

This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

Developing IPR solutions for academic author self-archiving

Charles Oppenheim*, Elizabeth Gadd, and Steve Proberts Department of Information Science, Loughborough University, UK

*c.oppenheim@lboro.ac.uk

Abstract

This paper describes the work of the UK JISC-funded RoMEO (Rights Metadata for open archiving) project. It reports on a survey of 542 academic authors and an analysis of 80 journal publishers' copyright transfer agreements, and how they have informed the development of some simple rights metadata by which academics can protect their research papers in an open access environment. It also reports on a survey of 22 OAI Data Providers and 13 OAI Service Providers, and how the results have informed the development of a "metadata protection solution" that describes the conditions of use of freely available metadata.

Keywords: Intellectual Property Rights; Self-archiving; Metadata

1 Introduction

Universities all over Europe are looking for a solution to the so-called 'serials crisis'. Universities are producing journal articles and assigning copyright in them to publishers for publication in their journals. The same Universities are buying those journals for their libraries, often at high prices. Yearly journal cancellations are the norm for most University Libraries. Not surprisingly, there are many calls for change to this scholarly communication process. One much-publicised solution is that academic authors 'self-archive' their research papers on freely available web servers [11, 4]. A protocol that has made this solution viable is the Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH) [15]. The protocol allows Data Providers (DPs) to disclose metadata about research papers (or other resources) which Service Providers (SPs) can then harvest and provide access to.

Thus the OAI-PMH provides a technical framework by which scholarly communication could be revolutionised through the self-archiving (or institutional-archiving) of so-called "e-prints" and other materials. However, as has been documented many times, and as has always been the case with digital libraries, the barriers to the success of this revolution are not technical but cultural [22]. One of the key cultural barriers to the use of the OAI-PMH for self-archiving is the issue of intellectual property rights [2]. For authors to self-archive, they need to have retained the right to do so [12]. For authors to retain the right to self-archive they need to ensure they do not assign that right to publishers [5]. For authors to feel

confident about self-archiving, they need to be reassured that their intellectual property will be protected in some way in the OAI environment. For Data Providers to make metadata about their documents freely available, they, too, need to be reassured that their metadata won't be misused (e.g. harvested and then sold). For Service Providers to harvest and enhance that metadata they need to be sure that they are allowed to do so. And for users to make legitimate use of self-archived works, they need to be informed as to how they may be used.

There are a wide range of commercial digital rights management systems available [13]. However, they are aimed at enabling e-commerce, not protecting the so-called "give-away" literature. The RoMEO (Rights metadata for open-archiving) Project has been funded by the UK Joint Information Systems Committee (JISC) to investigate the rights issues affecting the self-archiving of academic research papers under the OAI-PMH, and to suggest solutions to some of the problems [21]. In particular, it has been tasked with developing a set of machine-readable rights metadata elements to describe the rights status of documents disclosed under the OAI-PMH, and also to create a means of describing the rights status of the metadata itself. This paper describes the processes informing the development of the RoMEO rights metadata and the metadata protection solution.

2. Methodology

To inform the design of the rights metadata and metadata protection solutions, Project RoMEO undertook online surveys of the key stakeholders in the self-archiving process: publishers, academic authors, and OAI Data Providers and Service Providers. It also performed an analysis of 80 journal publishers' copyright transfer agreements (CTAs).

2.1 Academic author survey

One of the principal aims of the academic author survey [7, 8] was to understand how academics wished to protect their open-access research papers. To provide a framework for collecting such data, we used the Open Digital Rights Language [19] (ODRL). This provides a model of permissions, restrictions and conditions 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 requirement that must be fulfilled before the permission may be performed (e.g., you may print four times, but only if you pay a fee). Only those ODRL terms that were considered to be of any possible relevance to self-archived e-prints were incorporated into the survey. The survey was advertised via the Emerald Literati Club email list to about 16,000 authors as well as to other discussion lists.

2.2 Publisher CTA Analysis

Two approaches were taken to select journal publishers for the CTA analysis: a targeted approach; and a self-selecting approach. A target list of 84 publishers was drawn from publishers of high-impact journals and those publishing large numbers of refereed academic titles. A general call for agreements was also sent out via two professional bodies for academic journal publishers: the Association of Learned and Professional Society Publishers (ALPSP) [1] in the UK and the Society for Scholarly Publishing [23] in the USA. Agreements were collected between August and December 2002. An initial examination of a small number of agreements was undertaken and a list of criteria against which all agreements would be analysed was developed.

2.3 OAI Data and Service Provider Survey

The Data and Service Provider surveys used a mixture of open and closed questions to ascertain their views on the intellectual property rights (IPR) protection afforded to metadata in an open access environment.

3 Academic author surveys

3.1 Response rate

Responses were received from 542 academics from 57 countries. The majority (one-third) came from the UK. Fifty per cent of respondents came from a pure science discipline, with 38% from 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 mode was ten papers or less.

3.2 Permissions

Respondents were asked, "What

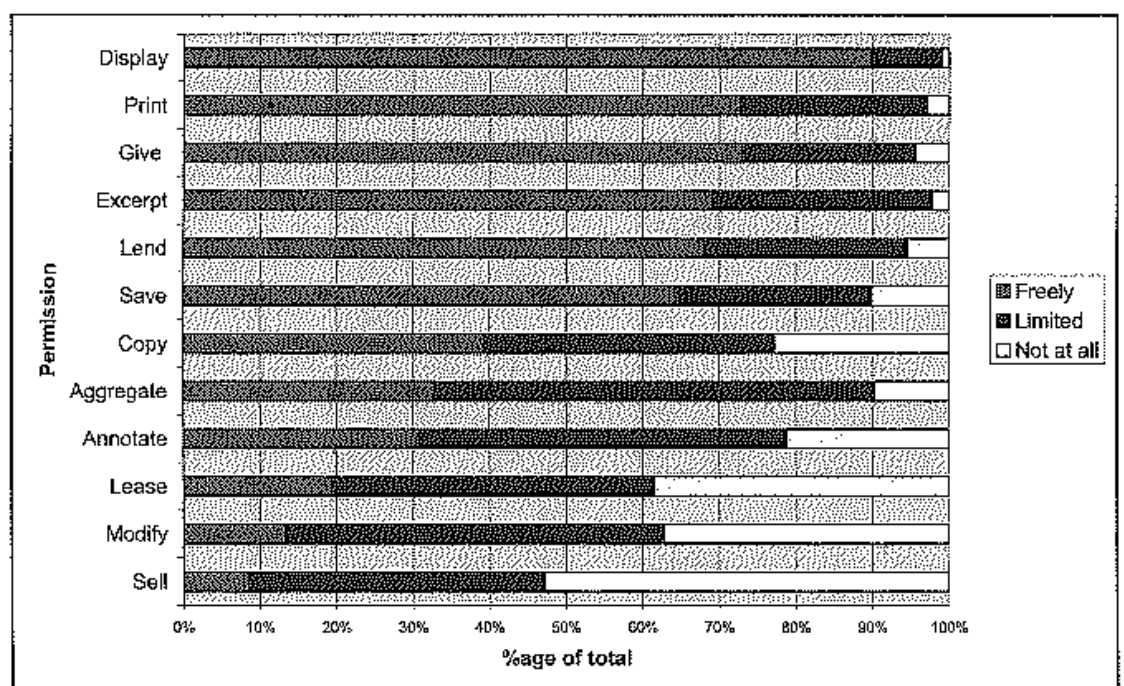


Figure 1 Activities academics would allow with their own research papers

would you consider to be acceptable use of any articles you made freely available on the web?” and were given a list of twelve ODRL permissions. They were requested to indicate for each permission whether they would be happy for others to perform that activity with their work “freely”, “with limits”, or “not at all”.

Table 1 indicates in bold the most common preference for each permission.

	Freely	Limited	Not at all	TOTAL
Aggregate	166	292	49	507
Annotate	157	245	108	510
Copy	202	195	117	514
Display	471	48	5	524
Excerpt	355	148	11	514
Give	375	116	23	514
Lease	97	211	192	500
Lend	347	135	28	510
Modify	68	252	189	509
Print	378	127	15	520
Save	334	130	52	516
Sell	43	196	268	507

Table 1 Permissions data table

Table 2 lists the permissions in order of greatest agreement amongst respondents.

Permission	Percentage agreeing	Option	Rank
Display	90%	Freely	1
Give	73%	Freely	2=
Print	73%	Freely	2=
Excerpt	69%	Freely	4
Lend	68%	Freely	5
Save	65%	Freely	6
Aggregate	58%	Limited	7
Sell	53%	Not	8
Modify	50%	Limited	9
Annotate	48%	Limited	10
Lease	42%	Limited	11
Copy	39%	Freely	12

Table 2 Permissions in order of agreement

The majority vote for seven of the 12 options was to allow that activity freely. The permission on which most respondents agreed was display. 90% of respondents felt

this should be allowed freely. Print and give were the next most commonly agreed upon terms. Seventy-three per cent of respondents held that printing and giving should be allowed freely. This is not surprising. E-journal user surveys have shown that printing is consistently the most common action with an e-journal article [24, 26]. The example provided for 'give' was that "copies may be forwarded to colleagues". Again, this is a common activity amongst academics and may be the reason why most academics make their works available on their personal web pages.

The majority of respondents were happy to allow others to excerpt from, and lend their works freely. The number of academics that were unhappy with the idea of others' saving their works (e.g., saved to disk) was perhaps surprising. Ten per cent said that this should not happen at all. One-quarter felt that saving should be limited or conditional. It may have been that these respondents share publishers' fears of the potential for infringement offered by an electronic copy as opposed to a printed copy. Although 73% of respondents wished to allow printing freely that left just over one-quarter that felt it should be limited or conditional.

Aggregate was the first activity that the majority of respondents felt should be limited or conditional. Fifty-eight per cent agreed on this. One-third was happy for others to freely aggregate their works. The example given for aggregate was "may be compiled into an anthology" which may include commonplace course pack compilations or re-publishing. The uncertainty may therefore have stemmed from a perceived lack of control over the nature of the aggregation, in particular, what other works the compilation would contain.

Not surprisingly, the majority of respondents (53%) were not happy with the idea of others selling their works without permission. Less expected was the eight per cent that felt this could be done freely. Similarly, leasing was not a popular activity amongst respondents. Forty-two per cent felt that this should be a limited or conditional permission. Thirty-eight per cent said they would not be happy with it at all. This is perhaps not surprising, as leasing offers commercial gain to a third-party.

Modify and annotate met with a similar response from academics. Fifty per cent and 48% respectively thought such activities should only be done under certain limits or conditions. Annotation in the print environment (scribbles in the margin of a library text book) may cause little concern to academics. However, they may fear loss of control over an electronically annotated work that can then be quickly and easily disseminated.

An interesting result was the response to the term "copy". Although 39% said they would allow this freely, 38% per cent felt it should be limited or conditional. The example given for copy was "mount another copy on the web". This was to distinguish it from a printed copy, described by "print" or a personal electronic copy, described by "saves". The distinctions may have caused problems of interpretation. Those allowing it freely may have viewed copying in the same terms as print or save,

both of which the majority allowed freely. Those that felt it should be limited may either have seen copying as a form of re-publication (mounting my article on their web site without my permission) or they may have equated copying with plagiarism. “Copy” is a word with many meanings. If I copy from another in a test, I have cheated. If I copy an article from the library, I have pursued an educational goal.

3.4 Restrictions and conditions

Respondents were then asked, “What (if any) restrictions or conditions would you want to place on the use of your self- archived works?” They were given the list of ODRL restrictions and conditions and were asked to select from it. In total, 528 respondents answered this question.

3.5 Restrictions

The list contained ten restrictions (including ‘no restrictions’). The results are given in Figure 2.

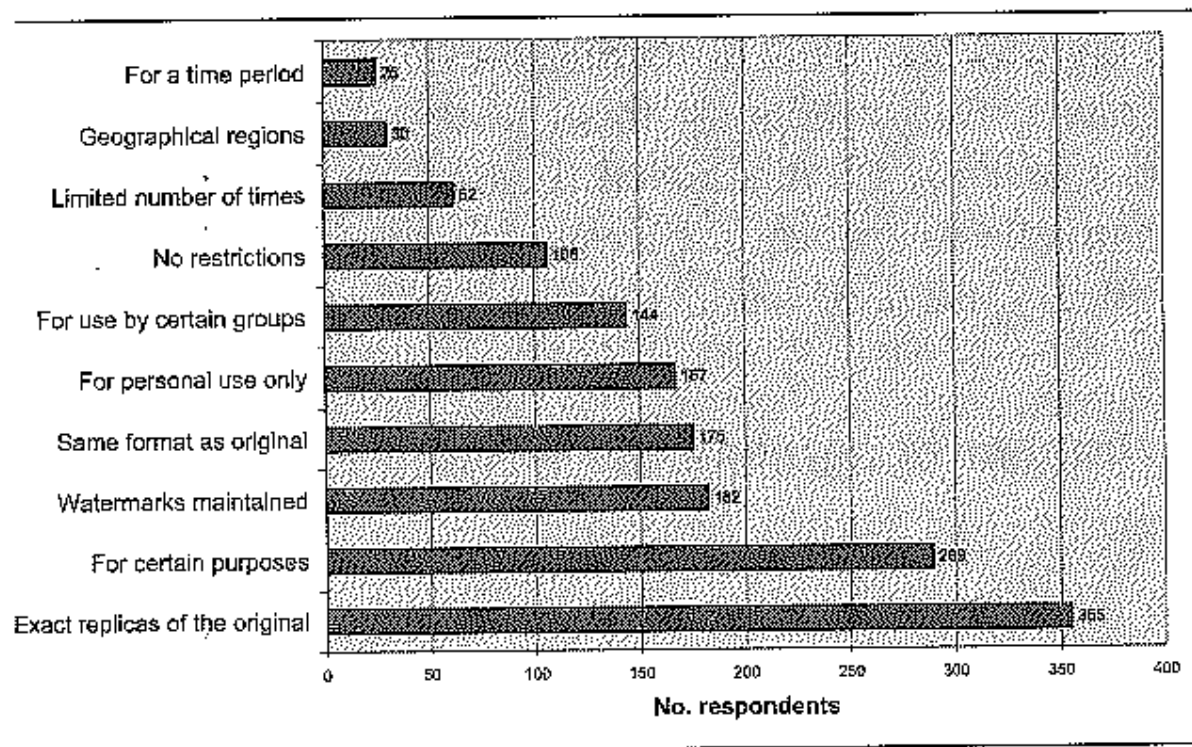


Figure 2: Restrictions academics wanted to place over their own research papers

Just over 20% (106) of all respondents selected the “no restrictions” option. However, 37 (35%) of these then did also select a restriction. This may mean that some respondents did not view the restriction options on the list as ‘restrictive’. The total number of respondents selecting only the “no restrictions” option was 69, or 13% of the total. The remaining 88% of respondents required some restriction over the use of their works. This indicates that simply releasing

works into the 'public domain' for use by anyone for any purpose is not an option for academics.

The largest group of respondents (67%) wanted reproductions of their work to be exact replicas of the original ("the text must not be altered in any way"). This is in line with authors' moral rights. It could have been that some respondents interpreted this question to mean that they wanted to limit reproduction of their work to the entire work (i.e., not parts). However, the fact that just under 73% of respondents indicated in the previous question that users may excerpt from their works freely implies that the interpretation of this question was as intended.

Fifty-five per cent of respondents wanted the use of their work to be limited to certain purposes (e.g., teaching, research or non-commercial). This was a considerable number and one of only two restrictions on which more than 50% of the respondents agreed.

About one-third (34%) of respondents felt that existing watermarks or security features should be maintained. Slightly fewer (33%) wanted reproductions to adopt the same format as the original (e.g., PDF). These results may further echo respondents' strength of feeling about copies being exact replicas of the original.

Thirty-one per cent believed their works should only be used for personal purposes. Twenty-seven per cent felt that usage should be limited to certain specified groups (e.g., only for use by educationalists). Much smaller numbers (12%, 6% and 5%) expressed an interest in restricting usage of their works to a limited number of times, to certain geographic regions, or for limited time periods respectively.

3.6 Conditions

The list contained six conditions (including "no conditions") and respondents were asked to select those they would like to apply to the usage of their open-access research papers. The responses are given in Figure 3 below.

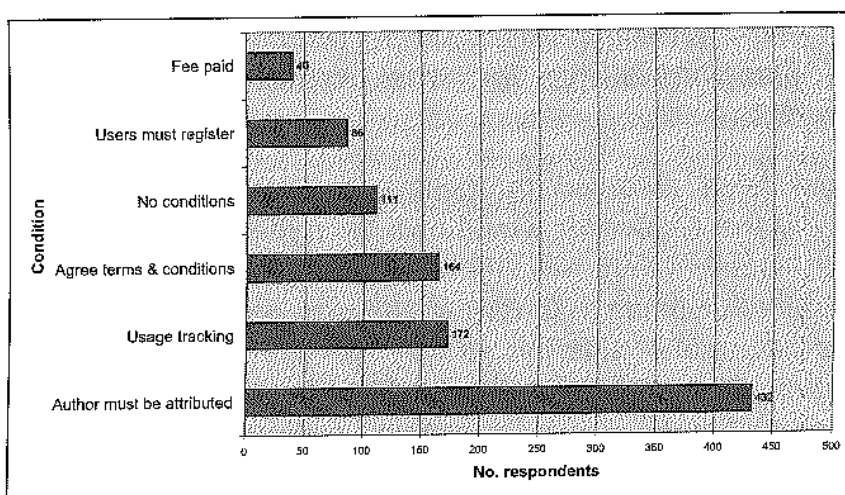


Figure 3 Conditions academics wanted to place on the use of their own research papers

A similar number of

respondents stated they required 'no conditions' (111) on the use of their works as those requiring 'no restrictions' (106). However, as with the 'no restrictions' option, a large number of respondents (in this case 50% (56)) did actually select a condition from the list. Of this 56, 52 (93%) selected the 'attribution' condition.

Overall, 432 respondents (81%) specified that they expected to be attributed as the author of their research papers. No other condition or restriction received such consensus. Just over 30% felt that all usage of their materials should be tracked. A similar number asked that users agreed to certain terms and conditions (as in a licence agreement) before being allowed to use the work. It will be important that any rights metadata associated with a paper is enforceable and agreeing to terms and conditions may be the most appropriate way of making it so. Sixteen per cent felt that users should at least register before using the work, and a much smaller figure, eight per cent, thought that users should pay a fee.

3.7 Implications for rights metadata

The results showed that whilst there was some agreement on a number of permissions, restrictions and conditions (P, R and C's) required by academics, the agreement in many cases was under 50%. This indicates that academics have a wide range of views on the protection their self-archived works require, and so a certain amount of flexibility in the metadata would be appropriate. At the same time, the Project Advisory Board advised the team to 'lead' rather than 'follow' in the design of a rights metadata solution. A middle-ground balancing simplicity and flexibility was clearly needed. It was decided that any P, R and C's receiving over 50% agreement would be selected for the metadata solution, and any receiving over 60% agreement would be made a mandatory element.

We therefore made display, give, print, excerpt, lend and save mandatory permissions, aggregate an optional permission, and prohibited sell. In terms of restrictions there would be one mandatory one: exact replicas, and for certain (non-commercial) purposes would be optional. Just one condition, attribution, would apply to all permissions.

3.8 Responsibility for the creation of rights

It had been assumed throughout the survey design and analysis that academics would own the rights in their research papers and that they would therefore be generating the rights metadata pertaining to those papers. However, there are three other scenarios. Firstly, it may be that academics have assigned their copyright to publishers, but that the publisher allows self-archiving under certain conditions. In this case, the publishers' conditions would need to form a part of the rights metadata. Secondly, it may be that the author's employer owns copyright. Traditionally, universities have not asserted copyright in academics' work. However, a number of groups are lobbying universities to stake their claim over the copyright in research

outputs. In such a scenario, the employer would have a say in the creation of the rights metadata.

Finally, it may be that institutions establishing e-print archives decide that they will only accept documents that can be used in certain ways (i.e., users must be allowed to print, excerpt and save). In this scenario, the archive may restrict the rights metadata options open to the author – or refuse to host papers that don't meet the required standards. This would certainly make the implementation and usage of rights metadata far simpler for Data Providers and Service Providers.

4 Publisher CTA Analyses

It was hoped that the CTA analysis would provide data to answer the first scenario above. In total, 80 agreements were received from publishers. Forty-eight agreements were collected from the target group of 84 publishers. This was a response rate of 57.1%. Between them, the target group of respondents published 6,960 academic journal titles [25]. A further 32 agreements were collected from other sources (e.g. the ALPSP mailing), representing a further 342 titles.

4.1 Copyright assignment

Morris has written, “it is...hard to find a justification, other than convenience, for insisting on taking the author's copyright [14].” However, 72 of the 80 agreements (90%), representing 94% of the journal titles, and asked authors for copyright assignment, although four of these (2.4% of titles) gave authors the option of signing an exclusive licence agreement instead. Five of the remaining eight asked for an exclusive licence (4.9% of journal titles) and three asked only for a non-exclusive licence (0.9% of titles).

This implied that most academics have to assign copyright to publishers, thus rendering them unable to self-archive their work unless the CTA allowed them to. It would also render them unable to specify the terms and conditions under which others may use their works via rights metadata.

4.2 Number of agreements offering exceptions

The agreements were therefore studied to see what provisions were made for authors to use their own works once copyright had been assigned. For the three non-exclusive licences, this part of the analysis was not applicable. Of the remaining 77 agreements, 22 (28.5%) representing 651 journal titles (8.9%) made no exceptions at all for authors to use their own works. Such authors would only legally be able to re-use their works in accordance with exceptions to copyright in the national law. Three agreements representing 122 titles (1.6%) were at the other end of the spectrum and allowed authors unlimited copying of their own works. The remaining 52 agreements allowed limited copying.

4.3 Types of exception

Eleven different types of exception were identified. Some agreements stated that they would allow authors to use their work in a certain way, but only if the publisher's permission was sought. Such allowances were not considered true exceptions and were omitted from the analysis. The number of agreements listing the various exceptions is illustrated in Figure 4.

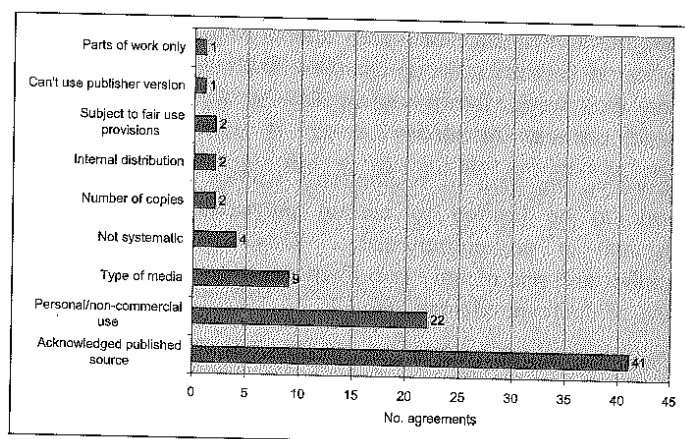


Figure 4 Exceptions to copyright assignment

Of the 52 agreements allowing limited exceptions, 36 allowed authors to reuse their article as a chapter of a book (69.2%). Slightly fewer (34) specified that such a publication must be a collection of the author's own writings. Fewer again (27), specified that the work could be used in a dissertation or thesis. Teaching and general 'reproduction' were allowed by 29 and 28 agreements respectively, and 'distribution to colleagues' and uses for 'scholarly or academic purposes' were mentioned by far fewer (15 and 13 respectively).

4.4 Self-archiving exception

Self-archiving (making the work available on a freely-accessible web server, not just on an intranet) was permitted by 28 agreements. Including the three non-exclusive licences and the three unlimited exceptions, a total 42.5% of the agreements overall allowed self-archiving. These agreements represented approximately 3,590 journal titles – 49.1% of the total represented by this survey. A further four agreements were unclear on this point and may have allowed self-archiving if clarification was sought. They used ambiguous phrases such as "authors need not seek permission to use their own materials in other publications appearing under their own name".

4.5 Conditions over self-archiving

The 31 agreements allowing self-archiving did so under the following conditions.

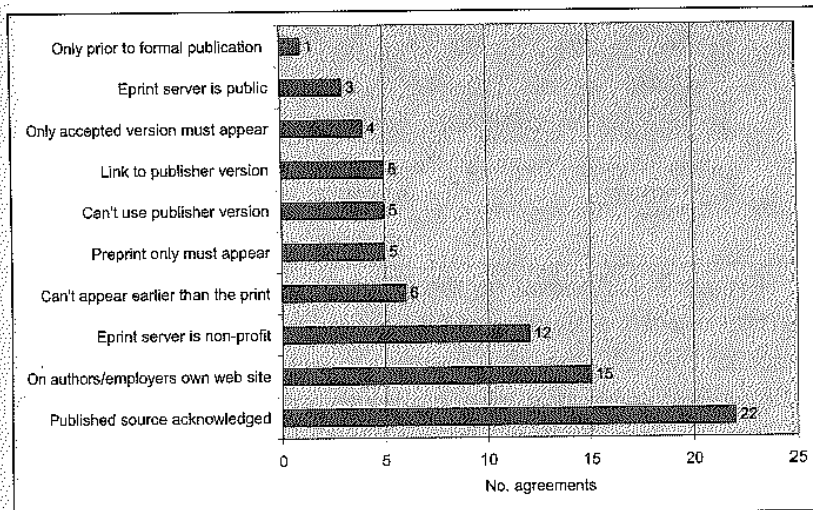


Figure 5: Conditions placed over self-archiving

Acknowledging the published source was important to the largest group of publishers (22 or 70.9%) and five publishers asked for a link from the self-archived version back to the publisher version. The type of e-print server was important to a good proportion of the publishers. About half would only allow self-archiving on either the author's personal web site or on the institutional site. Twelve specified that the server had to be non-profit and three stated that it must be publicly accessible.

Timing was also important to publishers – although different views on this emerged. Six said that the e-print cannot appear earlier than the print, one said the e-print must only appear earlier than the print (i.e., it should be removed once the published version was available). Five said that only the preprint (pre-refereed version) must appear, and four said that only the post-print (the accepted version) must appear. Five specified that the publisher version (i.e., the publisher pdf) must not be used for self-archiving (although one publisher said that this was permissible).

Implications for rights metadata

Just fewer than half the journal titles represented by the survey allowed author self-archiving in some form. Thus, authors submitting to the other 51% of journals would not even be able to self-archive, let alone specify the usage of their self-archived paper through rights metadata.

Those publishers allowing self-archiving placed a wide range of conditions on the actual mounting, but made little reference to conditions governing the end-use of the paper by others – the purpose of the rights metadata. The main instruction for end-users would come via the publisher acknowledgement required by most respondents. However, such a statement would be expected to appear on the document itself, not in the metadata.

It is probably safest to conclude that even those assigning copyright who are still able to self-archive, would not be in a position to use the RoMEO rights metadata, as they have probably assigned the right to tell others how their work can be used.

5 OAI Data and Service Providers Survey

OAI Data Providers and Service Providers face a similar copyright protection dilemma to academic authors, but with respect to their metadata. There are a number of schools of thought concerning the rights status of a single metadata record, but EU law is quite clear that a collection of metadata records is certainly afforded database right, and possibly copyright protection. An individual record probably qualifies for copyright protection [10].

5.1 Response rate

Twenty-two Data Providers responded to the DP survey. In April 2003, there were 79 Data Providers registered with the OAI site [16]. Thus, the respondents to this survey represented 27.8% of those registered. Thirteen Service Providers responded to their survey. Again, in April 2003 there were only 12 Service Providers registered with the OAI, so the thirteen responses we received actually represents a very good response rate.

5.2 Ownership of rights in metadata

The owner of copyright in an individual metadata record is the creator of that record, and the owner of a database is the maker of the database. In order to ascertain the ownership of rights in metadata, DPs were asked, "How do you generate the metadata records that you disclose under the OAI?"

Sixteen respondents stated that metadata was generated either entirely by, or with the input of, the Data Provider. Only five Data Providers stated the metadata was created exclusively by their depositing authors.

SPs were asked whether they enhanced or added any value to the metadata they harvested from DPs, thereby becoming joint copyright owners. Nine said that they did (75%) and three did not. Those that did were asked to list the ways in which they enhanced the metadata. The responses are given below.

- Normalize field values (e.g. date) (4)
- Subject classification (2)
- Citation analysis (2)
- Name authority
- Update provider information
- Add domain addresses to URLs that are lacking them
- Map to local data model

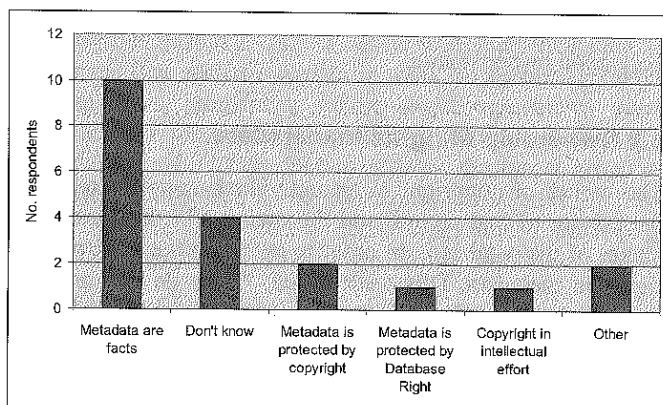
The important question is, do their enhancements merit copyright protection? Arguably, the enhancements made to a metadata record by the SP would qualify them for joint copyright ownership.

However, as one of the requirements for copyright is the demonstration of “sweat-of-brow” (in the UK) and skill and effort in other countries, it would seem logical that the enhancements would also need to fall into this category in order to qualify. Thus, enhancements such as normalising field values and adding in domain addresses to URLs that lack them may not qualify for copyright protection, but subject classification and the addition of name authority would.

5.3 Data Providers’ views on the copyright status of metadata

As the IPR status of metadata is a grey area, DPs were asked their opinions on the rights status of both individual and collections of metadata records. The first question asked, “How do you view the copyright status of an individual metadata record?” Five options were given plus an ‘Other’ category. Twenty responded to this question. The results are illustrated in Figure 6 below.

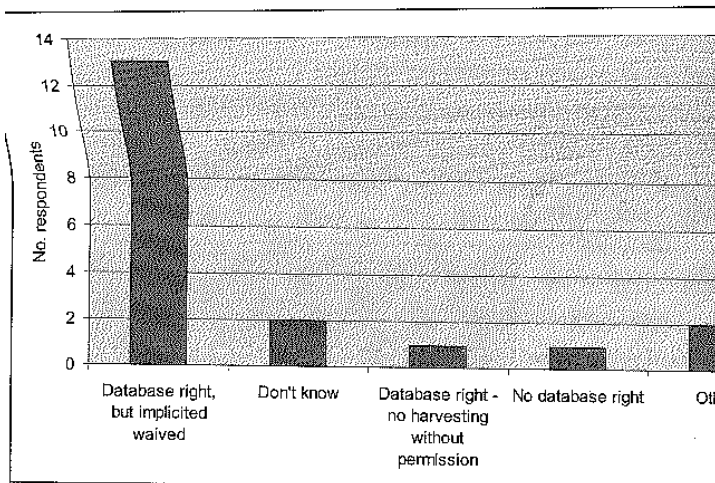
Figure 6 DP views on the IPR status of an individual metadata record



It can be seen that half of the DPs (50%) believed that metadata were facts, and that there was therefore no copyright in a metadata record. Four did not know. Two felt a record was protected by copyright although one thought the copyright rested only in those parts of the record that represented intellectual effort. One thought a record was protected by database right. Two selected the ‘Other’ category of which one wrote, “Don't care. This is a visibility and accessibility service to authors; the content is not ours.” The other admitted that they had not “thought about it much”; although no usage to date had given them cause to worry.

Respondents were then asked, “How do you view the rights status of your collection of metadata records?” Four options were given, plus an ‘Other’ category. The responses of the nineteen answering this question are illustrated in Figure 7.

Figure 7 DP views on the IPR status of a collection of metadata records



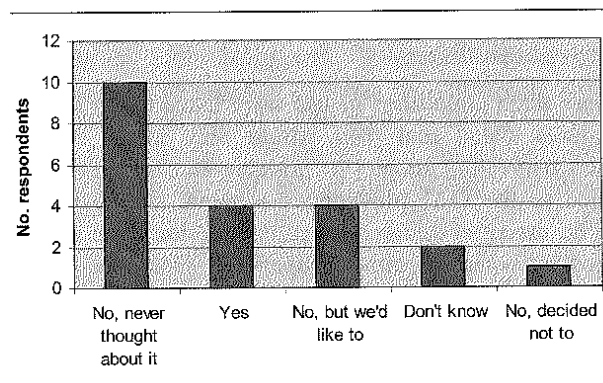
It can be seen that the majority of respondents, while believing that collections of metadata records enjoyed database right, thought that this right was implicitly waived in the OAI community. Two did not know what the situation was. One thought that database right did not apply, and another thought it did apply and that as a result data could not be harvested without permission.

5.4 Assertion of metadata rights by Data Providers

DPs were then asked whether they asserted the rights they believed they enjoyed in their individual and collections of metadata records.

The first question asked, “Do you assert the copyright status of your individual metadata records in the records themselves?” Respondents were given five options; “Yes”, “No, never thought about it”, “No, but we’d like to be able to do this”, “No, we considered it but decided against it”, and “Don’t know”. Respondents answering yes were then invited to give details of how they did this. The 21 responses to this question are illustrated in Figure 8 below.

Figure 8: Did DPs assert rights in individual records?

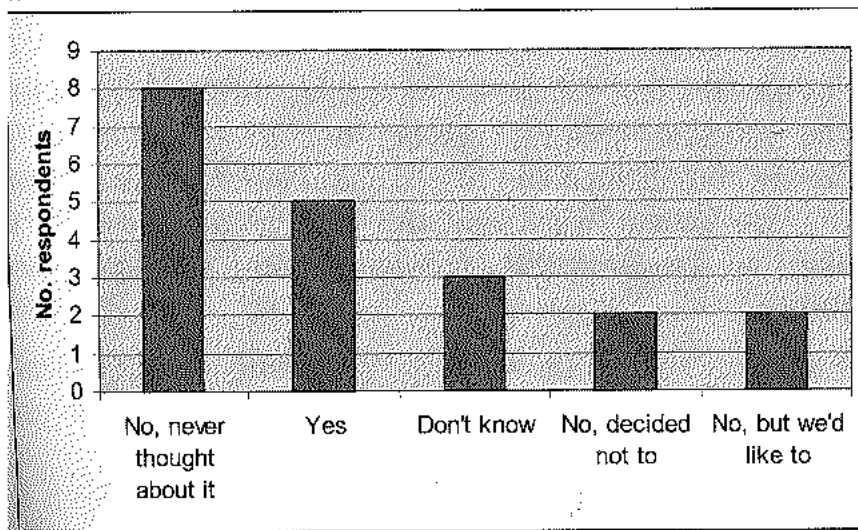


The largest group of respondents (47.6%) had never thought about this and did not assert rights in individual records.

However, four did assert rights at record level, and another four would like to. Two did not know, and one had considered this but had decided not to. Of the four answering “Yes”, three gave details. One wrote that they had “added the right container into the metadata record”, presumably referring to the optional about field that each OAI metadata record may contain. Another stated that they did not have to do anything to assert the rights in an individual record: “US copyright law says we create it, we own--without registration, fees, filing, etc.”

The next question asked DPs, “Do you assert the copyright status of your collection of metadata records?” Again, those answering, “Yes”, were asked to give details as to how they did this. Respondents were given the same five options as for the previous question. Results from the twenty respondents are given below.

Figure 9: Did Data Providers assert rights in collections of records?



Again, the largest group of respondents (40%) had never thought about this issue. Five did assert the rights in their collection and two expressed a desire to do this. Two had considered it but had decided against it, and three did not know whether they asserted rights in their collection or not. Of the five answering “Yes”, three gave details. One repeated their statement from the previous question that their rights were protected under US copyright law and were thus ‘asserted’. Another stated that they used the <metadataPolicy> element of the eprints Description container in response to the Identify verb.

5.5 Service Providers views on the copyright status of metadata

A good indication as to how SPs viewed the rights status of metadata was whether they checked the rights status before harvesting. SPs were therefore asked, “Before harvesting metadata do you check if it can be freely harvested and to what use it may be put?” They were given a choice of five options: “Yes”, “Sometimes”, “No, metadata is implicitly free under the OAI”, “No, never thought about it”, and

“Don’t know”. Respondents answering yes or sometimes were asked to explain how they did this.

No respondent selected the “Sometimes”, or “Don’t know” options. However, an equal number (4) selected one of the other three categories: that they did check; that metadata was implicitly free under the OAI; or that they had never thought about it.

Of the four that said that they did check, three stated that they communicated with the Data Provider(s) and one checked “OAI data policy statements”. One had a written agreement with data providers “to ensure that harvesting protocols are observed”, and one had email permission: “Each record is badged with a link back to the [original] server” they explained.

Two of the respondents that said they had “never thought about it” offered further information that showed that they did check for metadata policies. The first stated that they “personally visit the machine that the service is running on and look for any information provided about its intended use.” The second stated that they “usually check the site to see whether or not the subject area fits our service...In the process of doing this, we will check to see whether the metadata can be freely harvested. We don't usually check to see what use the metadata may be used for. If there are restrictions we assume that they will be included in the Rights element of each piece of metadata.”

5.6 Service Providers views on the copyright status

Having considered SP’s views on the copyright status of others’ metadata, they were then asked what they considered the copyright status of their own metadata enhancements to be. A choice of three options was given: “There’s no copyright in them”, “There probably is copyright in them, but we’re not concerned about protecting it”, and “There is copyright in them, and we are concerned about protecting it.” All respondents were asked to explain their answer.

Eight of the nine that had enhanced metadata responded, of which, seven (87.5%) said they thought there was copyright in the enhancements. However, whilst four were concerned about protecting it, three were not. The other respondent felt that there was no copyright in their metadata enhancements.

5.7 How Data and Service Providers wanted their metadata to be used

DPs were asked, “What uses do you expect others to make of your metadata?” Respondents could select one of two options: 1) “We have no objections to other organisations harvesting our metadata, as long as it is...” or 2) “We would object to other organisations using our metadata without our permission”. Those selecting the first option were given a list of six conditions (including ‘Other’) and asked to select as many as they may wish to apply to the use of their metadata. The list included;

- “freely available”,

- “used for non- commercial purposes”,
- “not altered in any way”,
- “attributed to our organisation”,
- “Pointed to the original publisher source”.

Seventeen respondents stated that they had no objections to others harvesting their metadata, however a further four respondents selected conditions they would like to apply to the harvesting of their metadata, which implies that they fall into the “no objections” category. Only one respondent indicated that they objected to others using their metadata without permission. Of the 21 that did not object, nineteen cited conditions under which they would want their metadata used. The results are illustrated in Figure 10.

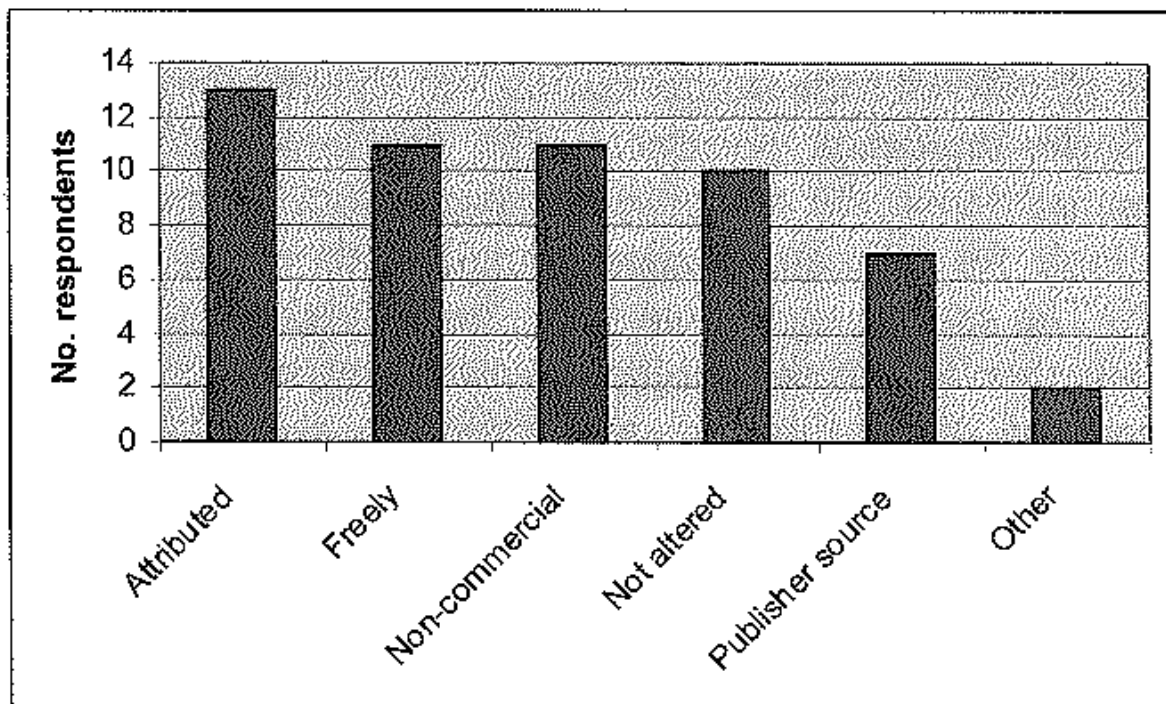


Figure 10 Conditions required by DPs on the use of metadata

The majority of respondents (68.4%) wanted the metadata to be attributed to their organisation. An equal number (57.9%) wanted the metadata to continue to be made freely available after it had been harvested, and for non-commercial purposes. A surprising 52.6% wanted to specify that their metadata remained unaltered. Of course, this would severely inhibit SPs wishing to enhance the metadata in order to build better services upon it.

Only seven respondents (36.8%) thought it should point to the original publisher source. What the survey meant by this was that the later metadata should provide a link to the original source as opposed to just acknowledging the original source. However, it may have been that respondents misunderstood this question –

particularly as the word, “publisher” was used, as most Data Providers would not see themselves as publishers.

Just three main conditions of use were listed by SPs. Three respondents stated that they would only allow enhanced metadata harvesting subject to prior agreement. Two made reference to the OAI provenance schema, stating that this should be used “so that service providers don't end up being confused by where records originated.” One specified that “the commercial service allows their web service metadata to be harvested.”

5.8 Usefulness of standard metadata rights protection

Most DPs and SPs did not want their metadata to be used unconditionally, but nor did they wish to be as restrictive over the use of their metadata as copyright law would stipulate. It would seem logical. Therefore, that they found a way of communicating to third-parties exactly how they wanted their metadata used. Thus, respondents were asked if they saw any benefit in having a standardised way of describing the rights status of metadata. 77.2% of DPs and 76.9% of SPs thought that such a standard would be useful.

The consensus was that a standard would provide clarity for end-users, save duplication of effort by DPs, and enable SPs to build services based on the information more easily. One wrote “there will always be exception but if a large percentage of data could be covered by the standard levels then it may be worthwhile developing a set of standard levels and their associated terms and conditions”. Two pointed out that the solution would have to be machine-readable. As one stated, “There is [a] right statement of [the] OAI protocol. But we don't know how we can implement it, because the right statement is machine unreadable.”

One of the DPs that did not know, wrote, “If the standard had all the relevant options, maybe, but if it just made it needlessly complicated and fussed about unimportant matters, no.” Two others were doubtful that it was possible to develop a standard that provided both simplicity and flexibility.

6 Rights metadata solution

Although the academic author survey asked respondents to select from ODRL terms to describe the permissions, restrictions and conditions they wanted to be in place over their works, it is not assumed that the final solution will use ODRL. There are other Digital Rights Expression Languages (DREs) available, including XrML [10] (Extensible Rights Markup Language). However, at the current time, XrML does not have a Rights Data Dictionary of terms, thus rendering it unusable for the purposes of the research. The other option would be to adopt or adapt the Creative Commons licence solution [3] to describe academics' giveaway works.

Creative Commons (CC) provides a rights protection model consisting of four elements. A 'Choose a Licence' application allows creators to select the terms and conditions under which their work may be made available. Each set of terms and conditions is available in simple human-readable language in the form of a 'Commons Deed'. A URI with explanatory text inserted in a work links to the Commons Deed. Each Commons Deed is linked to a full legally enforceable licence and rights metadata is available in XML/RDF along with an RDF schema.

However, the licences available under CC do not fully match the requirements of academic authors as specified in the RoMEO survey. For example, all CC Commons Deeds use the term "Copy". However, the RoMEO survey showed that this term can be easily misinterpreted to mean 'plagiarise'. Although the CC licences themselves use the term 'reproduce', the repeated use of "Copy" may not encourage non-archivers to archive, and may be misinterpreted by end-users. All of the CC licences allow copying and interpret Copy as "to incorporate the Work into one or more Collective Works". However, 67% of RoMEO respondents wanted to limit or prohibit such aggregation activity. There was also concern that the highly-valued restriction, Exact Replicas, may not be the same as the nearest CC equivalent, No Derivative Works.

It may be therefore, that whilst the CC solution is a good one in its own right, it is not as suitable for academics' give-away texts as a specialist solution would be. The other option would be to create an application profile of ODRL to describe the uses that may be made of academics' open-access research papers.

In addition to the P, R & Cs chosen by the academics, two additional elements could be included in the application profile. The first would be a <TransferPermission> restriction element. This would ensure that any P, R and C's specified in the rights metadata would be transferred with subsequent copies of the work. The second would be an <Accept> condition element, ensuring that anyone using the work had agreed to abide by the terms set out in the rights metadata.

The application profile would allow academics to select from four options.

- i) RoMEO Core: Allowing display, give, print, save and excerpt as long as users accept the terms and conditions, attribute the author, make only exact replicas and transfer the permissions metadata with subsequent copies.
- ii) RoMEO Core Plus Aggregate: As above with an optional Aggregate permission
- iii) RoMEO Core Plus Non-Commercial: As i) with an optional non-commercial purposes condition

iv) RoMEO Core Plus Aggregate and NonCommercial: a combination of options ii) and iii).

If this option is chosen, an application profile of the ODRL XML schema for the purposes of the project would be made available on the project web site in addition to XML instances of the four RoMEO options for end-users simply to import into their metadata records. Further information on the rights metadata direction chosen by RoMEO will be presented at the conference.

7 Metadata protection solutions

The RoMEO metadata protection model expects to allow metadata to be protected at whole collection and individual record level. This will enable prospective Service Providers to check the whole collection policy at-a-glance, whilst ensuring that the usage policy of individual records, if separated from the collection, is still communicated to end- users. At the whole collection level the eprints containers <MetadataPolicy> element could be used. At an individual record level, we propose to use a metadata records optional about container.

7.1 Expressing rights at collection level

In response to the OAI Identify verb, a Data Provider may provide a description container. A description container that describes the content and policies of repositories in the e-print community has already been created [17]. The schema allows for a <metadataPolicy> element which, in turn, contains a <text> element and a <URL> element. Data Providers wishing to use the RoMEO metadata protection solution may use something like the following <text> element based on ODRL:

“This metadata record may be [harvested] [displayed, saved, duplicated, printed, given, excerpted from, aggregated and modified] for non-commercial purposes, as long as all subsequent incarnations of this record retain their associated rights metadata and provenance information.”

Alternatively, it could use the URI associated with a particular Creative Commons Licence.

7.2 Expressing rights at individual record level

To express the rights status of each individual record, the RoMEO solution expects use a specific XML instance that conforms to the standard ODRL expression and data dictionary schemas. This instance would be used within a record's optional about container, in the same way as the optional provenance element [18].

As DPs and SPs were not given lists of P, R and Cs to choose from as the academic authors were, we had to interpret which permissions best would characterise to the activities necessary for “harvesting”. We also had to consider which of the restrictions and conditions equated to the ‘conditions of use’ suggested by the two groups.

It was decided that the following permissions would allow the necessary harvesting activities: Display, Print, Save,

Duplicate, Give, Excerpt, Aggregate, and Modify. The conditions seemed to equate to two restriction elements: Non-commercial purposes and Transfer permission, and two condition elements: Attribution (of the metadata copyright owners) and Accept (terms and conditions). Again, further information on the rights metadata direction chosen by RoMEO will be presented at the conference.

8 Conclusions and next steps

The number of academics assigning copyright to publishers may mean that the RoMEO rights metadata solution only becomes widely adopted in the long-term as more academics retain their rights in their work. However, our assessment of how academics wish to use open-access works demonstrates that the rights metadata we propose should provide academics with all the permissions they need in order to get the full benefit from open access research [8].

The metadata protection solution should have a more immediate impact, as the copyright in metadata records and the database right in metadata collections both (usually) belong to the Data and Service Providers that generate them. There are still issues to be resolved regarding the metadata protection solution, particularly the practicalities of adding in additional rights holders as metadata is harvested and enhanced by Service Providers. It may be that the OAI provenance verb can be utilised in some way for this purpose. Also, DPs supplying metadata may want some way of demanding that they receive a copy of any enhanced versions for their own purposes.

Once the ‘alpha’ solutions to the rights metadata and metadata protection problem are complete, the next step will be to put them out to the community for consultation before generating a beta version. It is hoped that they will encourage the uptake of author self-archiving by academics and the safe disclosure and harvesting of metadata by OAI Data and Service Providers

Acknowledgements

The RoMEO Project team wishes to acknowledge the UK Joint Information Systems Committee for providing the funding for this research.

References

- [1] Association of Learned and Professional Society Publishers. URL: <http://www.alpsp.org/> (2003)
- [2] S. Bird. White paper on machine-readable rights information. Pennsylvania, University of Pennsylvania. <http://www ldc.upenn.edu/sb/oai/rights.html>. (2001).
- [3] Creative Commons. URL: <http://creativecommons.org/> (2003)
- [4] R. Crow. The case for institutional repositories: a SPARC position paper. Washington DC, The Scholarly Publishing and Academic Resources Coalition: <http://www.arl.org/sparc/IR/ir.html>. (2002)
- [5] M. S. Frankel. Seizing the moment: scientist's authorship rights in the digital age. Report of a study by the American Association for the Advancement of Science, American Association for the Advancement of Science. <http://www.aaas.org/spp/sfrr/projects/epub/finalrept.html>. (2002).
- [6] E.A.Gadd, C.Oppenheim and S.Probets. "RoMEO Studies 1: The impact of copyright ownership on academic author self-archiving." *Journal of Documentation*. 29(3) [In Press] (2003)
- [7] E.A.Gadd, C.Oppenheim and S.Probets. "RoMEO Studies 2: How academics want to protect their open-access research papers." *Journal of Information Science*. 29(5) [In Press] (2003)
- [8] E.A.Gadd, C.Oppenheim and S.Probets. "RoMEO Studies 3: How academics expect to use open-access research papers." Submitted to *Journal of Librarianship and Information Science*. (2003).
- [9] E.A.Gadd, C.Oppenheim and S.Probets. "RoMEO Studies 4: The author-publisher bargain: an analysis of journal publisher copyright transfer agreements." Submitted to *Learned Publishing*. (2003).
- [10] E.A.Gadd, C.Oppenheim and S.Probets. "RoMEO Studies 5: The IPR issues facing OAI Data and Service Providers." Submitted to *Journal of Information Law and Technology*. (2003).

[11] S. Harnad. 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>. (2001).

[12] S. R. Kulkarni. "All professors create equally: why faculty should have complete control over the intellectual property rights in their creations". *Hastings Law Journal*, 47 (November 1995): p. 22-256.

[13] N. McAllister. "Freedom of expression: emerging standards in rights management". *New Architect Magazine* March <http://www.newarchitectmag.com/documents/s=2453/new1011651985727/index.html>. (2002).

[14] S. Morris. "Guest Editorial. Authors and copyright." *Learned Publishing* 13(2). pp. 75-76. (2000).

[15] Open Archives Initiative. URL: <http://www.openarchives.org/> (2003)

[16] Open Archives initiative. Registered Data Providers URL: <http://www.openarchives.org/Register/BrowseSites.pl> (2003)

[17] Open Archives initiative. XML schema to describe content and policies of repositories in the e-print community. URL: <http://www.openarchives.org/OAI/2.0/guidelines-eprints.htm> (2003)

[18] Open Archives initiative. XML schema to hold provenance information in the "about" part of a record. URL: <http://www.openarchives.org/OAI/2.0/guidelines-provenance.htm> (2003)

[19] Open Digital Rights Language. <http://odrl.net/> (2002) [20] Principles for emerging systems of scholarly publishing,

Association of Research Libraries.

<http://www.arl.org/scomm/tempe.html>. (2000).

[21] Project RoMEO. URL: <http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html> (2003)

[22] C. Rusbridge. "After eLib", *Ariadne* 26. <http://www.ariadne.ac.uk/issue26/chris/intro.htm>. (2001).

[23] Society for Scholarly Publishing. URL: <http://www.sspnet.org/> (2002)

[24] C. Tenopir and D. W. King. "Reading behaviour and electronic journals." Learned Publishing 15(4): pp. 259-265. (2002).

[25] Ulrich's Periodicals Directory Database. (2003). [26] H. Woodward, C. McKnight, C. Pritchett, and F.

Rowland, Cafe Jus: commercial and free electronic journals

User study. London, British Library. Research and Innovation

Report. pp. 1-52. (1997).

[27] XRML (eXtensible Rights Markup Language) URL:

<http://www.xrml.org/> (2003)