey of faculty at Louisiana State University[14] showed that usage of subject repositories was one of the main reasons for low institutional repository use. Interviews with 25 users of institutional repositories[15] ascertained that users of institutional repositories, although not yet loyal and devoted, recognized their value and unique nature. A few studies have looked at the use of different types of material. For example, an investigation of use of digital material deposited at Ohio State University IR showed that articles and undergraduate e-theses are the most frequently used type of material.[16]

Methodology

To provide the necessary reach and international coverage, as in the two previous Charleston Observatory studies,[17] an online questionnaire was the primary means of collecting data. The survey was distributed across lists owned by IoPP to 85,000 email addresses between 19 and 31 December 2011. Fully useable completions totalled 1,685, a response rate of just below 2%. 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 ± 2.4 percentage points at the 95% confidence level: so a value of 50% should be interpreted as lying in the range 47.6–52.4.

Responding researchers were quite widely dispersed geographically; nearly a 100 countries were represented. The US had the highest representation with 16% of respondents; this was followed by China (10%), Italy, India and Germany (all 6%), Russia and the UK (both 5%). Fifty-five per cent of respondents were academic faculty, 32% salaried researchers and the rest students (mainly Ph.D. students). With regard to age, 15% were in their twenties; 36% in their thirties; 22% in their forties; 15% in their fifties, and 12% were older. Eighty per cent 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	%
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

we know very little about academics as users of repositories, and what we do know is rather old

we cannot claim, of course, that a sample dominated by physicists constitutes a representative group of researchers

We cannot claim, of course, that a sample dominated by physicists constitutes a representative group of researchers, especially as physicists have probably gone the furthest in terms of embracing subject repositories (notably arXiv, but also the Astrophysics Data Service). Although it is possible, and a few of their free-text comments suggest this, that physicists might not quite see it that way, because arXiv was set up as an automated preprint (hardcopy) exchange system, which pre-dated the open access movement. However, what physicists do offer 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.

Results

Depositing rates

A definition of a digital repository was provided to help respondents understand the question and nearly two-thirds said they had deposited in such a repository (Table 1).

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

	n	%
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).

there are no great surprises why some respondents have not deposited, the modal response being that they are not aware of the existence of institutional repositories

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

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

	n	%
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

(about 7% 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 institutional repositories. These responses are uniformly distributed across the three age groups.

Reasons for depositing

The categories in Table 3 are overlapping to an extent (so, for example, someone whose

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	%
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	0.2
Other	26	2.4
Total	1,088	100.0

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.

Type of repository used

Respondents were provided with a definition of the various repository types and it is interesting to discover that more respondents say they have deposited in an institutional

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

Only respondents who have actually deposited

	n	%
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 specialized 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 specialized 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.

(44.1%) rather than a subject repository (39.7%), especially given the iconic status of the arXiv in the physical sciences (Table 4). Of course, only some types of physicists use ArXiv but those who do use it always use it and so the result may reflect authors depositing twice. Again, there are strong age-related effects here. The under 30s are 34% more likely than the 50 and older age group to have deposited in an institutional repository and 31% 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.

Type of content deposited

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

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 III (publishers' final edited version) seem almost equally acceptable.

There are large and significant age-related differences here. Older (50 plus) researchers

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

Only respondents who have actually deposited

	n	%
Journal article	694	64.1
Ph.D. 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	0.9
Video recording	4	0.4
Patent	4	0.4
Image or photograph	3	0.3
Metadata-only record	1	0.1
Other	21	1.9
Total	1,083	100.0

these findings suggest that younger researchers are taking institutional repositories (and the mandates that underpin them) very seriously

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

Only respondents whose last deposit was a journal article

	n	%
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

around three-quarters of depositors also make additional arrangements to provide access via their personal or institutional website

are 50% more likely than the youngest researchers to deposit a stage I manuscript and 73% 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.

Website availability

Around three-quarters of depositors also make additional arrangements to provide access via their personal or institutional website (Table 7). It would be interesting to find out more about this through qualitative research: why bother to post material on a website *and* a repository, as well as via the publisher – a case of touching all bases? Or to ensure that the formal and informal channels of distribution are covered? Surprisingly, there are no age-related differences in response to this particular question.

Table 7. Do you provide access to your research outputs from your personal or departmental website?

Only respondents who have actually deposited

	n	%
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

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 'website' 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 website approach (grey) is often forgotten.

Type of repository preference

Nevertheless, despite the popularity of websites, overall preference is to deposit via a more formal route: a subject or institutional repository (Table 8). Over 60% did so. This is somewhat at odds with the finding in Table 4 which showed that for most researchers the institutional repository was the place in which they deposited their most recent publication. Nevertheless, the subject repository was the preferred location. Just over a fifth (22%) expressed no preference, which again is interesting and needs to be followed up. Are they floating depositors? Younger researchers expressed a slightly stronger preference for institutional repositories and a slightly weaker preference for subject repositories than the older age

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	%
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 website	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

groups, but this is not statistically significant.

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 in this regard. A large majority, 84%, had (Table 9), and most of the others were not sure they had used repositories as a source of information.

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

All respondents

	n	%
Yes	1,418	83.7
I'm not sure	275	16.2
No	2	0.1
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, even though the questions again reminded them of the differences. It may also reflect the strength of support for arXiv.

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

Relative repository performance

Table 11 clearly shows 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

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

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

we should perhaps not assume that respondents fully understand the difference between these repository types

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

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.

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'; there is no difference by age group here. There are, however, some interesting and statistically significant age differences in response to this question:

- 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).

With 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

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

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 open access)	+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

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

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).

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 the attention of researchers 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 open access 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.

In a parallel question (Section 2.1.16), we asked library directors for their views on the likely trends for institutional and subject repositories over the next three years. There is a clear divergence of views with regard to institutional repositories: 49.8% of research-

ers believe they will become more, or much more important, compared with 76.6% 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.

Researchers are much more bullish about the future prospects for subject-based repositories, with 70.2% (as against with 49.8% 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%, respectively).

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

Free-text comments

The questionnaire ended by asking researchers to add anything they liked or disliked in connection with repositories. And they did, in droves – there were over 500 comments. It was certainly not the stolid topic we thought it would be. Instead it seemed the kind of issue guaranteed to draw out every platitude, worry, concern, and compliment that could possibly made in regard to the functioning of the scholarly communication system. Because of the general quality and wide-ranging aspects of the remarks and their off-the-cuff quality, we have reproduced a large number of them in the final

the main issue that divides the generations is a tendency for younger researchers to agree with the proposition that gold open access will gradually make digital repositories redundant

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

Free-text comment

Librarians and other 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, aged 50–54, USA)

report[5] and provide an example below. 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 several 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.

the resources associated with repositories are very modest indeed

Key findings from the library director survey

The library director study contained some important findings for publishers:

1. The resources associated with repositories are very modest indeed, certainly by publishing standards. Two-thirds of repositories can be categorized as 'small', having only one or two people working on 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. Repositories are mainly funded within the library. Recurrent spending on digital repositories is minuscule, averaging only 1.8% of library operational budgets.

2. Repositories are not just seen as a second-rate collection of journal articles. Collection policies and ambitions are much wider than this.

3. Repositories are not thought, by library directors, to herald a major reform of scholarly communication and publishing system.

4. 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.

Conclusion

High deposit rates were found but this can probably be put down to the high proportion of physical scientists in the sample. Of those that did not deposit, the main reason given was that they did not know about repositories. Voluntary deposit was the main reason for deposit, and this was followed by mandates. It does appear that institutional funding mandates and peer pressure are extremely helpful in terms of driving deposition behaviour. Young researchers were more likely to abide by a mandate; demographics could therefore work to the advantage of a mandated repository system in future.

Journal articles and e-theses were the main types of document deposited, and articles most commonly deposited at stage I, especially among older researchers. The stamp of peer review appears to be much more important for younger researchers, who perhaps have less confidence depositing materials that have not been reviewed.

The large majority of depositors made their articles available on their own or departmental websites, but only a minority made all of them available. The overall preference, though, was to deposit via a more formal route: a subject or institutional repository.

Anecdote has it that repository websites, especially institutional repositories, perform poorly compared to the best of the publisher platforms, but in fact researchers thought the quality of content, speed of response, and ease of use to be very good. It was only

the breadth, depth, and ease of navigation that they were not so happy with. Of course, they might have been answering the question with arXiv and PubMed Central in mind.

Researchers viewed 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'. Researchers, especially the younger ones, thought that repositories could be beneficial to publishers by virtue of creating greater digital visibility for research content. Young researchers were more inclined to believe that repositories contributed to the reform of scholarly communication and publishing.

There were two main complaints in respect to repositories: the variable quality of materials and insecurity over their long-term viability. Younger researchers were much more concerned about the potential for confusion caused by different versions of the same article being in circulation.

The main issue that divides the generations of researchers is a tendency for younger researchers to agree with the proposition that gold open access will gradually make digital repositories redundant.

Comparing the responses of the users of repositories with those that provide them (library directors) provides some interesting results:

- 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 any 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.

What, then, is the answer to the research question we posed for ourselves at the outset – 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, though digital repositories as a whole are still in a formative stage, they are 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. However, it does appear, from the free-text comments provide by respondents (and, indeed, CIBER log studies[18]) that some subject repositories, like arXiv and PubMed Central, certainly have come of age. Finally, as far as institutional repositories are concerned, in spite of the mandates, and even in spite of the fact that they are getting some traction, a scan of the recent literature shows the vocal movement associated with their implementation seems to have gone quiet for the last six years, leastways in the UK. This is puzzling.

digital repositories 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

References

1. Poynder, R. 2011. Open Access by Numbers, Open and Shut, http://poynder.blogspot.com/2011/06/open-access-by-numbers.html (accessed 20 March 2012). Laakso, M., Welling, P., Bukvova, H., Nyman L., Björk B.-C., et al. 2011. The Development of Open Access Journal Publishing from 1993 to 2009. *PLoS ONE*, 6:6, e20961. doi:10.1371/journal.pone.0020961
2. www.opendoar.org/countrylist.php (accessed 12 January 2012).
3. Covering dissertations or data, for instance.
4. www.peerproject.eu/ (accessed 12 February 2012).
5. CIBER. 2012. Coming of Age? Strategic Developments for Digital Repositories. Third Charleston

Observatory Report. http://ciber-research.eu/download/20120306-Charleston_Coming_of_Age.pdf (accessed 20 March 2102).

6. Conference of European National Libraries. 2005. Statement of the Development and Establishment of Voluntary Deposit Schemes For Electronic Publications. Paper presented at the Annual Conference of European National Libraries, Rome. http://ec.europa.eu/information_society/activities/digital_libraries/doc/consultations/replies/fep_a302619–1.pdf (accessed 12 April 2012).

7. Jones, R., Andrew, T., and MacColl, J. 2006. The Institutional Repository. Witney, Chandos Publishing.

8. Sale, A. 2007. The patchwork mandate. *D-Lib Magazine*, 13(1–2). www.dlib.org/dlib/january07/sale/01sale.html (accessed 20 March 2012).

9. Davis, P.M. and Connolly, M.J.L. 2007. Institutional repositories: evaluating the reasons for non-use of Cornell University's installation of DSpace. *D-Lib Magazine*, 13(3–4). www.dlib.org/dlib/march07/davis/03davis.html (accessed 20 March 2012),

10. Cullen, R. and Chawner, B. 2010. Institutional repositories: assessing their value to the academic community. *Performance Measurement and Metrics*, 11(2): 131–147.

11. Creaser, C., Fry, J., Greenwood, H., Oppenheim, C., Probets, S., Spezi, V., and White, S. 2010. Authors' awareness and attitudes toward open access repositories. *New Review of Academic Librarianship*, 16(S1), 145–16.

12. McKay, D. 2007. Institutional repositories and their other users: usability beyond authors. *Ariadne*, 5. www.ariadne.ac.uk/issue52/mckay/ (accessed 20 March 2012).

13. University of California. 2007. Faculty Attitudes and Behaviors Regarding Scholarly Communication: Survey Findings from the University of California, Berkeley, University of California Office of Scholarly Communication and the California Digital Library eScholarship Program. http://osc.universityofcalifornia.edu/responses/materials/OSC-survey-full-20070828.pdf (accessed 20 March 2102).

14. Lercher, A. 2008. A survey of attitudes about digital repositories among faculty at Louisiana State University at Baton Rouge. *Journal of Academic Librarianship*, 34(5): 408–415.

15. Jean, B.S., Rieh, S.Y., Yakel, E., and Markey, K. 2011. Unheard voices: institutional repository end-users. *College & Research Libraries*, 72(1), 21–42.

16. Connell, T.H. 2011. The use of institutional repositories: the Ohio State University experience. *College and Research Libraries*, 72(3): 2532274.

17. Nicholas, D., Rowlands., I, Jubb, M., and Jamali, H.R. 2010. The impact of the economic downturn on libraries: with special reference to university libraries. *Journal of Academic Librarianship*, 36(5): 376–382. Nicholas, D., Rowlands, I. 2011. Social media use in the research workflow. *Information Services and Use*, 31(1–2): 61–83.

18. Nicholas, D., Williams, P., and Rowlands, I. 2010. Researchers' e-journal use and information seeking behaviour. *Journal of Information Science*, 36(5): 494–516.

David NICHOLAS,
Anthony WATKINSON, David BROWN
CIBER Research Ltd
I Westwood Farmhouse
Newbury RG14 7RU, UK
Email: office@ciber-research.eu
http://ciber-research.eu

Ian ROWLANDS

David Wilson Library
University of Leicester, PO Box 248
University Road, Leicester LE1 9QD, UK
Email: ir46@leicester.ac.uk

Hamid R. JAMALI
Department of Library and Information Studies
Faculty of Psychology and Education
Kharazmi University,
No 49, Mofateh Ave, POBox 15614
Tehran, Iran
Email: h.jamali@gmail.com