

RoMEO Studies 3 - How academics expect to use open-access research papers

Elizabeth Gadd, Charles Oppenheim, and Steve Proberts

Department of Information Science, Loughborough University, Loughborough, LE11 3TU

This article has been accepted for publication by the Journal of Librarianship and Information Science.

Abstract

This paper is the third in a series of studies emanating from the UK JISC-funded RoMEO Project (Rights Metadata for Open-archiving). It considers previous studies of the usage of electronic journal articles through a literature survey. It then reports on the results of a survey of 542 academic authors as to how they expected to use open-access research papers. This data is compared with results from the second of the RoMEO Studies series as to how academics wished to protect their open-access research papers. The ways in which academics expect to use open-access works (including activities, restrictions and conditions) are described. It concludes that academics-as-users do not expect to perform all the activities with open-access research papers that academics-as-authors would allow. Thus the rights metadata proposed by the RoMEO Project would appear to meet the usage requirements of most academics.

Introduction

The UK JISC-funded RoMEO (Rights Metadata for Open-archiving) project has been tasked with investigating the rights issues associated with the self-archiving of research papers by academics, and the subsequent disclosure and harvesting of metadata about those research papers using the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH). One of the key aims of the project is to produce some simple rights metadata elements that academics can use to describe the terms and conditions they want to impose on the use of their research papers in an open-access environment. To inform the development of this metadata, the project team performed a survey of 542 academic authors to ascertain both how academic authors wished to protect their own self-archived research papers, and how they expected to use such research papers. The survey is described in full in the first two papers in the RoMEO Studies series (Gadd, Oppenheim and Proberts, 2003a, and 2003b), hereinafter referred to as RoMEO Studies 1 and RoMEO Studies 2.

RoMEO Studies 2 reported on how academics wished to protect their *own* open-access research papers. This, the third in the series, looks at how academics wish to use *others'* papers. It also considers whether there is any significant difference between the way academics-as-authors wish to protect open-access research papers, and how academics-as-users wish to use them. This will show, in turn, whether the rights metadata developed by the project to protect research papers, informed by academics-as-authors, will actually meet the usage needs of academics-as-users.

Previous research paper usage studies

The term “usage” is a broad one that can be interpreted in a number of different ways. Copyright law defines a range of usage activities (copy, broadcast, lend, etc.) some of which exclusively belong to the copyright owner, and some that may be performed by non-copyright owners under certain terms and conditions. Lessig (2003) calls the former set of activities “copyright” activities and the latter set, “regulated”. There is also a third category of activities with which copyright law does not concern itself and Lessig terms these “unregulated” activities. The usage activities of interest to the RoMEO Project were those performed with *open-access electronic research papers* that fell into Lessig’s “copyright” and “regulated” categories. (For the purposes of this paper, such activities will be referred to as “copyright-regulated”). It was not concerned with unregulated activities. This was because we anticipated that the resulting rights metadata would provide less restriction on use than that provided by copyright law, but more than that provided by simply releasing a work into the public domain. If the RoMEO rights metadata began regulating activities that were unregulated under copyright law, this would go against the spirit of open-access.

A search of the literature showed that, to date, investigations into the usage of open-access research papers on eprint archives have tended to focus on the depositing activities of authors, rather than usage by end-

users (Carr et al, 2000; Pinfield, 2001; and Hunt, 2002). Thus, it was necessary to turn to the literature on the use of *closed access* electronic research papers, in particular, electronic journal articles, to find some data on how research papers were used by academics.

The usage of electronic journals has been a pre-occupation of librarians and publishers alike since the beginning of the digital age. As early as the 1980's, projects like BLEND (Birmingham and Loughborough Electronic Network Development) (Shackel, 1991) began to investigate the feasibility of the electronic journal as a scholarly communication medium. The Quartet project (Tuck, 1990) followed on with a broader scope of study, examining the impact of a wide range of electronic communication technologies on the scholarly communication process. However, key to the development of the e-journal was the human factor: how were end-users likely to use electronic journals? Would this differ from their use of printed journals?

In the mid-1980's Simpson (1988) performed a study of the usage of printed journals in an effort to provide an answer to this question. She used a questionnaire to ascertain the source of researchers' references; the number of personal journal subscriptions held; methods of obtaining articles (i.e., through inter-library loan, reprints, etc.); how articles were selected for reading; and reading habits (time, location, reading strategy, etc.). However, only one question, "How articles are used" gathered any information on copyright-regulated activities that were performed with journal articles. The most common responses were: taking a photocopy, making notes, and writing a brief abstract.

With similar aims in mind, Olsen (1994) performed a study six years later to answer the research question: "What are the particular processes carried out by scholars with journal literature which are so fundamental to scholarship that they must be accommodated by the electronic version of those journals?". She performed a series of 46 interviews with researchers at two major US universities. Her interview schedule was divided into five main headings, as follows:

- Reasons and techniques for locating journal literature;
- Methods of reading journal literature;
- When and where journal literature is read;
- Which journal literature is most useful;
- Speculation on the advantages and disadvantages of printed journal literature compared with the electronic journal.

Again, only one question considered researchers' usage of journal articles in broader terms than reading. That was, "What do you do with the information in the article as you are reading it, and after you have read it? E.g. underline...photocopy parts of the full article,...use it in your class as a handout, use parts of it as overhead transparencies, put it on reserve." However, very few results were reported: 91.3% of scholars photocopied important articles, and 69.0% annotated article or underlined articles. No further findings were discussed.

One of the first studies of usage of actual e-journals was the Café Jus project (Woodward et al, 1997). The project used structured questionnaires and log sheets to try to ascertain how academic researchers used e-journals, and what they thought of them. They were disappointed with the response rate to their log analysis, but managed to ascertain, via the questionnaire, what users thought about the content, appearance, and facilities of e-journals. Some of their findings have been outdated by the increase in the availability of e-journals and associated technology. However, one of their key results, to be triangulated many times over by subsequent research, was that users did not like reading on screen, and would instead print the relevant article out. Again, this was the only result reporting usage regulated by copyright law.

An overlapping study, the SuperJournal Project (1996-1998) performed a three-year analysis of:

“the factors that will make electronic journals successful. The objective was to identify the features that deliver real value, and to explore the implications with stakeholders in the publishing process: readers, authors, libraries, and publishers.” (Superjournal, 1999).

The project established clusters of e-journals in various subject disciplines and analysed usage patterns and user preferences through a number of different methodologies. One method was an analysis of transaction logs (Eason, Richardson and Yu, 2000; Yu and Apps, 2000) which gave detailed information on who was accessing what, when and from where, but did not focus on any other aspects of usage. Indeed, transaction log analyses rarely focus on copyright-regulated activities other than the printing of whole documents (Nicholas and Huntington, 2002).

Ten per cent of their users were asked about their usage through interviews, questionnaires and focus groups which provided further data. However, this tended to focus on the interaction with the e-journal itself (use of browse and search to locate articles, depth of use in terms of mining to table of contents, abstract or article level) and whether articles were printed, read on-screen or both (Eason and Harker, 2000). Their “questionnaire study of repeat users found...that 89% would print rather than read the article on-screen. This result was replicated in an interview study of 52 repeat users”. This led them to conclude that “the electronic journal revolution is largely an exercise in printed document delivery, shifting the printing from the publisher to the end user”.

More recent work on the usage of e-journals has been undertaken by Tenopir and King (2000, 2001, 2002). Their seminal monograph on the history, use and economics of scholarly journals, *Towards Electronic Journals*, reported data from (amongst others) three national surveys of scientists in 1977 and 1984 on the readership of scholarly journals, and a series of statistical surveys of scientists from 1981 through 1998 on user activities such as reading and communication patterns. They have provided the community with impressive and comprehensive results on the reading activities of university scientists including:

- Average number of scholarly article readings per year;
- Average time spent reading scholarly scientific journals;
- Readers’ goals in reading – e.g. current awareness, to support a grant proposal, writing a paper;
- How the articles are read (e.g., abstract first, etc – all or parts).

With regard to readers’ goals, they write:

University scientists were asked several questions concerning the last article they read. One critical incident question concerned the purposes for which they used or planned to use the information. Over one-half of the readings were for current awareness or professional development. When applied to work, the information is frequently used to support research (75 percent) and teaching (41 percent). Administration accounted for 13 percent of the total. Brown (1999) also found that scientists at the University of Oklahoma relied on journals more for research than for teaching.

The purpose of reading provides some insight into the copyright-regulated activities that may be performed with research papers. For example, papers used in teaching may be photocopied and distributed to students, whereas papers used for research may be more likely to be retained and annotated. Tenopir and King (2000) have also done a lot of work on information-seeking and readership patterns amongst university scientists. In particular, they have focussed on:

- How readings are identified (e.g., browsing, searching, citations, other people etc.)
- The source of the journal used (e.g., library, personal subscription, etc.)
- The proportion of articles in a journal read per year

However, notwithstanding a section on the photocopying of scholarly articles (75% of respondents photocopied journals), there was no real data on the copyright-regulated usage of *electronic* journal articles. Other work by Tenopir and King (2001) addresses this matter to a certain extent. Their paper, *Electronic journals: how user behavior is changing*, compares the usage of journals by researchers at the Oak Ridge National Laboratory (ORNL) in 1984 with a survey of the same organisation in 2000-1. They consider the take-up of electronic journals in 2000-1 as opposed to printed journals in 1984; methods of identifying relevant journal articles; the volume of reading (an increase in reading of recent articles was attributed to an increase of material available electronically); and the time spent browsing, searching, downloading and printing out. They make an interesting observation that,

“the proportion of print articles photocopied was about 50% compared with 62% of electronic/digital articles downloaded/printed out. The time spent photocopying was about 3 min compared with 4.5 min spent downloading/printing.”

Although they do not consider the activities performed with electronic journal articles themselves, they cite Richardson (1981), Schauder (1994), and Simpson (1988) all of which found that respondents preferred to read print-outs rather than articles on screen.

Despite the comprehensive approach to their studies on electronic journals, Tenopir and King, like others working in this field, appear not to have given much consideration to the activities regulated by copyright law that are being performed with electronic research papers.

E-journal statistics

Usage of electronic research papers is currently occupying librarians and publishers in another regard, as they battle to generate meaningful usage statistics for electronic journals (Kidd, 2002). However, as Luther (2000) writes, “Librarians seek to collect usage data that justify the library’s investment in electronic resources.” This influences the statistics required and subsequently gathered. In 1998, the ICOLC (2001) produced some *Guidelines for statistical measures of usage of web-based indexed, abstracted and full-text resources* which were revised in 2001. The data they suggested should be gathered included:

- Number of sessions (Logins);
- Number of queries (searches);
- Number of Menu Selections (e.g. an electronic journal site provides alphabetic and subject-based menu options in addition to a search form. The number of searches and the number of alphabetic and subject menu selections should be tracked);
- Number of full content units examined, downloaded, or otherwise supplied to the user;
- Number of turn-aways.

It was not clear whether the number of full content units examined should be reported as a single figure, or divided up according to activity. In 2000, Luther recommended that four categories of usage information should be collected. These included:

- What is being used? (content)
- Who is using the content? (user)
- How is the database used? (activity)
- When is the content being used?

However, she does not suggest what ‘activities’ (the third bullet point) are measured.

The most recent investigation into usage statistics has emanated from Project COUNTER (2003), supported by a coalition of library and publisher groups. With respect to electronic journals, they recommend that only two figures are provided: the number of full-text article requests and the number of ‘turnaways’. Print-outs, downloads, and so on, are not listed. However, it should be pointed out that the focus of this project was to

develop a Code of Practice that a wide range of publishers were able to sign up to, rather than a ‘wish list’ generated by librarians.

Thus, the primary focus of usage statistics is not to be the copyright-regulated activities performed with the full-text journal articles. One of the reasons for this may well be a semantic difficulty with describing the various activities. As Cowhig (2001) notes, “definitions and measures of ‘download’ may vary; definitions and measures of a ‘paper’ may vary...”. Indeed, while he later suggests that “The key figure for any journal might be *average downloads per paper* over a defined time”, it is not clear whether, by ‘download’, he means display, save to disk, print out, or all three. Strictly, it means saving to disk.

Although information on the activity performed by an end-user may not be of interest to librarians, such measures could be used by e-journal service providers to police illegal use of full-text. However, most services would only be able to detect what happens to the e-version of the article hosted at the vendor’s site. They could not (obviously) provide information on what happens to resulting electronic or printed copies – with which any illegal transactions are most likely to take place. It is perhaps for this reason that most electronic journal services are governed by a combination of licence agreements and technical measures rather than solely by technical measures.

Summary of previous e-journal usage studies

We can conclude that despite the interest in e-journal usage as a research topic, there has been no study to date focussing on the activities performed with research papers that may or may not be regulated by copyright law (printing, saving, annotating, lending, etc). Most e-journal usage research seems to focus on nine main areas:

1. Who uses e-journals;
2. What types of e-journals are the most useful (e.g. trade, scholarly, etc.);
3. How they are located (source of references and the source of full-text);
4. When they are used;
5. Where they are used;
6. How often they are used (frequency and volume of use);
7. Why they are used (reasons and purposes for using e-journal literature);
8. User interaction with the electronic version (navigation techniques etc);
9. Reading methods (progression through the article – abstract first, for example).

Apart from question seven “Why they are used”, to which the data reported has been fairly thin, such studies have not focused on the types of activity performed with a journal article once located and read. This is partly due to the aims of many early studies to inform the development of the e-journal, and later usage statistics studies to inform collection management activities. It may also be due to the lack (until recently) of a usable taxonomy by which to classify such activities.

Methods adopted

One of the beneficial side effects of e-commerce and the ensuing need to develop Digital Rights Management (DRM) systems, has been the efforts taken to develop digital rights expression languages (DREs) incorporating rights models and data dictionaries. To provide a framework for collecting information on the ways academics wished to protect and use open-access research papers, we used the Open Digital Rights Language (2002) (ODRL) which provides a model of permissions, restrictions and conditions (P, R and C’s) over works. A restriction is a limit on the extent of the permission being offered (e.g., you may print, but only four times), whereas a condition is a prerequisite under which the permission may be performed (e.g., you may print four times, if you pay a fee).

The ODRL terms (P, R and C’s) considered of most relevance to the use of academic research papers were selected (see RoMEO Studies 2 for full details of the selection process) for use in an online survey. The terms selected are listed in Tables 1, 2 and 3 below.

Permission (ODRL)	Wording in survey (including example)
--------------------------	--

Display	Display (e.g. may be viewed on screen)
Print	Print (e.g. copies may be printed out)
Modify	Modify (e.g. may be translated or a derivative work created)
Excerpt	Excerpt (e.g. a short passage may be quoted)
Annotate	Annotate (e.g. editorial or peer commentary may be inserted)
Aggregate	Aggregate (e.g. may be compiled into an anthology)
Sell	Sell (e.g. either on a cost-recovery basis or as a commercial enterprise)
Lend	Lend (e.g. a printed copy may be loaned by a library)
Give	Give (e.g. copies may be forwarded to colleagues)
Lease	Lease (e.g. may be loaned or licensed for a fee)
Duplicate	Copy (e.g. mount another copy on the web)
Save	Save (e.g. may be saved to disk)

Table 1 Permissions used in the survey

Restriction (ODRL)	Wording for survey (including example)
Individual	For personal use only (e.g. user can't distribute copies to others)
Group	For use by certain groups (e.g. only for use by educationalists)
Count	Limited number of times (e.g. print up to four copies, or save once)
Spatial	Limited to certain geographical regions (e.g. not to be used in countries with sanctions)
DateTime	Only available for a certain period of time (e.g. must be removed after 5 years)
Quality	Copies must be exact replicas of the original text (e.g. the text must not be altered in any way)
Format	Copies must be in the same format as the original (e.g. only PDF copies can be made)
Watermark	Existing watermarks or security features must be maintained (e.g. all copies must retain the security features of the original)
Purpose	Only for certain purposes (e.g. teaching, research or non-commercial)

Table 2 Restrictions used in the survey

Conditions (ODRL)	Wording for survey (including example)
Payment	Fee paid (e.g. a fee per copy, or pay for access)
Accept	Users must agree to certain terms and conditions (e.g. a click-through licence)
Register	Users must register (e.g. with a service provider)
Attribution	Author must be attributed (e.g. your name should always be clearly displayed on the article)
Tracked	Usage tracking (e.g. all usage of your article should be tracked)

Table 3 Conditions used in the survey

With regard to the permissions, respondents were invited to say whether they would expect to perform them freely, with limits or conditions, or not at all. For the restrictions and conditions, respondents either selected them or left them blank. The survey was advertised via a range of email discussion lists to academics across the world. RoMEO Studies 2 describes how we obtained our respondent sample.

Survey response rate and demographics

The number of survey respondents and their demographics has been reported in full in RoMEO Studies 1. However, to summarise, 542 responses were received from 57 countries. The majority (one-third) came from the UK. The USA, Australia, Canada and Germany contributed the next largest tranche of

respondents representing 17%, 4%, 3% and 3% of the total. It was impossible to calculate the response rate as the sample size was unknown.

Fifty per cent of respondents came from a pure science discipline, with 38% from a social sciences or humanities discipline, and 12% from engineering. The largest group of respondents (39%) had had more than 15 years service in academia, but there was a fairly even distribution of respondents with differing lengths of academic service. Each respondent had published an average (mean) of 42 papers, although the modal average was ten papers or less.

Use of open-access research papers

In RoMEO Studies 2, we reported that of 536 respondents, 310 had made research papers freely available on the web and 226 had not. (Those that had were referred to as ‘archivers’, and those that had not as ‘non-archivers’). Respondents were later asked; ‘Have you used other people’s research papers that have been made freely available on the web?’. Exactly 530 responded, of which 67 (12%) said no and 463 (88%) said yes. Thus, considerably more respondents had used others’ freely available research papers than had made papers available themselves.

A cross-tabulation was performed between the data from these two questions (“Have you made research papers freely available on the web?” and “Have you used others’ freely available research papers on the web?”) to see if those that self-archived their own papers were more likely to use others’ self-archived works. This showed that of the 310 that had previously self-archived, 293 (or 95%) had also used others’ self-archived work. Of the 226 that had not self-archived, 167 (or 74%) had used others’ self-archived work. The data therefore suggests that those who have previously self-archived are (perhaps not surprisingly) more likely to have used self-archived materials than those who have never self-archived. However a high proportion of those that had not self-archived themselves (75%) were clearly reaping the benefits of the self-archiving activities of others.

Location of open-access research papers

Respondents were asked where they had located others’ freely available research papers. Unfortunately, whereas we had provided respondents with five options for a previous question, “Where have you made your own research papers available?”, only four options were provided for this question, “Which archives or services have you used to access such papers”. The “e-journal” option was missing. The responses that were given are illustrated in Figure 1.

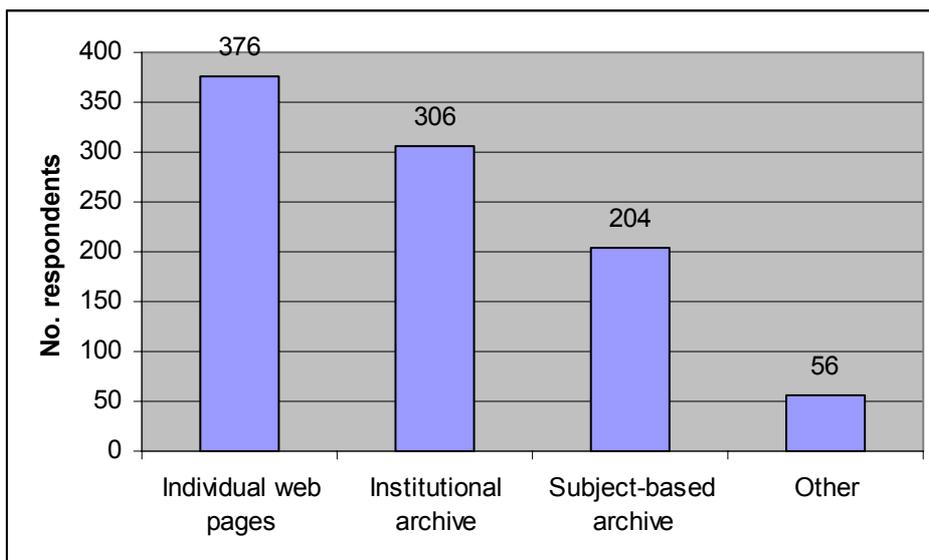


Figure 1 Where respondents had located freely available research papers

The usage of papers via individuals' web pages was most common. Eighty-one per cent of respondents that had used freely available research papers had accessed them through individuals' own web pages. Interestingly, whereas just 48 respondents had made papers available on their own institutional archives, 306 respondents (66% of responding users) had used papers from this source. This anomaly may result from the slightly different wording of the two questions. The earlier question asked if respondents had made papers available on "a repository or archive run by *your* institution" whereas this question asked if respondents had used papers on "a repository or archive run by *an* institution". It did not specify an academic institution. Just over 200 respondents (44%) had used papers from a subject-based archive.

Fifty-six respondents indicated that they used some 'other' source for locating research papers. Seventy-one gave further details of the sources they used. Not surprisingly 31 respondents made mention of electronic journals of some description. However, interestingly, 11 were explicitly talking about subscription-based e-journals rather than free e-journals. Sources such as JSTOR, Ingenta and Science Direct were mentioned. A further 13 respondents just mentioned 'online journals', which could of course mean free or subscription-based journals. The remaining seven were explicitly talking about free online journals such as First Monday, BMJ, and BioMed Central. Four respondents said they used publisher sites that made journals freely available for a time.

Other sources included email newsgroups through which authors circulated preprints; online conference proceedings; and 'portals' or 'gateways' such as those run by professional bodies or government organisations. Four respondents simply used web search engines to locate research papers and hadn't taken note of the source of those papers.

Usage of open-access research papers

Respondents were asked how they expected to use other peoples' freely available research papers. They were given the same matrix of activities and options as in a previous question regarding how they would like to protect their own papers. The results are illustrated in Figure 2 below.

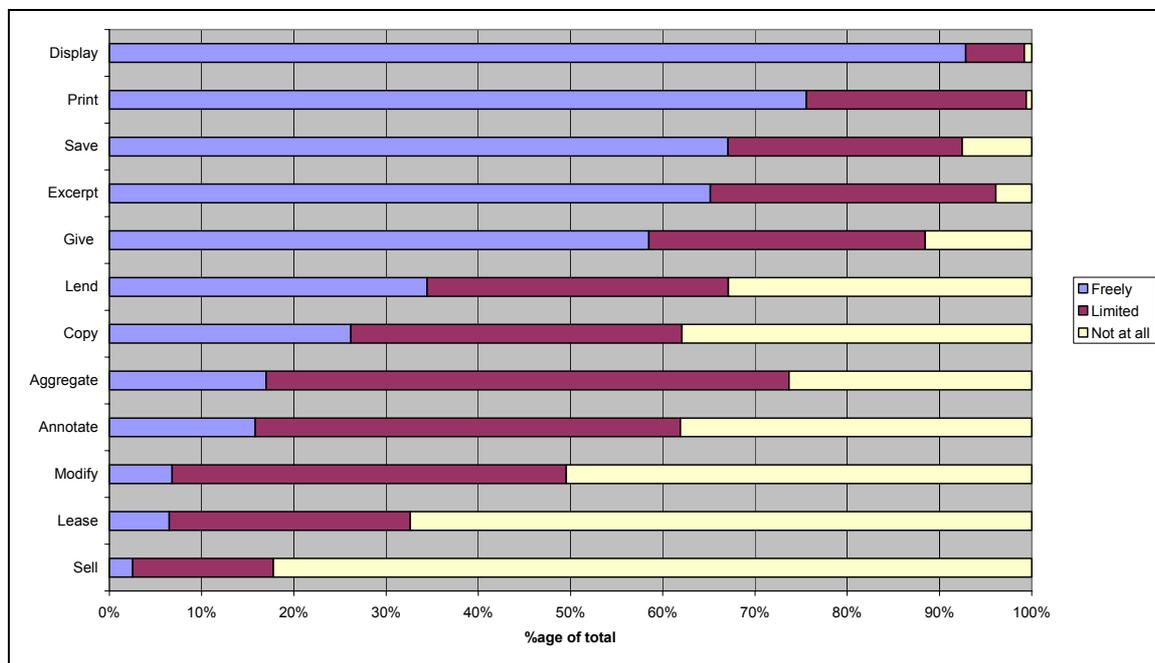


Figure 2 How respondents expect to use others' research papers

The data table below indicates in **bold** the most common preference (i.e., freely, limited or not at all) for each permission. Response rates varied from 432 for the 'Lease' option to 508 for the 'Display' option. These figures may themselves indicate the importance or relevance of a particular activity to academics.

	Freely	Limited	Not at all	TOTAL
--	---------------	----------------	-------------------	--------------

Aggregate	75	250	116	441
Annotate	71	207	171	449
Copy	116	159	168	443
Display	468	32	4	504
Excerpt	318	151	19	488
Give	279	143	55	477
Lease	28	113	291	432
Lend	154	146	147	447
Modify	30	188	222	440
Print	384	121	3	508
Save	330	125	37	492
Sell	11	66	356	433

Table 4 Permissions data table

Table 5 lists the permissions and options in order of greatest agreement amongst respondents.

Permission	Percentage agreeing	Option	Rank
Display	93%	Freely	1
Sell	82%	Not at all	2
Print	76%	Freely	3
Lease	67%	Not at all	4=
Save	67%	Freely	4=
Excerpt	65%	Freely	6
Give	58%	Freely	7
Aggregate	57%	Limited	8
Modify	50%	Not at all	9
Annotate	46%	Limited	10
Copy	38%	Not at all	11
Lend	34%	Freely	12

Table 5 Permissions in order of agreement

Figure 2 shows that over 90% of respondents expected to be able display, print, save and excerpt either freely or under limits or conditions. Just under 90% expected to give works away (the example used was “copies may be forwarded to colleagues”). Figure 2 also illustrates that aggregate, despite having a small number that expected to do this freely, ranked fifth after ‘give’ when the ‘freely’ and ‘limited’ percentages were combined. Aggregate and annotate were the only two permissions that the largest group of respondents expected to perform with limit or condition (Table 5), but never-the-less, over 60% of respondents expected to do so with open-access research papers (Figure 2). Table 5 shows that there were four permissions that the largest group of respondents agreed that they would not perform at all. These were sell (82%), lease (67%), modify (50%) and copy (38%). As discussed in RoMEO Studies 2, we believed that there was a semantic problem with the term ‘copy’. It could mean either to cheat, or to make a reproduction.

Permissions comparison: academics-as-users and authors

A comparison of this data on how academics expected to use others’ works (i.e., academics-as-users) with the data from an earlier question as to how academics expected others’ to use *their* works (i.e., academics-as-authors) was performed by completing chi-square tests on all the permissions data (i.e., the name of the permission and the numbers selecting freely, limited, and not at all) collected from each group. Any result at the 0.05 level or below was deemed to be significant. The results are listed in Table 6.

Permission	Chi-square value	Significance (2 degrees of freedom)
Save	2.0799	Insignificant
Display	2.9327	Insignificant
Excerpt	3.5253	Insignificant
Print	8.0534	0.025
Modify	21.7917	0.01
Give	28.6931	0.01
Copy	30.948	0.01
Annotate	46.1658	0.01
Aggregate	60.52	0.01
Lease	83.505	0.01
Sell	90.613	0.01
Lend	152.2122	0.01

Table 6 Differences between permissions data from academics-as-authors and as-users

The tests revealed highly significant differences (mainly to the 1% level) between the two groups on all permissions except display, excerpt and save. As the rights metadata solution was based on the option (freely, limited or not at all) that the largest group of respondents had selected for each permission, a comparison of the percentage of academics-as-authors, and as-users agreeing with the ‘majority’ option for each permission is given in Table 7.

Permission	Academics as users		Academics as authors	
	%age agreeing with majority option	Rank	%age agreeing with majority option	Rank
Display	93% Freely	1	90% Freely	1
Sell	82% Not at all	2	53% Not at all	8
Print	76% Freely	3	73% Freely	2=
Lease	67% Not at all	4=	42% Limited or conditional	11
Save	67% Freely	4=	65% Freely	6
Excerpt	65% Freely	6	69% Freely	4
Give	58% Freely	7	73% Freely	2=
Aggregate	57% Limited or conditional	8	58% Limited or conditional	7
Modify	50% Not at all	9	50% Limited or conditional	9
Annotate	46% Limited or conditional	10	48% Limited or conditional	10
Copy	38% Not at all	11	39% Freely	12
Lend	34% Freely	12	68% Freely	5

Table 7 Comparison of permission data

The results of this analysis can be broadly divided into two categories: permissions on which respondents to both questions roughly agreed, and permissions on which academics-as-authors were far more liberal in what they would allow compared to what academics-as-users actually expected to do.

Agreed

There were six permissions on which the largest proportion in each group agreed. However, on three of them (display, print and save), slightly more academics-as-users expected to perform them, than academics-as-authors were willing to allow them (3%, 3% and 2%). On the other three (excerpt, aggregate and

annotate) slightly more academics-as-authors were willing to allow them, than academics-as-users expected to perform them (4%, 1% and 2%).

Academics-as-authors more liberal

Overall, academics-as-authors tended to be far more liberal about the permissions they would allow others to perform with their works than academics-as-users expected to be permitted to make use of. In addition to the three categories above (excerpt, aggregate and annotate), the remaining six permissions found authors more generous than users required. Academics-as-users generally concurred that they did not expect to sell other people’s papers at all. However, just 53% of academics-as-authors disapproved of their papers being sold. Again, while 79% were happy for others to give their works away freely, just 58% expected to do this freely with others’ works. Similarly, 68% were happy for their works to be loaned, but only 34% expected to lend works.

For the remaining three permissions, the largest proportion in the author group selected a more liberal option (e.g., freely or limited), whilst the largest proportion in the user group stated that they did not expect to perform that option at all. Thus, while 67% of academics-as-users agreed that they would not lease others’ works, a 42% ‘majority’ of academics-as-authors were happy to allow limited leasing of their own works. Fifty per cent of academics-as-users did not expect to modify works at all, whilst 50% of academics-as-authors were happy to allow modification under limit or condition. An even starker contrast was found in response to the copy permission, where 38% of academics-as-users did not expect to copy at all, but 68% of academics-as-authors were happy to allow copying freely. There may however have been some confusion over the meaning of the term “copy” as previously discussed.

Restrictions and conditions

In total, 504 respondents answered the question inviting them to specify what restrictions and conditions they would expect to be in force over other people’s self-archived works.

Restrictions

Figure 3 illustrates the restrictions that academics-as-users expect to adhere to when using others’ freely available research papers.

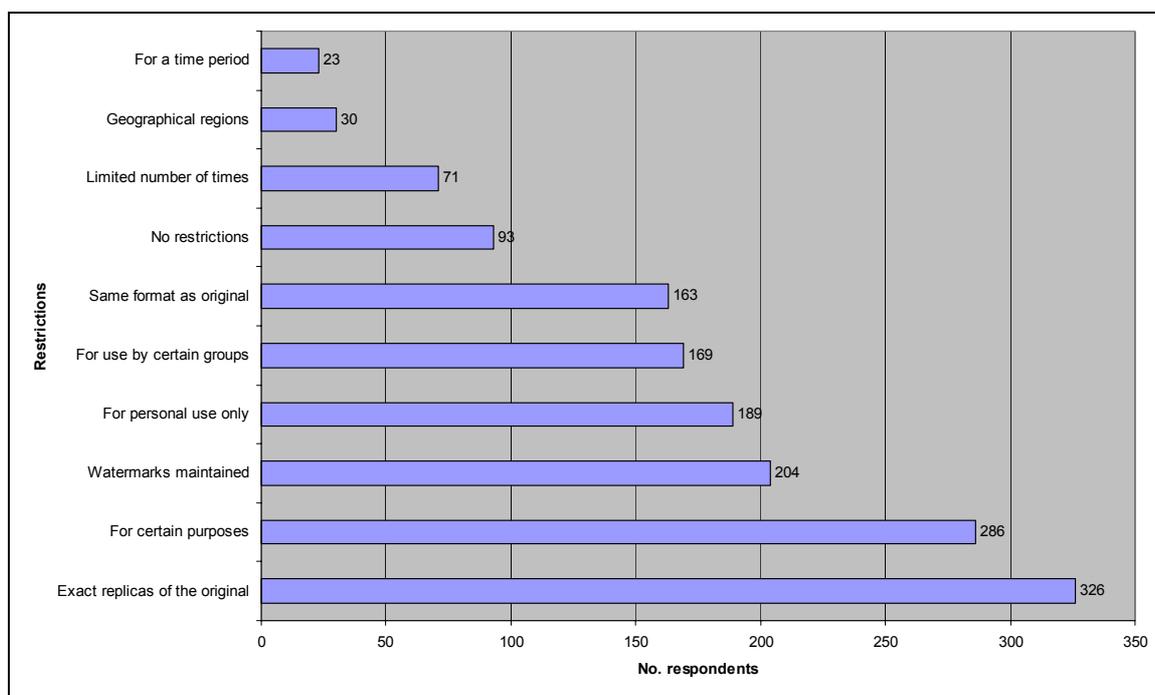


Figure 3 Restrictions academics expected to be in place over others’ self-archived works

Only 18% of respondents stated that they expected there to be no restrictions on the use of open-access works. The remaining 82% believed there would be restrictions on use. The largest groups of respondents (64%) expected that all copies should be exact replicas of the original, and that they should only be used for certain purposes (56%). Less than 50% of respondents expected the other restrictions to be in place. However a significant 40% expected to maintain watermarks, and 38% expected to use the paper for personal use only. About one-third thought that usage was limited to certain groups and that copies should take the same format as the original. Less than 15% expected usage to be limited by time (time period or number of uses) or space (geographically).

Restrictions comparison: academics-as-users and authors

Table 8 compares responses from academics-as-users with academics-as-authors.

	Academics as users			Academics as authors		
	No.	%	Rank	No.	%	Rank
Exact replicas of the original	326	64	1	355	67	1
For certain purposes	286	56	2	289	55	2
Watermarks maintained	204	40	3	182	34	3
For personal use only	189	38	4	167	31	5
For use by certain groups	169	34	5	144	27	6
Same format as original	163	32	6	175	33	4
No restrictions	93	18	7	106	20	7
Limited number of times	71	14	8	62	12	8
Geographical regions	30	6	9	30	6	9
For a time period	23	5	10	25	5	10
Total	504			528		

Table 8 Comparison of restriction data

It can be seen that academics-as-users agreed with academics-as-authors on the importance ranking of restrictions on research papers in seven out of ten cases. The main differences resulted from academics-as-users placing retaining the format of the original in sixth place, whereas academics-as-authors valued that slightly higher at fourth place. This was understandable, as authors have the greater interest in retaining the original format (e.g., PDF) of a paper, particularly if the format contains security features. Just two per cent more academics said they'd be happy with no restrictions on the use of their papers, than expected there to be no restrictions on the use of others' papers.

Again, chi-square tests were performed to compare responses from authors and users to see if there were any significant differences between the restrictions academics required over their own works compared with their actual usage of others' works. The results are given in Table 9.

Restrictions	Chi-square value	Significance (1 degree of freedom)
For a time period	0.017	Insignificant
Geographical regions	0.0344	Insignificant
Same format as original	0.0754	Insignificant
For certain purposes	0.4227	Insignificant
No restrictions	0.4366	Insignificant
Exact replicas of the original	0.7484	Insignificant
Limited number of times	1.2629	Insignificant
For personal use only	3.9337	0.05
Watermarks maintained	3.9734	0.05
For use by certain groups	4.7806	0.05

Table 9 Differences between restrictions data from academics-as-authors and as-users

In the majority of cases, there was no significant differences between the restrictions required by authors and by users. However, there was a difference at the 5% level for the “For personal use only”, “Watermarks maintained”, and “For use by certain groups” restrictions. A higher proportion of academics-as-users than academics-as-authors selected these restrictions. This may indicate that users show a respect for either written or unwritten regulations on the use of research papers, that they are not concerned about as authors. The one restriction that academics-as-authors were more concerned about than academics-as-users was that copies be exact replicas of the original. Three per cent more authors selected this restriction than users.

Conditions

The bar chart below illustrates the conditions that academics-as-users expect to apply when using others’ freely available research papers.

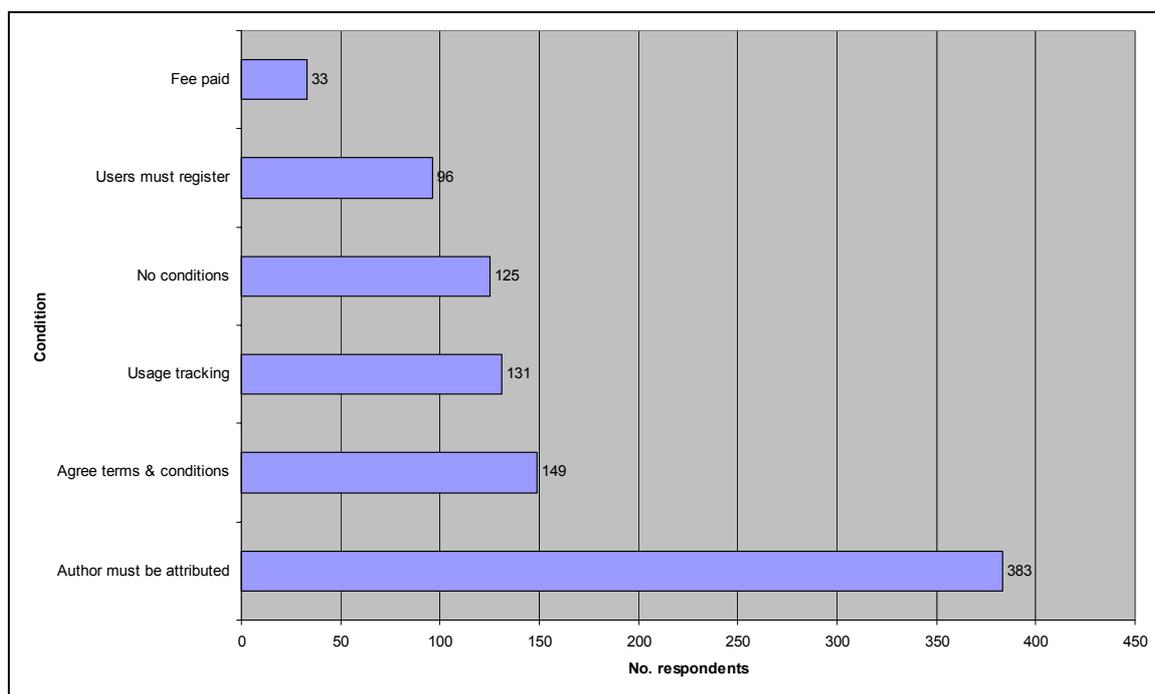


Figure 4 Conditions academics expected to be in place over others’ research papers

Slightly more respondents expected there to be no conditions on the use of open-access research papers (25%) than expected there to be no restrictions (18%). However, that still left the majority of academics (75%) expecting to adhere to conditions. Indeed some selecting “no conditions” must also have selected a condition as 76% believed the author should be attributed. No other condition received as much agreement as attribution. Under one-third expected to agree to terms and conditions and just over one-quarter expected their usage to be tracked. Nineteen per cent expected to have to register to use open-access works. Perhaps not surprisingly, only 7% expected to have to pay a fee.

Conditions comparison: academics-as-users and authors

Table 10 compares conditions required by academics-as-users to those required by academics-as-authors.

Condition	Academics as users			Academics as authors		
	No.	%	Rank	No.	%	Rank
Author must be attributed	383	76	1	432	81	1
Agree terms & conditions	149	30	2	164	31	3
Usage tracking	131	26	3	172	33	2
No conditions	125	25	4	111	21	4
Users must register	96	19	5	86	16	5

Fee paid	33	7	6	40	8	6
Total	504			528		

Table 10 Comparison of conditions data

Table 11 shows the results of chi-square tests to determine the level of difference between academics-as-authors and as-users on their choice of conditions.

Conditions	Chi-square value	Significance (1df)
Usage tracking	5.389	0.025
Author must be attributed	5.2709	0.025
No conditions	2.0875	Insignificant
Users must register	1.352	Insignificant
Fee paid	0.4146	Insignificant
Agree terms & conditions	0.2735	Insignificant

Table 11 Differences between conditions data from academics-as-authors and as-users

The results indicate that on only two conditions were there significant differences: on the matter of usage tracking and author attribution. Table 10 shows that whilst both groups ranked author attribution as top of their list of conditions, five per cent more academics-as-authors selected this condition, than academics-as-users. On the issue of usage tracking, academics-as-authors ranked this second with one-third of them selecting it. However, academics-as-users ranked it third with just over one-quarter selecting it.

What was surprising was that three per cent more academics-as-users expected to register to use research papers, than academics-as-authors. This may be based on academics' experience of registering for free or subscription-based e-journals. Interestingly, although it was not a significant difference, four per cent more academics (25%) expected there to be no conditions over their use of others' freely available research papers, than were happy for there to be no conditions on the use of their own papers.

Discussion

Use of freely available research papers by non-archivers

Seventy-four per cent of those that had not self-archived papers themselves had actually used other peoples' freely available works. This could be taken as an encouraging sign for advocates of self-archiving in that three-quarters of non-users are actually seeing and making use of the benefits of freely available research. As such, they may themselves be won over to self-archiving. The ALPSP have reported a similar phenomenon (Swan, 2002). Although only 11% of their author survey respondents had deposited with preprint archives themselves, 32% thought they were important to their subject discipline. Similarly, 11% had deposited with reprint archives, but 62% felt they were important to their discipline. This indicates that a considerable percentage of academics are waiting to be won over.

Location of open-access research papers

The largest proportion of respondents (81%) had located freely-available research papers through individuals' own web pages. This was a surprising response considering the value many place on 'one-stop-shop' search engines and services to improve accessibility of information. However, just because the full-text was located on an individuals' web page does not mean that the reference was not sourced via a search engine or subject-based index. Indeed, four respondents indicated that they were not aware of the final location of their papers, only that they found them using free web-based search engines such as Yahoo and Google. Assuming that many respondents did go directly to known individuals' web pages to source papers, this could be seen as the electronic equivalent of reprint distribution amongst colleagues. As such, it should provide reassurance to publishers who are considering allowing authors the right to self-archive. It demonstrates the importance of subject communities and inside knowledge about the work of individuals, research groups and projects.

Academics' comments to this question revealed a misunderstanding of term 'freely available' that may have influenced the overall results. Thirty-five per cent of respondents that said they had located freely available

research papers via e-journals actually cited subscription-based e-journals, rather than free ones. The misunderstanding stems from the fact that e-journal services are often free at the point of use as a result of library or institutional subscriptions. Academics therefore believe they are 'freely available' to all. There have been many debates in the open-access community around the ambiguous use of the terms "free" and "open". Just because something is free at the point of use does not mean it costs nothing to purchase or provide.

How academics expect to use research papers

Permissions

Almost every academic (99%) expected to display and print open-access works either freely or under limits or conditions. Those that did not expect to display a work may either have been non-users, or those that print a work without displaying it first. It was not surprising to find that most academics expected to print. This triangulates with many e-journal usage studies (Woodward et al, 1997; Schauder, 1994). Indeed, printing was slightly more popular than saving, perhaps illustrating that hard copy is still the medium of choice for academics. However, over 90% did expect to save either freely or with limits. Displaying, saving and printing are the three activities that e-journal statistics often measure, by virtue of their being "full content units examined, downloaded or otherwise supplied to the user" (ICOLC, 2001).

Most academics expected to excerpt from open-access works (96% freely or with limits and conditions). This is not surprising given the practice of citation in scholarly research literature. Excerpting from works in this way for "criticism or review" falls within the fair dealing exceptions of UK copyright law.

Just under 90% of respondents expected to give works away (e.g., forward to colleagues), although only 58% expected to do so 'freely'. In contrast, 95% of academics-as-authors expected to give works away, however, a far higher proportion - 79% - expected to do so freely. Of course, the open-access research literature is often referred to as the 'giveaway literature' (Harnad, 2001). It is given away to publishers, (and thereby colleagues and the general public) for peer recognition and as a contribution to the advancement of knowledge for the public good. It is perhaps because academics-as-authors give their works away freely, that they expect to do the same with others' works as users. It is also undoubtedly an example of the prevalence of information sharing amongst communities of science, as Sir John Sulston said after winning the Nobel Prize for medicine in 2002, "Research is hastened when people share results freely" (Meeks, 2002). Projects developing Virtual Research Environments all build in the ability for research communities to share papers of interest (Virtual Research Environment Project, 2003; Torii, 2003).

Interestingly, far more academics expected to give papers away rather than lend them out. This may come back to the strong desire to own a copy, explaining why most academics wish to print or save. Tenopir and King (2000) found that,

Most readings of articles (about 60 percent) take place within six months of publication. Nevertheless, many of those articles are re-read later for scientific research or teaching purposes. For example, 37 percent of the scientists who read articles more than two years old were not reading them for the first time...Most quick reading involves recently published articles, while most in-depth reading is of older articles.

Thus whilst borrowing a copy may be satisfactory for an initial reading, it is common academic practice to re-read material, hence the need to own a copy to refer back to.

A large number of academics expected to aggregate a work. The example given for aggregate was "may be compiled into an anthology". Of course, perhaps the most common type of anthology to be compiled in an academic setting is the coursepack: a compilation of readings for use by students. Seventy-three per cent of respondents expected to compile open-access research papers into an anthology, although only 17% expected to do so freely. Annotation also appears to be a common practice, but again, while 62% expected to do this, only 16% expected to do so freely. It may have been that academics did not expect to perform

these activities very often. However, it could also have been that they saw the need for moderation or discretion in performing activities that *could* affect the moral rights of the author.

Academic users clearly saw a difference between annotation and modification, as the majority group (just over 50%) did not expect to modify open-access works at all. What was perhaps more surprising was that just under 50% did expect to modify the work. However, this may have something to do with the example given for modify, namely, “may be translated or a derivative work created”. In one sense all academic research builds on what has gone before, it is the extent to which it does so that makes the difference between a subsequent work being a ‘derivative work’ or not.

The fact that only two per cent of respondents expected to be able to freely sell open-access works, and six per cent to lease them, sends a clear signal that commercial gain is not a consideration for academics in their usage of such research papers. What is perhaps more surprising is that anyone expected to do so at all. This may indicate a lack of understanding about the nature of open-access works, or of the question itself.

Restrictions and conditions

It is perhaps good news for authors of open-access research papers that 82% of respondents recognised that there were likely to be restrictions on the use of such works. It may also be good news that 64% expected that all copies should be exact replicas of the original – something highly prized by authors and enshrined in UK moral rights legislation (Great Britain, 1988). The fact that the second largest group of respondents (56%) thought that open-access works should only be used for certain purposes may stem from a knowledge of copyright law which defines the legitimate purposes for copying very narrowly. Similarly, copyright law may also have influenced those respondents that expected usage of open-access works to be restriction to personal use, or for use by certain groups. Experiences of e-journal licences and technical protection measures may have caused 40% of respondents to state that they expected to maintain watermarks. This response may also have resulted from an inability, or lack of motivation, to remove them. Respondents’ knowledge of the web may have influenced their expectation that web-based materials should not be limited by time (time period or number of uses) or space (geographically).

Fewer respondents expected there to be conditions of use on open-access research papers, although 76% expected to attribute the author. Again, this is a condition imposed by UK copyright law’s moral rights provisions (Great Britain, 1988). Those selecting the other conditions such as agreeing to terms and conditions, registering, and expecting their usage to be tracked, may have been influenced by their experience of either fee-based or free electronic materials. It is not uncommon for users of even freely available web sites to be asked to register before using. Certainly, many subscription-based full-text services expect users to ‘click-thru’ a licence specifying certain terms and conditions before they are allowed access to the content. Although only seven per cent expected to have to pay a fee before using an open-access work, this again indicated that some respondents were not aware of the meaning of the term “open-access”.

Comparison of protection and usage required by academics

The permission, restriction and condition (P, R & C) data produced by academics-as-users provided a very interesting insight into how academics expected to use research papers. Clearly the RoMEO survey only shows how academics *think* they use freely available electronic research papers, and not their actual usage. A more accurate study in this area might take the form of a tracer study which traces academics’ actual usage of specific papers over time. However, this would have to be a very lengthy study and may be methodologically difficult.

It should also be stressed that the data on the activities allowed by authors, and those performed by users was collected from the same group and at the same time. Respondents may have been aware that the two sets of data would be compared. Thus, having just given serious thought to the permissions and restrictions they would allow over their own works, may have made them less liberal in their expectations of others’ works. Indeed, in most cases academics-as-users expected to make less use of others works than academics-as-authors were happy to allow.

Permissions

With regards to permissions, in only three cases did academics-as-users expect to be slightly more liberal with others' works than academics-as-authors were happy with. Two of these were insignificant differences. However, in the other nine cases, authors were more liberal than users demanded. It appears that whilst academics are happy to permit others to do certain activities, it is not within their normal practice to do those things themselves. Some difference in responses is not surprising as academics may well have had different 'end-users' in mind when answering the question about the use of their own paper (e.g., libraries or publishers) whereas the question about academics' own use of papers obviously only referred to usage by academics! Thus whilst respondents may be happy to allow librarians to lease or lend printed copies of their open-access works in certain circumstances, they were aware that they were unlikely to want to do this themselves. Indeed, three-quarters of the permissions responses by academics-as-users and academics-as-authors demonstrated significant difference.

Restrictions and conditions

There were far fewer differences in the 'rankings' of restrictions and conditions amongst users and authors, compared with the permissions ranking. Only one restriction and one condition specified by academics-as-users fell into a different position in the list generated by academics-as-authors' responses. There were also far fewer significant differences on restrictions and conditions between the two groups. Only three restrictions showed a significant difference (and then only at the 5% level) and in each case more academics-as-users expected to adhere to the restriction than academics-as-authors expected to impose it. Just two of the six conditions demonstrated a significant difference between the two groups: author attribution and usage tracking. However, in both of these cases more authors wished to impose the condition than users expected to adhere to it. Surprisingly, just 76% of users expected to attribute the author although this was naturally important to 81% of authors. Only one-quarter of users expected their usage to be tracked whereas one-third of authors desired usage-tracking. This illustrates the tension between information and privacy: just because usage data *can* be collected does not mean that it should (Sturges, 2002; Gutwirth, 2002).

Effect on RoMEO Rights Metadata

RoMEO Studies 2 used a somewhat arbitrary cut-off point to decide which of the P, R and C's specified by academics-as-authors should be encapsulated in the rights metadata solution. Where over 60% of respondents agreed on a P, R or C, the response was made a mandatory element. Where over 50% and under 60% of respondents agreed, the response became an optional element. Less than 50% agreement was taken as an indication of a lack of perceived relevance of the P, R or C to the protection of academic research papers. Using the same cut-off points, we performed the same analysis on the data specified by academics-as-users. The results from both activities are illustrated in Table 12 below.

Protection required by academics-as-authors	Usage required by academics-as-users
Permissions	
Display	Display
Give	Give (optional)
Print	Print
Excerpt	Excerpt
Save	Save
Aggregate (optional)	Aggregate (optional)
Sell (prohibit)	Sell (prohibit)
Restrictions	
Exact replicas	Exact replicas
For certain purposes (optional)	For certain purposes (optional)
Conditions	
Attribution	Attribution

Table 12 A comparison of the rights metadata solution proposed by RoMEO and the usage needs of academics

Although there were significant differences overall between the options chosen for each permission by the two groups, the permissions and options on which over 50% agreed were almost identical. The only difference was that users were less likely to want to 'give' freely available research papers away. This indicates again that it is seen as the authors' prerogative to distribute their paper amongst colleagues, and is not something that tends to be done by others. The restrictions and conditions on which over 50% of users agreed were identical to those on which over 50% of authors agreed.

This is an important finding both for the RoMEO rights metadata solution and for open-access generally. It seems that academics can put their works on open-access without fearing that other academics will take liberties with their material. Also, those adopting the RoMEO rights metadata solution to protect their works can be reassured that it seems to meet the needs of the majority of academic authors. However, it is an obvious but important point that academics are not the only end-users of research papers. Whilst this study only looked at academics' use of open-access research papers, students are another large user group, as are researchers from commercial organisations, and perhaps even aggregators and other secondary publishers. Further research into the expected usage of open-access research papers by these groups would be of significant interest and benefit.

Bibliography

- Brown, C. M. (1999). "Information seeking behavior of scientists in the electronic information age: astronomers, chemists, mathematicians, and physicists." Journal of the American Society for Information Science. **50**(10): 929-43.
- Carr, L., Steve Hitchcock, Wendy Hall and Stevan Harnad (2000). *A usage-based analysis of CoRR: a commentary on "CoRR: a Computing Research Repository" by Joseph Y. Halpern*. [Self-archived] URL: <http://cogprints.ecs.soton.ac.uk/archive/00001645/00/harnad00.halpern.htm>.
- Cowhig, J. (2001). "Electronic article and journal usage statistics (EAJUS): proposal for an industry-wide standard." Learned Publishing **14**(3): 233-236.
- Eason, K., Sue Richardson, and Liangzhi Yu (2000). "Patterns of use of electronic journals." *Journal of Documentation* **56**(5): 477-504.
- Eason, K., and Susan Harker (2000). "Psychological processes in the use of electronic journals." *Serials* **13**(2): 67-72.

- Gadd, E., Charles Oppenheim and Steve Proberts (2003a). "RoMEO Studies 1: The impact of copyright ownership on academic author self-archiving." Journal of Documentation **59**(3): 243-277.
- Gadd, E., Charles Oppenheim and Steve Proberts (2003b). "RoMEO Studies 2: How academics want to protect their open-access research papers." Journal of Information Science **29**(5): [In Press].
- Great Britain, 1988. Copyright, designs and patents Act 1988, Elizabeth II. London : HMSO
- Gutwirth, S. (2002). *Privacy and the information age.* Lanham, Md; Oxford, Rowman & Littlefield.
- Harnad, S. (2001). "For Whom the Gate Tolls? How and Why to Free the Refereed Research Literature Online Through Author/Institution Self-Archiving, Now." <http://cogprints.soton.ac.uk/documents/disk0/00/00/16/39/index.html>.
- Hunt, C. (2002). *Archive User Survey.* Electronics and Computer Science Department. Southampton, Southampton University. URL: <http://www.eprints.org/results/>.
- ICOLC (2001). Guidelines for statistical measures of usage of web-based indexed, abstracted and full-text resources. URL: <http://www.library.yale.edu/consortia/2001webstats.htm>
- Kidd, T. (2002). "Electronic journal usage statistics in practice." Serials **15**(1): 11-17.
- Lessig, L. (2002). "Free Culture. [Keynote paper presented at the Open Source Convention. 22-26 July, 2002. San Diego, CA]." <http://www.oreillynet.com/pub/a/policy/2002/08/15/lessig.html>.
- Luther, J. (2000). White Paper on Electronic Journal Usage Statistics. Washington, DC., Council on Library and Information Resources. URL: <http://www.clir.org/pubs/reports/pub94/contents.html>.
- Meek, J. (2002). Nobel prize for British and US scientists who used worms to decode the book of life. The Guardian. London. URL: <http://www.guardian.co.uk/Print/0,3858,4517292,00.html>.
- Nicholas, D., and P. Huntington. (2002). *Big deals: results and analysis from a pilot analysis of web log data.* London: Ciber.
- Open Digital Rights Language. (2002) URL: <http://odrl.net/>
- Olsen, J. (1994). *Electronic journal literature: implications for scholars.* London, Mecklermedia.
- Pinfield, S. (2001). "How Do Physicists Use an E-Print Archive? Implications for Institutional E-Print Services." D-Lib Magazine **7**(12). URL: <http://www.dlib.org/dlib/december01/pinfield/12pinfield.html>.
- Project COUNTER (2003). URL: <http://www.projectcounter.org/>
- Richardson, R. J. (1981). "End-user online searching in a high-technology environment." Online **5**(4): 44-57.
- Schauder, D. (1994). "Electronic publishing of professional articles: attitudes of academics and implications for the scholarly communications industry." Journal of the American Society for Information Science. **45**(2): 73-100.
- Shackel, B. (1991). *BLEND-9: overview and appraisal.* British Library Research Paper 82. London, British Library Research and Development Department.
- Simpson, A. (1988). "Academic journal usage." British Journal of Academic Librarianship **3**(1): 25-36.

Sturges, P. (2002). Privacy in the digital library environment. Boston Spa, Resource.

SuperJournal Project. (1999) URL: <http://www.mimas.ac.uk/sj/>

Swan, A. (2002). Authors and electronic publishing: the ALPSP research study on authors' and readers' views of electronic research communication. West Sussex, ALPSP: 83pp.

Tenopir, C., and Donald W. King (2000). *Towards electronic journals: realities for scientists, librarians and publishers*. Washington, Special Libraries Association.

Tenopir, C., and Donald W. King (2001). Electronic journals: how user behaviour is changing. Online Information, London, Online Information. URL: http://www.online-information.co.uk/proceedings/online/2001/tenopir_c.asp.

Tenopir, C., and Donald W. King (2002). "Reading behaviour and electronic journals." Learned Publishing 15(4): 259-265.

Torii: (2003) Access to the digital research community. URL: <http://tips.sissa.it/docs/booklet.pdf>

Tuck, B., et al. (1990). *Project Quartet. Library and Information Research Report 76*. London, The British Library.

Virtual Research Environment Project. 2003. URL: http://www.external.ucl.ac.uk/elpt/VRE_Project/index.asp

Woodward, H., McKnight, C; Pritchett, C; Rowland, F (1997). Cafe Jus: commercial and free electronic journals user study. London, British Library. Research and Innovation Report: 1-52.

Yu, L., and Ann Apps (2000). "Studying E-Journal User Behavior Using Log Files: The Experience of SuperJournal." *Library & Information Science Research* 22(3): 311-338.