RoMEO Studies 5: IPR issues facing OAI Data and Service Providers

Elizabeth Gadd, Charles Oppenheim, and Steve Probets
Department of Information Science, Loughborough University, Loughborough, Leics, LE11 3TU

Email: C.Oppenheim@lboro.ac.uk

This article is an unrefereed preprint.

Abstract

This paper is the fifth in a series of studies emanating from the UK JISC-funded RoMEO Project (Rights Metadata for Open-archiving). It reports the results of two surveys of OAI Data Providers (DPs) and Service Providers (SPs) with regards to the rights issues they face. It finds that very few DPs have rights agreements with depositing authors and that there is no standard approach to the creation of rights metadata. The paper considers the rights protection afforded individual and collections of metadata records under UK Law and contrasts this with DP and SP's views on the rights status of metadata and how they wish to protect it. The majority of DP and SPs believe that a standard way of describing both the rights status of documents and of metadata would be useful.

1 Introduction

The Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH) (2003a) provides a means by which metadata can be disclosed by Data Providers (DPs) and 'harvested' into services by Service Providers (SPs). The minimum metadata standard for OAI-compliance is unqualified Dublin Core (2003), although items disclosed under the OAI-PMH may have any number of associated metadata records in different formats. The development of the OAI-PMH was driven by the eprints community, although its application is by no means limited to the disclosure of metadata about eprints. Indeed, it is being adopted by a wide range of communities, both commercial and non-commercial, as an extensible and interoperable metadata exchange protocol. However, most of the early adopters belong to the e-prints or e-theses.

In 2002, the UK Joint Information Systems Committee (JISC) funded 14 projects under it's FAIR programme (Focus on Access to Institutional Resources). One of the objectives of the FAIR programme is to explore the OAI protocol as a mechanism for disclosure and sharing a range of resource types (2002). These include eprints (Pinfield and McColl; 2002) and e-theses (Nixon; 2003) of course, but also other resources such as museum objects, video clips and images**.** One of the projects funded under the FAIR programme is the RoMEO Project (Rights Metadata for Open archiving). The purpose of the RoMEO project is to investigate the IPR issues related to the self-archiving of research papers under the OAI-PMH, and the subsequent disclosure and harvesting of metadata about those papers by Data Providers and Service Providers. The project has two main aims: to develop some simple rights metadata by which academics can protect their research papers in an open-access environment, and to develop a means of protecting the metadata disclosed by DPs and harvested by SPs under the OAI-PMH.

To inform the development of the rights metadata and metadata protection solutions, a series of surveys of key stakeholders were carried out. Our survey of 542 academic authors has been described elsewhere (Gadd, Oppenheim, and Probets; 2003a, 2003b). A survey of journal publishers has been performed, and a related analysis of 80 journal publishers' Copyright Transfer Agreements has also been described elsewhere (Gadd, Oppenheim, and Probets; 2003c, 2003d). This paper, the fifth in the RoMEO Studies series, reports on the results of our surveys of Data and Service Providers.

The aim of the DP survey was to ask Data Providers about their agreements with authors; any copyright protection afforded the documents themselves; and their views on the rights status of any

metadata they create. The aim of the SP survey was to ask Service Providers about the rights issues they face when harvesting, enhancing and disclosing metadata.

2 Methods adopted

Two online questionnaires (Project RoMEO; 2003a, 2003b) were designed in consultation with the Project Advisory Board, a Service Provider and a Data Provider. The DP questionnaire was divided into five sections: A – About You, B – Your agreement with authors, C – Copyright status of documents, D – Copyright status of your metadata and E - Other. Section A collected information on the name, location, and type of Data Provider. Section B collected information on the relationship the DP had with their depositing authors. Section C asked questions about the copyright status of documents and means of protecting copyright through rights metadata or technical protection measures. Section D looked at the IPR status of metadata and how they might want that protected. Section E asked if they had encountered any other rights issues not covered by the questionnaire.

The SP questionnaire was also divided into five sections: A – About You, B – Harvesting metadata, C – Enhancing metadata, D – Rights metadata (and rights of metadata) and E - Other. Section A collected information on the name, location, and type of Service Provider. Section B collected information on the harvesting activities of SPs and whether they checked the rights status before harvesting. Section C asked questions about the copyright status of any enhancements made to harvested metadata. Section D asked whether SPs thought that standardised rights metadata and metadata protection solutions would be useful. Section E asked if they had encountered any other rights issues not covered by the questionnaire.

The surveys were advertised in November 2002 via a number of email discussion lists and newsgroups including:

- OAI eprints – for those involved in using the OAI-PMH to disclose and harvest metadata about eprints;
- FAIRProjects – aimed at UK JISC-funded FAIR Projects that are mainly using the OAI-PMH to disclose institutional resources;
- Open Archive Forum – for European projects involved in open-access projects using the OAI-PMH;
- OAI-General – a forum for discussing OAI issues of a non-technical nature;
- OAI-Implementers – is a forum for technical discussions about the OAI protocol;
- American Scientist September-98 Forum- for the discussion of open-access and scholarly communication issues.

A list of Data Providers registered with the OAI was obtained from the OAI and the survey circulated to them. A link to the online survey was also placed on the home page of the Open Archives Initiative (2003a) web site.

3 Response rate & demographics

3.1 Data Providers

Twenty-two Data Providers responded to the survey. In April 2003, there were 79 Data Providers registered with the Open Archives Initiative site (2003b). The respondents to this survey thus represented 27.8% of those registered. Respondents were based in seven countries (see Table 1). The largest groups were based in the US (36.3%) and the UK (22.7%).

US	8
UK	5

Germany	3
Italy	2
Sweden	1
France	1
Austria	1
Everywhere'	1

Table 1 Countries in which respondents were based

Respondents were asked whether they disclosed only metadata, or whether they both disclosed metadata and provided access to the full-text. The majority of respondents (19 or 86%) both disclosed metadata and provided access to the full-text of documents. The other three (14%) that disclosed only metadata were asked to move straight on to Section D on the copyright status of metadata.

3.2 Service Providers

Thirteen Service Providers responded to the survey. In April 2003 there were only 12 Service Providers registered with the Open Archives Initiative, so the thirteen responses we received, although a small number, represented a very good response rate. (It is not a requirement for Data and Service Providers to be registered with the OAI). Respondents were based in six countries (see Table 2). The largest groups were based in the US (38.4%) and the UK (30.7%).

US	5
UK	4
Germany	1
Netherlands	1
Australia	1
France	1

Table 2 Countries in which respondents were based

Respondents were asked what kind of OAI Service they provided and were given a choice of three answers: "We just provide access to our own data", "We provide access to our own data and harvest other data", or "We just harvest other data". The first answer would indicate that the respondent was really a Data Provider, rather than a Service Provider (that by definition harvests others' data) and would be directed to Section D for their views on Rights Metadata. The results are illustrated in Figure 1 below.

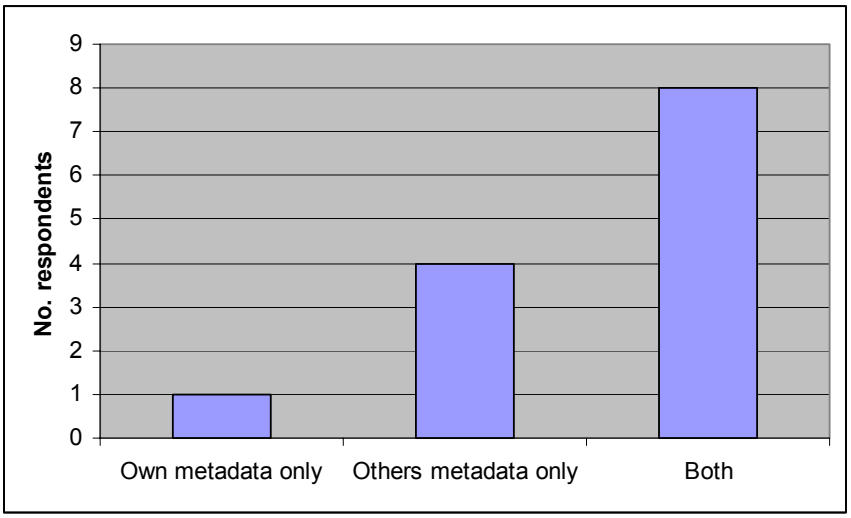

Figure 1 The types of metadata services provide access to

It can be seen that the majority of respondents provided access to both their own metadata whilst also harvesting others' data. Four harvested only others' metadata and, perhaps not surprisingly, only one respondent categorised themselves as a Data Provider by stating that they harvested only their own metadata.

4 Data Providers agreements with authors

4.1 Licence agreements

The Data Providers were firstly asked about the rights agreements they held with authors (depositors). The first question asked "Do you ask authors to sign a licence agreement before they can start depositing documents with you?". Despite the fact that only 19 respondents said they provided access to full-text, there were 20 responses to this question. Presumably the unexpected respondent answered 'No'! The responses are illustrated in Figure 2.

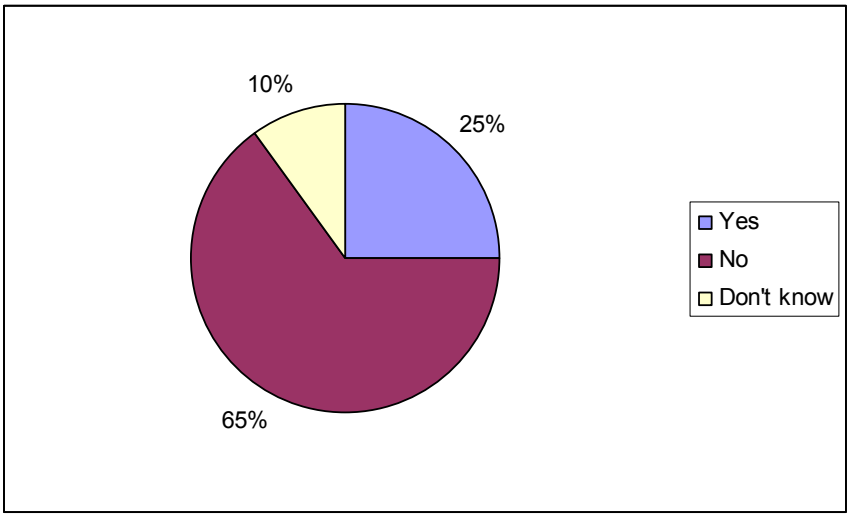

Figure 2 Do respondents ask authors to sign licence agreements?

It can be seen that the majority of respondents (13) did not have licence agreements with their authors. One-quarter (5) did so and two respondents did not know.

4.2 Rights warranties required of authors

The next question asked DPs "Do you ask authors to warrant that they have the right to deposit a document with you (i.e. copyright has not previously been assigned)?". They were given a list of four options plus an 'other' category and asked to select as many as applied. The options were:

- No, we **trust** that authors will only deposit documents they have the right to
- No, but we provide a **general warning statement** explaining the IPR implications of deposit
- Yes, our **licence agreement** asks authors to warrant only to deposit documents they have the right to
- Yes, **every time** an author deposits a document they must warrant that they have the right to do so

There were 20 responses to this question which are illustrated in Figure 3.

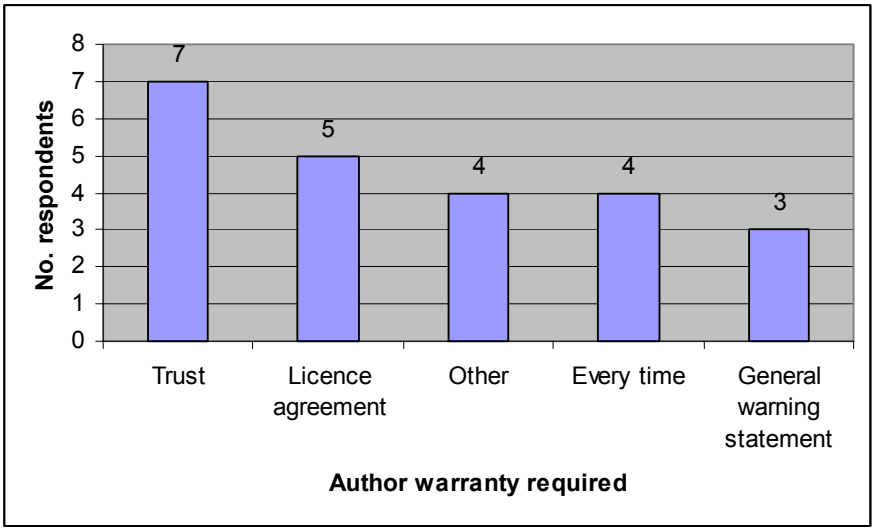

Figure 3 Rights relationship between author and Data Provider

The total does not add up to 20 as three respondents selected more than one option. The results show that the largest group of respondents (31.8%) took it on trust that the author had the right to deposit a document with them. Five asked authors to warrant they had the right to deposit via their licence agreement and four asked authors to do so every time they made a deposit. Three provided a general warning statement about the IPR implications of deposit. Four selected the 'Other' option. Two stated that they were currently investigating their options. One stated that, "The warning statement is the one provided by the software". Another commented, "we own all the content that we make available", thus warranties from depositors were not required.

Respondents were also asked to rate the success of their approach as very successful, successful, unsuccessful or very unsuccessful. Of the seven that stated they trusted their depositors, five rated this approach as very successful and two as successful. Those requesting just a licence agreement had a mixed response, one rating being very successful, one successful and one unsuccessful. The respondent rating it unsuccessful explained, "There are relatively few deposits of full-text theses so far. We fear that the legal wording of the license agreement might frighten the authors". However, the Provider that rated it very successful stated, "We have 700 records and have not had any complaints re copyright".

The two respondents that selected both a licence agreement and a warranty every time the author deposited rated this approach as successful or very successful. One commented, "The agreement text is displayed about the DEPOSIT button to remind the authors and that seems to work fine. Also, in some cases, a department head or other significant authority may provide a blanket permission for all records, such as a closed technical report series." The other said, "Authors like it, without exception."

Another respondent that selected only "every time" also rated this approach very successful. They wrote, "No one knows what needs to be done. Forcing authors to read think and click at least alerts them to the question." The fourth respondent in this category didn't answer this question.

All three respondents that offered just a general IPR warning statement found this approach successful. However one wrote, "I say "successful" but the papers are deposited by the archive's staff, so we can't rate the success of this approach." Another stated, "We have received no complaints to date regarding IPR issues."

4.3 Double checking

The next question asked Data Providers, "Do you double-check the copyright status of documents deposited with you (e.g. with publishers)?". (The eprints.org software allows only e-prints administrators to make content publicly available, giving them the opportunity to check it's IPR status (eprints.org; 2003; Nixon; 2002) Respondents could select from Always, Never, Sometimes and Don't Know, and were invited to give any comments. Eighteen responded to this question. None selected 'always'. The results are illustrated in Figure 4 below.

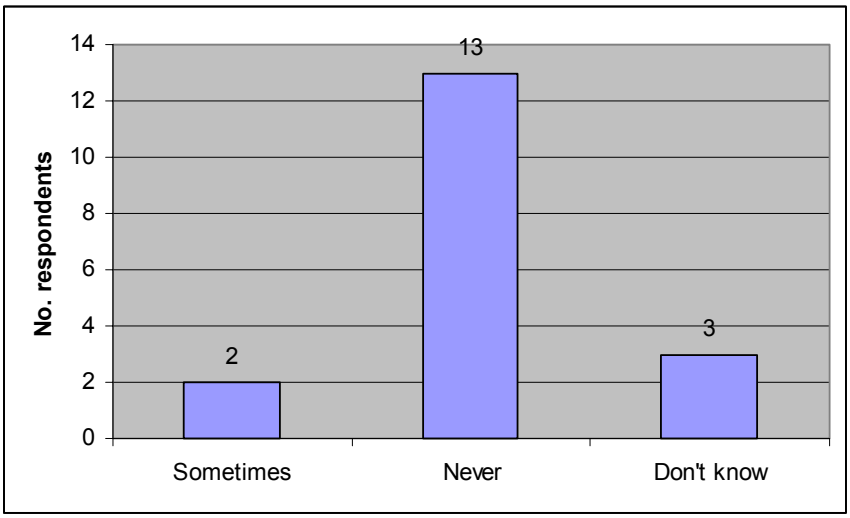

Figure 4 Do Data Providers double-check the copyright status of items deposited?

Only two DPs stated that they sometimes double-checked the copyright status of documents deposited. One explained, "We check PDF/PS submissions for copyright statements, author supplied TeX and HTML not checked". The majority (72.2%) said they never did so. Three explained that they didn't need to. This was either because the documents came directly from publishers, or because staff had already clarified the rights situation, or because they relied on the author's click-through rights warranty. However, one wrote, "Don't care. If publisher approaches us to remove a document, we ask for proof, and if the proof seems convincing, we would remove. But it has never happened yet (perhaps because we have only 1500 documents so far)."

4.4 Rights authors license to Data Providers

The final question in this section asked DPs if they asked authors to license them any rights to manage their work. They were given a list of three options and an "other" category. The options were to distribute the work; permanently store the work; and to make back-up copies of the work. Responses are illustrated in Figure 5.

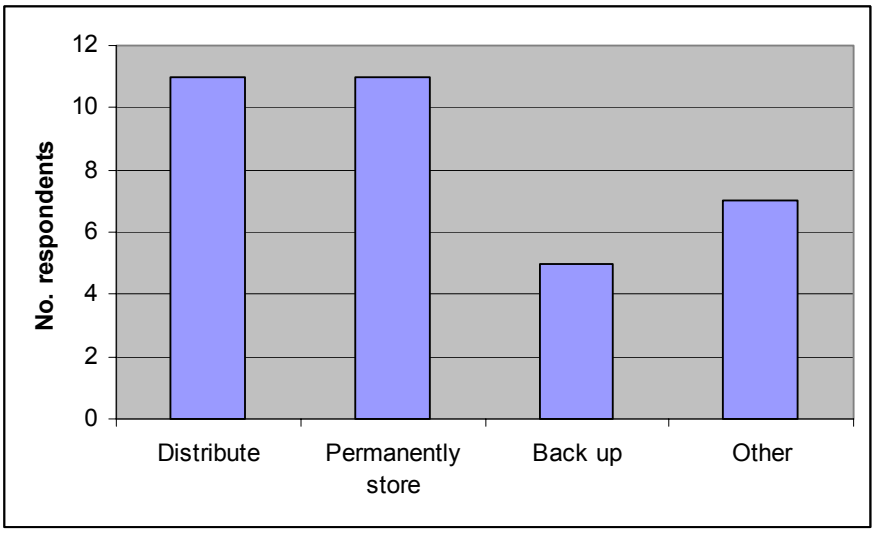

Figure 5 Permissions that DPs require of authors

Five respondents selected all three options. A further five selected both 'distribute' and 'permanently store'. It can be seen that these two options most commonly appeared in Data Provider/author agreements. Four of those selecting the 'Other' category said that these requirements were "implicit" in their agreement with authors. "Well, we don't ask, its kinda implied", wrote one respondent. Two asked for additional rights. One asked for rights to make copies "available to others library-wise", which may have meant via Inter-Library Loan. Another quoted from their agreement thus:

> Our agreement states: "I hereby grant to the [Data Provider] the irrevocable, non-exclusive royalty free right to reproduce, distribute, display, and perform this work in any format..."

4.5 Do Data Providers add value to documents?

Copyright law states that the copyright owner in a literary work, such as an eprint, is usually the author (Great Britain; 1988). However, if a third-party makes a significant contribution to the work, they could become a joint copyright owner. Also, publishers may own copyright in the typographical layout of a work. DPs were therefore asked, "Do you add any value to the documents yourselves (i.e. editing, formatting, abstracting etc.)?". Nineteen responded to this question, of which eight (42.1%) said 'No' and 11 (57.9%) said 'Yes'. Respondents were asked to specify the value they added to the documents and eight did so. These are listed below:

- Reference linking
- Cross-linking to references (using CrossRef)
- Format conversion
- Formatting SGML (EAD, TEI-Lite) for html output
- Abstracting
- Subject indexing
- Editing
- Word-indexing of text documents
- Entering identifying metadata for images.
- Presentation of author supplied metadata

It is arguable whether some of these enhancements add value to the documents themselves or to the metadata. This is discussed below.

5 Protecting documents with rights metadata

As one of the proposed outputs of the RoMEO Project was the development of some simple rights metadata describing the conditions of use of eprints, DPs and SPs were asked about their existing use of rights metadata.

5.1 Creation of rights metadata by Data Providers

Data Providers were asked, "Do you create (or disclose) any rights metadata describing the copyright status of the documents you point to?" Seven said that they did, twelve said they did not and one did not know. The seven that said they did create or disclose rights metadata were asked: 1)who created it, 2) what schema they used, and 3) did they know of any Service Providers that used it.

In answer to the first question, five said that the Data Provider created it, and two said the depositors did. Only five responded to the second question. One specified 'OAI', one 'DC' (Dublin Core, which has a rights element), and another 'ORyX.xsd' (a schema developed for the Open Language Archives Community (OLAC) (2003)). One used a 'proprietary' schema and one used 'none'. Four answered question three, all of which said that that they did not know of any Service Providers that utilised their rights metadata.

5.2 Harvesting of rights metadata by Service Providers

However, when SPs were asked, "Have you ever knowingly harvested rights or permissions metadata describing what use may be made of documents (e.g. must only be used for non-commercial purposes)?" the majority (7) said yes. Four thought they had not harvested rights metadata and one did not know. Of the seven that had, three explained that by virtue of harvesting Dublin Core metadata that has a rights field, they had harvested rights metadata. Another said they had harvested "data policy statements" which equated to rights metadata. One stated that a university-based Data Provider, "provides metadata describing restricted access material."

5.3 Communicating rights metadata to end-users

Those SPs that had harvested rights metadata were asked, "do you have a means of communicating this information to end users?" Although only seven stated that they had harvested rights metadata in the previous question, nine responded to this one. Respondents were asked to explain their answer.

Five indicated that they had communicated rights information to end-users, three said they had not, and one did not know. Three of those that had explained that the rights information was displayed on the search results page. Another worked with a different model whereby, " Institutions that use the service sign a license agreement in advance. The terms and conditions are made available to end users through the interface." The other assumed that "restrictions are explicated in the Rights element of the metadata." However it was not clear how the end-user had access to that metadata.

5.4 Usefulness of rights metadata

Both DPs and SPs were asked whether they saw any benefit in having a standard way of describing how documents may be used (i.e., rights metadata)?".

5.4.1 Views of Data Providers

Twenty DPs responded. Sixteen respondents (76.1%) thought a standard way of describing how their documents should be used would be useful. One wrote that "clarity and consistency are always helpful". Another thought that "it would make it easier to explain the need for a license agreement to authors". One recommended,

> A closed set of attributes… Also a set for who *may* access it would be good, e.g., freely available online, requires free registration, only available to a limited set (e.g., University members), pay-to-view full text.

However, another thought that "Other than to say there is a copyright holder, this statement would be difficult to generalize."

Four did not know whether such rights metadata would be useful, although two qualified this. One wrote:

> Yes, provided that the papers deposited share the same nature (preprints), in order to "force" the academic publishers to permit this alternative form of publishing of the giveaway literature; No (I think), if the papers deposited are not preprints (dissertations, departmental [web pages], in general grey literature).

The other wrote:

> Probably yes. Our material can be used freely and gratis by anyone, so in a way it's not relevant to us. We indicate on the documents that they are 'Open Access'.

Only one respondent thought it would not be useful, and they specified why:

> I can't see how the various and unknown document usage scenarios could translate well into a standard description. The best thing for now, [in my honest opinion], is to describe the document (and metadata) usage policy in natural language, put that in the OAI identify request, assume service providers will read, interpret and follow the stated policy and rely on technical means to enforce it, if necessary.

5.4.2 Views of Service Providers

Thirteen SPs responded to this question, eleven of which thought a standard rights metadata solution would be useful. One thought it might 'possibly' by useful and one thought not. A number of Service Providers explained their response Three saw the benefit of such standards to end-users. As one wrote,

> This would be extremely beneficial for our end-users and for managing the data. Currently, rights information is too variable. If there were a yes/no/in certain circumstances toggle, it would be easier to make that info available to end-users in our interface.

Another thought this would, "encourage authors to self-archive their material, and to make this information as clear and unambiguous as possible to service providers." A third saw the possibility for enabling "standard services depending on rights."

The respondent that did not think rights metadata would be useful wrote, "[As far as I know] there haven't been any copyright concerns raised by authors who place their own works on the (freely-accessible) Web." The one that thought it would possibly be useful commented, "In general, I imagine it would be a benefit" however, it would "not [be] necessary, or even desired, for our application since our agreements override what would be in the record".

6 Technical protection measures

It was recognised that technical protection measures are often used instead of, or as well as, rights metadata or licences to protect documents from misuse. DPs were asked, "Do you use any technical

means of physically preventing certain uses of your documents (e.g. disabling print in Acrobat)?". Of the 20 respondents, 17 did not use such measures, two did and one did not know. The one respondent that did not know explained that although they did "not do this, the author depositing a PDF could but probably won't". The two that did use technical measures used them to restrict access to electronic theses and dissertations or to digital images to a defined user group.

7 Copyright status of metadata

An IPR issue of perhaps more importance to many Data and Service Providers than the rights status of documents is the rights status of metadata itself. Indeed, a paper on machine-readable rights information in the OAI by Bird (2001) only discusses the protection of rights over metadata.

7.1 Is there copyright in metadata?

Although the copyright status of a document is indisputable, the IPR status of an individual metadata record and a collection of metadata records is not so clear cut.

7.2 An individual metadata record

In UK Law, copyright subsists in: "original literary, dramatic, musical or artistic works" (S.1)

S.3(1)"literary work" means any work, other than a dramatic or musical work, which is written, spoken or sung, and accordingly includes—
(a) a table or compilation other than a database,
(b) a computer program;
(c) preparatory design manual for a computer program, and
(d) a database

3A. - (1) In this Part "database" means a collection of independent works, data or other materials which -

(a) are arranged in a systematic or methodical way, and
(b) are individually accessible by electronic or other means.

3A (2) For the purposes of this Part a literary work consisting of a database is original if, and only if, by reason of the selection or arrangement of the contents of the database the database constitutes the author's own intellectual creation.".

So, could an individual metadata record be classified as a literary work? There are two categories of literary work that a single metadata record could fall into: 1) a compilation (of data), and 2) a database.

Compilations are protected by copyright. A database will be protected by copyright if "by reason of the selection or arrangement of the contents of the database the database constitutes the author's own intellectual creation". Failing this, a database may be protected by the *sui generis* database right if there is nothing innovative in the way the contents are arranged (Giavarra and Oppenheim, 1997).

7.2.1 Compilations

There is little doubt that a metadata record could be considered a compilation of data. However, is such a compilation original? What does copyright law mean by original? There is no definition in UK law, so it would be up to a court of law to decide. Previous cases have afforded originality to works involving 'sweat of brow' as long as the material has not been copied from somewhere else. Does creating a metadata record involve sweat of brow? Arguably it does. Depending on the type of metadata record, there may be more or less intellectual effort involved. However creating even the simplest compilation of intrinsic metadata such as author, title, journal, ISBN, etc, will involve some intellectual effort and decision making. Extrinsic, qualitative metadata such as subject classification, authority control, indexing and abstracting certainly involve considerable 'sweat of brow'.

We conclude that an individual metadata record is probably protected by copyright. However the key word here is "probably".

Incidentally, the fact that two persons could come up with the same metadata record independently has no bearing on whether it could be considered original. As Oppenheim et al (2001) point out, "original…implies 'not copied'….Thus if two people take a photograph of the Houses of Parliament from the same spot, each work is original and will be protected by copyright."

7.2.2 Databases

The definition of database is fairly broad, and it has been shown that even a journal or newspaper could be considered a database in that they are collections of i) independent works; ii) arranged methodically; and iii) are individually accessible. So the question is, could an individual metadata record be considered a database in itself? The first criterion is that a database is a collection of independent data. A metadata record is a collection of data (i.e., metadata elements) but are they 'independent'? In one sense metadata elements are not independent from one another: 'author' relates to 'work' relates to 'journal' relates to 'volume', and so on. However, you could argue that each metadata element is independent from the others because each element (e.g., "Author: Elizabeth Gadd") does have a meaning of its own. Some elements, such as an abstract, would certainly be considered to be independent.

Assuming that metadata elements are independent, databases also have to be arranged systematically or methodically. Using pre-determined formats such as Dublin Core, AACR2, and so on, metadata elements are in most cases arranged methodically. Finally, each datum needs to be individually accessible; electronic metadata elements always are. For example, a library catalogue will allow a user to extract small portions of the records – such as a list of authors.

7.2.3 Original databases

Thus it appears that an individual metadata record could be considered a database. The next question is would it be regarded as an 'original' database by virtue of the selection and arrangement of its contents, and thus enjoy copyright protection? Most metadata records consist of information selected and arranged according to a pre-existing formula and thus may not be considered the author's intellectual creation. Some might argue that if a metadata record enjoys copyright protection as a compilation by virtue of the intellectual effort or 'sweat of brow' involved in its creation, then that same intellectual effort should grant it the status of a database also. However, others have argued not.

7.2.4 Non-original databases

If a database is not deemed to be 'original' according to the above definition, then it could still protected by the *sui generis* database right as long as it has the following qualification:

13 (1) A property right ("database right") subsists, in accordance with this Part, in a database if there has been a substantial investment in obtaining, verifying or presenting the contents of the database.

Section 12 states that substantial "in relation to any investment, extraction or re-utilisation, means substantial in terms of quantity or quality or a combination of both" and investment, "includes any investment, whether of financial, human or technical resources". So is substantial investment required in obtaining, verifying or presenting a single metadata record? This is unlikely. Investment is certainly required to create an individual metadata record, but it is unlikely that investment could be considered "substantial". *Collections* of metadata records are another matter.

7.2.5 Summary

It can thus be concluded that an individual metadata record may be considered an original compilation or database and thus enjoy copyright. Without doubt, the more creative effort that goes into the record (such as abstracting, indexing, etc), the more likely it is that that record enjoys copyright. Indeed, abstracts certainly do so. Interestingly, although a metadata record may be considered an *original* database and thus enjoy copyright, it is unlikely that it would enjoy database right because of the lack of a 'substantial' investment in its creation.

7.3 A collection of metadata records

A collection of metadata records would certainly be considered a database and would thus benefit from database right. However, as previously discussed, some databases also qualify for copyright protection, "if, and only if, by reason of the selection or arrangement of the contents of the database the database constitutes the author's own intellectual creation.". The question as to whether a database of metadata records is "the author's own intellectual creation" would have to be decided on a case-by-case basis.

7.4 Protection afforded by copyright and database right

The next question is, what protection does copyright law and database right provide for individual and collections of metadata records?

7.4.1 Individual records

If an individual record was deemed to enjoy copyright protection by virtue of being a literary work, then copyright will subsist in the record for the life of its author (or last surviving author if jointly created) plus 70 years. If the record was created by someone employed to create such records, then the copyright in it would belong to the employer and the lifetime is 70 years from the end of the year in which the material was published. During this time, the copyright owner has the right to prevent, or authorise a variety of acts, notably copying the work, issuing copies to the public, or renting or lending the work to the public. The major exceptions to these so-called restricted acts that people may draw upon are:

- Fair dealing for research or private study (soon to be changed under a proposed Statutory Instrument to fair dealing for non-commercial purposes or private study);
- Fair dealing for criticism or review;
- Incidental inclusion.

The only exception that is likely to be relevant to the copying of metadata would be for non-commercial research or private study and this exception is very limited. Thus in practice, anyone wishing to copy or adapt an individual metadata record should ask the copyright owner's permission before they do so.

7.4.2 Collections of records

Database right belongs to the maker of the database, namely, "the person who takes the initiative in obtaining, verifying or presenting the contents of a database and assumes the risk of investing in that obtaining, verification or presentation" or their employer if employed to create the database. It lasts for 15 years from the end of the calendar year in which the database was created. Databases that are continually updated arguably retain copyright protection in perpetuity. Database right is infringed when someone, without the consent of the owner of the right, extracts or re-utilises all or a substantial part of the contents of the database.

It is clear that any Service Provider wishing to make use of a Data Provider's metadata should seek permission to do so. However, whilst this is the theoretical legal position, in practice, many rights owners may not be concerned by such use.

8 The creation and use of metadata by Data and Service Providers

8.1 How do Data Providers create metadata?

As we have shown, the owner of copyright in an individual 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?". They were given four options: Author input (authors may or may not be helped in this process by server-side software during the submission process); Data Provider staff; a mixture of authors and data provider staff; or another source. Responses are illustrated in Figure 6 below.

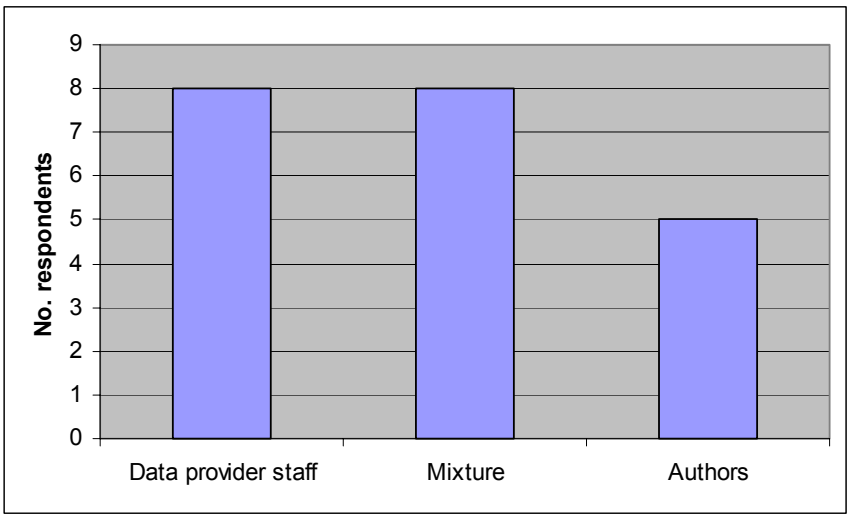

Figure 6 How metadata is generated

In the majority of cases, 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. Dspace's *Internal Reference Specification* (Bass et al, 2002) states that there are two categories of personnel that may alter the author's submitted metadata: 1) A*pprovers* that "can edit the…metadata to fix obvious errors", and ii) *Metadata Editors* that can perform other editing tasks.

8.2 Do Data Providers enhance metadata?

DPs were then asked if they enhanced or added any value to the metadata (e.g., check for accuracy, write abstracts, add subject indexing, etc.)?. Nineteen responded, of which nine said that they did and ten that they did not.

Of the nine that enhanced the metadata, four mentioned accuracy and/or quality control checks. Another four carried out abstracting and a further three indexing. One stated that they added in author link data (presumably so end-users can search easily for other material by a particular author). One wrote that they enhanced the record to the point that it could be converted to MARC 21 for addition to the library OPAC.

8.3 Do Service Providers enhance metadata?

SPs were also asked whether they enhanced or added any value to the metadata they harvested from DPs. 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 fell into the following activities. Where more than one respondent gave the same answer, the number doing so is indicated in brackets.

- 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

A fuller list of ways in which SPs can enhance metadata is given by McClelland, McArthur and Giersch (2002).

8.4 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. The options were: "Metadata are facts and there is no copyright in a fact", "There is only copyright in those parts of the record that involve intellectual effort such as the abstract", "The whole metadata record is protected by copyright", "The whole metadata record is protected by database right", and "Don't know". Twenty responded to this question. It must be borne in mind that database right only applies in the EU, thus non-EU respondents were unlikely to recognise the term. The results are illustrated in Figure 7 below.

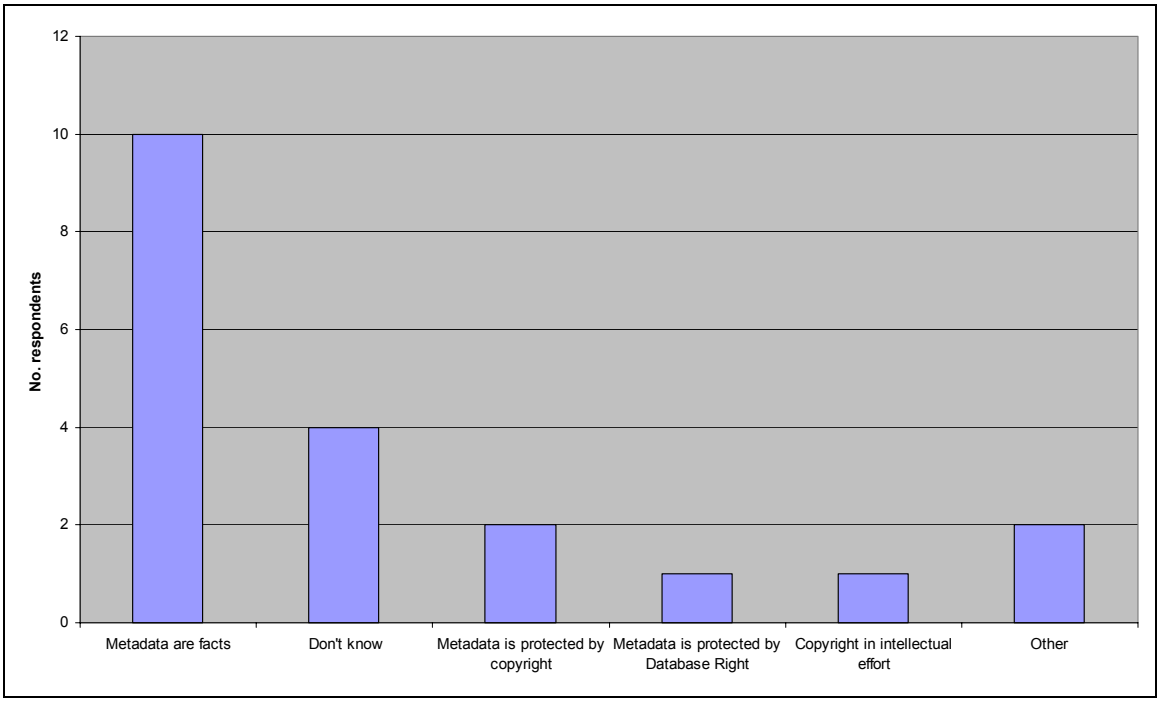

Figure 7 Views on the IPR status of an individual metadata record

It can be seen that half of the respondents (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 options were: "A collection of records is protected by database right, therefore no-one should systematically harvest them without permission", "A collection of records is usually protected by database right, but this right is implicitly waived in the OAI community", "There is no database right in our collection of records because they are not stored in a systematic way" and "Don't know". The responses of the nineteen answering this question are illustrated in Figure 8**.**

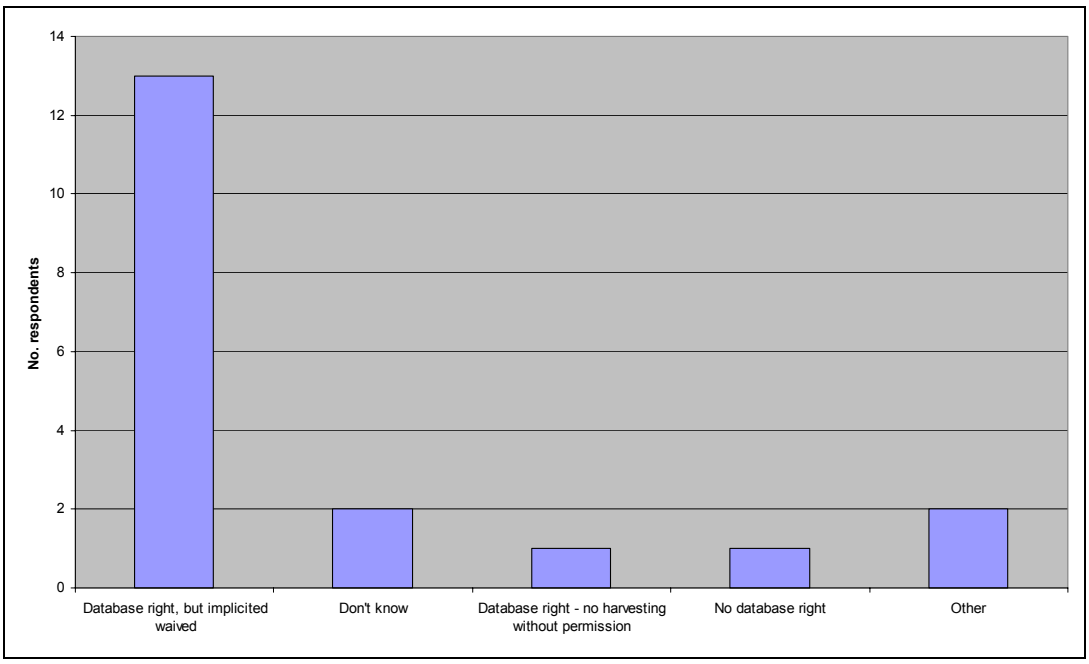

Figure 8 Views on the IPR status of a collection of metadata records

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.

8.5 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 give 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 9 below.

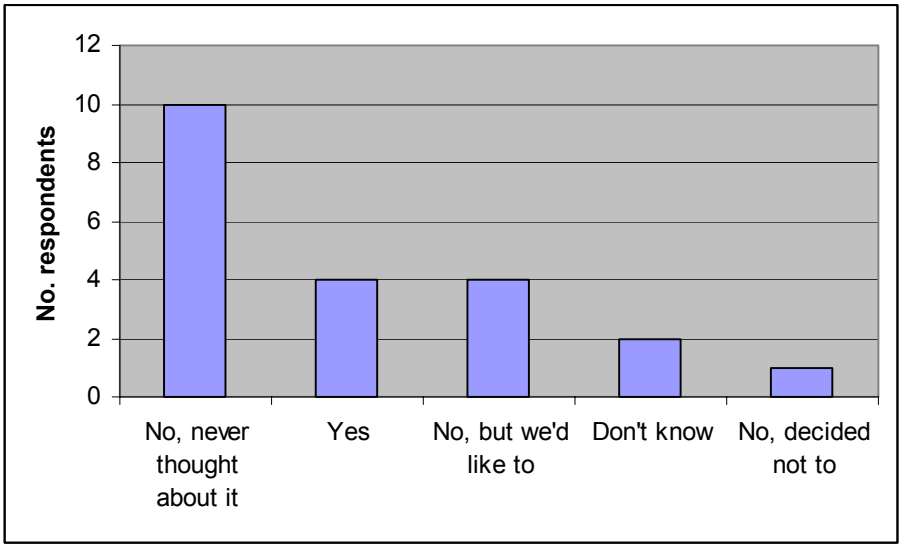

Figure 9 Did Data Providers assert rights in individual records?

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.

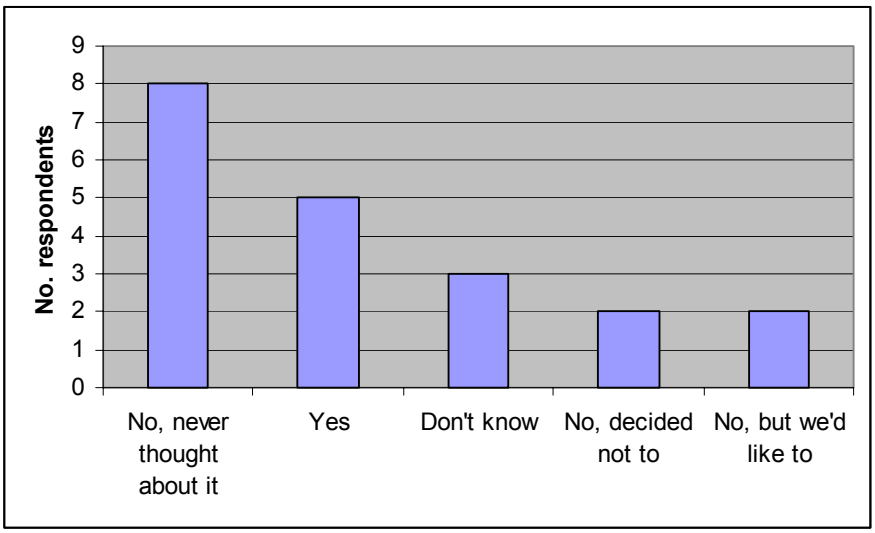

Figure 10 Did Data Providers assert rights in collections of records?

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. The third pointed to the URL of their copyright policy page which actually referred to the copyright status of documents rather than metadata.

8.6 Views on unacceptable use by Data Providers

DP respondents were asked what uses of their metadata would they consider unacceptable. Fifteen responded. Ten of those responses again mentioned commercial use. "Making it a part of a product available for sale", stated one respondent. "Selling access to our records", wrote another. One felt that any usage that "undermine[d] the open access ethos or philosophy" would be unacceptable, although they admitted, "it is difficult to see how we could prevent that". Three stated that they would not want the metadata altered or misrepresented. For example, "putting them in a deceptive context that implies they endorse or support something that they do not." Two felt it would be unacceptable if another organisation mirrored or re-published the whole collection "as is". Other unacceptable uses included claiming ownership, non-attribution of the original Data Provider, and the misuse of names and email addresses for spamming purposes.

The next question asked respondents whether they had had any incidents involving others using their metadata. Twenty-two responded of which only two (9%) answered that they had. One explained, "there was a part of the data once copied to another server, but with wrong attribution." The other said that they had experienced email address harvesting, although this was not via the OAI.

8.7 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) said that they did check, that metadata was implicitly free under the OAI, and 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. [They] Archive announcements about new collections to maintain a record of what is said with respect to its usage." 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.

8.8 Do Service Providers think there is copyright in their metadata enhancements?

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 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 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.

8.9 Did Service Providers want to protect those enhancements?

The four respondents that were concerned about protecting their rights gave explanations for their answer. However, none justified their response with reference to IPR law. One stated that "The automatic subject classification service is supplied by a commercial service who will require us to protect their IPR." Another felt they owned copyright in the compilation of enhanced records. They explained, "The individual records are the property of the Data Providers; the copyright to the compilation is ours." One believed they owned copyright in their thesaurus, although "its use is meant to be promulgated". The final respondent made the point that "The data is sensitive, e.g. citation impact", also, they did "not want data - which we are giving away - to be used for commercial purposes without our permission".

8.10 Were Service Providers happy for others to harvest their metadata?

The next question asked this group of respondents, "Would you be happy for other Service Providers to harvest your enhanced metadata?". The Arc service provider has experimented with implementing an OAI layer over harvested metadata so that other SPs can harvest it (Lui, Maley and Zubair, (2001). Four options were given; "Yes, unconditionally", "Yes, under certain conditions", "No" and "Don't know". Those selecting "under certain conditions" were asked to list the conditions under which they would be happy for the harvesting to take place. Eleven responded to this question.

No SP answered that they were happy for others to harvest their enhanced metadata unconditionally. However, over half (6) said they were happy for it to harvested under certain conditions. Three respondents said that they would not want others to harvest their enhanced metadata, and two did not know.

8.11 Were Service Providers happy for others to harvest and further enhance?

The final question on this subject asked respondents if they would be happy for other Service Providers to harvest *and further enhance* their enhanced metadata. The same four choices were given as for the previous question, and once again respondents were asked to list the conditions under which they would permit this activity. Twelve responded. Apart from an additional respondent in the "Yes conditionally" category, the results were identical to those for the previous question.

8.12 Conditions of use required by Data Providers

Having ascertained that neither Data nor Service Providers wished to allow the unconditional use of their metadata, questions were asked as to the conditions under which they wanted to make it available. 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”, and “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 11**.**

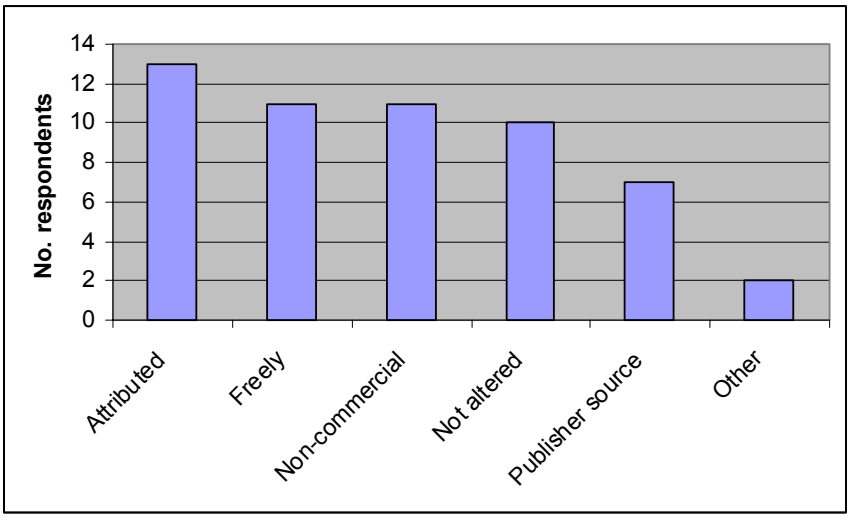

Figure 11 Conditions required by DPs on the use of metadata

The majority of respondents (68.4%) wanted the metadata to be attributed to their organisation. Fifty-eight per cent wanted the metadata to continue to be made freely available, and for non-commercial purposes. A surprising 52.6% wanted to specify that their metadata remained unaltered.

Only seven respondents (36.8%) thought the metadata 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. Two respondents selected the ‘Other’ category. They both added qualifications to the “Not altered” condition. The first said they would not want the record, “altered and re-distributed without our permission”, however, it “may be altered ‘in house’.” The second clarified that “non-alteration includes attributing authorship, locus, etc.”

8.13 Conditions of use required by SPS

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.” With regard to the conditions required for harvesting and further enhancing, the list was fairly similar. Two stipulated the use of the provenance verb. Two specified attribution or “credit given to source”. Two stated this could only be done subject to prior agreement and one repeated their previous comment that commercial services should allow their web service metadata to be harvested in the same way that theirs was freely harvestable.

9 Usefulness of metadata rights protection

Having seen that most DPs and SPs do not wish to be as restrictive over the use of their metadata as copyright law would stipulate, it would seem logical that they found a way of communicating to third-parties exactly how they wanted their metadata used. As McClelland, McArthur and Giersch (2002) have written:

> To some extent, the problems that arise from missing elements in imported metadata — or from elements in imported metadata whose values are from different vocabularies used in the importing digital library environment — might be solved if catalogers were free to modify or enrich imported metadata by adding values for unpopulated elements. The problem is the lack of clarity about the conventions and/or courtesies surrounding the use of, and freedom to change, the shared metadata.

Thus, respondents were asked if they saw any benefit in having a standardised way of describing the rights status of metadata.

9.1 Data Providers

Twenty-two DPs responded of which seventeen (77.2%) said yes. Three said they did not know, and two said no. One of those answering no explained, "As long as [the conditions of use are] met let's not stifle innovation." Of course, such conditions of use would be exactly the type of information contained in the metadata rights protection solution.

One of the respondents 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." Ten of those that thought a standard metadata usage code would be helpful gave additional comments. The consensus was that a standard would provide clarity for end-users, and save duplication of effort by Data Providers.

However, two were doubtful that it was possible to develop a standard that provided both simplicity and flexibility. Another recognised that "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 can we implement it, because the right statement is machine unreadable."

9.2 Service Providers

Thirteen SPs responded to this question. Ten said that they did see a benefit, two said possibly and one said no. The one respondent that did not think this would be useful said,

> I would place (freely-accessible) OAI service providers in the same category as Web search engines. By placing information on the Web, it is inferred that data will be harvested and re-exposed by Web services. Should the author not wish this there are a multitude of methods for preventing indexing by Web services.

Of the two that thought it might possibly be useful, one raised a question about the IPR status of metadata in the first instance and whether it was protected under law. Their argument was that if metadata is protected under law, then it would not need to be protected via metadata rights information. The other respondent pointed out that their harvesting activities were governed by licence agreements and thus, again, would not require protection via other means.

The majority of respondents that thought that a standard way of describing the rights status of metadata would be useful made similar comments to those about document protection. Two saw the benefits with regard to the protection of their own metadata on "reharvesting". One reiterated the

"commercial and intellectual rights invested in metadata" which "need to be acknowledged." Another thought it would help "avoid disputes".

10 Discussion

10.1 Agreements with authors

Only one-quarter of responding DPs had licence agreements with their depositing authors. Half of all respondents either trusted their depositors only to mount documents that they had the right to, or issued a general warning statement on this issue. In some cases this may have been because the DP was fairly new and had yet to finalise their policies and procedures. However, it is important that DPs protect their own interests even in an open-access environment. A DP making available a copyright-infringing work will be responsible (under UK law at least) for secondary infringement. Whilst there have been no cases to date where publishers have taken individual authors to court for infringing the copyright in their own paper, there are precedents for publishers taking action against infringing intermediaries such as online hosts.

Previous studies have shown that academic authors may ignore the terms of their agreements with publishers in order to self-archive their works (Pinfield, 2001; Hunt, 2002). DPs would do well, therefore, to develop a standard licence agreement that each depositor has to complete as part of the submission process. As one respondent warned, the wording of such an agreement would have to be clear and simple whilst covering all the appropriate issues. Such an agreement should also include a statement about the rights status and dissemination of metadata. The DP should ensure they retain those agreements for their records. It may beneficial if a model DP-depositor licence is developed to save each DP from having to 'reinvent the wheel'. In the meantime, CalTech (2001) and DSpace (2002) have developed such agreements that may provide a model for others to learn from.

10.1.1 Author warranties

The validity of the 'success' ratings of each of the approaches towards warranties from authors was affected by DPs' varying definitions of success. Success to some seemed to be a lack of complaints from third parties. However, this is clearly not the same as success in the face of legal action. It may be that institutional and subject-based repositories are not currently seen as threat to publishers and therefore they are unlikely to police them. However, in future this situation may change and DPs would be wise to be prepared.

10.1.2 Author permissions

Just under half the responding DPs did not ask depositing authors to license them any rights to deal with their works. Four stated that these permissions were "implicit" – just as it is implicit that a letter to a newspaper may be published. However, it is not wise to rely on such assumptions. For example, the DP may wish to adopt a policy that all deposited works will be permanently available. The depositor, whilst aware that her work will be made freely available, may not be aware of the permanency policy. It is important that all such rights are made explicit at the outset to both protect the DP and inform the author. SPARC (Crow, 2002) have recommended that institutional repositories,

> need each contributor to grant the institution an irrevocable, non-exclusive, royalty-free license to distribute the content, to translate its format for the purpose of digital preservation, and to maintain the content in perpetuity.

Our research showed that while 50% of DPs ask for rights to distribute and maintain the content in perpetuity, less than one-quarter considered the right to preserve (or 'back up' in our survey).

10.2 Copyright status of documents

A surprisingly high number of DPs added 'value' to the documents that were deposited. The important question is, does the value they add give them any rights over the work? As outlined above, UK copyright law states that an "author", "in relation to a work, means the person who creates it." However, " in the case of the typographical arrangement of a published edition, [this is] the publisher." There may also be works of joint ownership described as "a work produced by the collaboration of two or more authors in which the contribution of each author is not distinct from that of the other author or authors".

Of the list of ways in which DPs believed they added value to the documents, three would not result in the creation of new intellectual property. These include reference linking and word-indexing of text documents. In three cases (including format conversion and the presentation of author-supplied metadata) a new typographical copyright would be created. A new copyright would be generated by the writing of an abstract, and possibly by editing, depending on the nature of the edits. Heavy editing, including the re-writing of passages may qualify the editor for joint copyright ownership. In two other cases, new copyright may be created, but in the metadata record rather than the document itself. These were the addition of subject indexing and the provision of descriptive metadata for images. DPs would need to consider whether they wished to assert and protect any new copyrights they create.

10.3 Creation of rights metadata by Data Providers

Although just over one-third of DPs stated that they created or disclosed rights metadata describing the documents, no standard approach emerged. Two appeared to use Dublin Core, and whilst this has a rights element, there has been no guidance to date as to how it should be used. Some have used it to provide details of the rights holder; others have used it to describe permissions information. One community (OLAC) has developed its own metadata set which includes a rights element. Another DP has developed its own proprietary schema. However, the extensibility of such rights metadata solutions to other communities is doubtful. The fact that no DP knew of any Service Provider (SP) that used their rights metadata may be a reflection of the lack of standardisation in this area.

10.4 Harvesting and disclosing of rights metadata by Service Providers

Surprisingly, the majority of SPs (7) stated that they had harvested rights or permissions metadata "describing what use may be made of documents". However, three of the seven indicated that by virtue of harvesting DC metadata, they knew they had harvested rights information, because DC has a rights field. Of course, as the OAI specifies only unqualified DC as a metadata standard, not every DP will use the rights field.

Another surprise was that five SPs stated that they made the harvested rights metadata available to end users. This bodes well for any standardised approach to rights metadata, that services are already developing ways of disclosing such information to end-users.

10.5 Usefulness of rights metadata

Despite this lack of standardisation, an overwhelming majority of DPs believed that a standard way of describing how their documents may be used (i.e., rights metadata) would be useful. Some were not sure how this could be successfully achieved despite the potential benefit and others recognised that it would be easier for distinct subsets of material such as eprints. As the majority of respondents did not use technical protection measures to protect their full-text, there would certainly be advantages in providing a metadata solution.

As with the DPs, the overwhelming majority of SPs thought that having a standardised way of describing the rights status of documents and metadata would be useful. The consensus was that this would assist both in the SP's harvesting activities, and in alleviating confusion amongst end-users.

There was some misunderstanding about the purpose of rights metadata. One respondent clearly saw it as a means of communicating between the DP and the SP about the rights status of a document, thus rendering it irrelevant if the DP and SP had a licence agreement. However, one of the principal audiences for document rights metadata is, of course, the end-users.

10.6 Views on copyright ownership of metadata

10.6.1 Data providers

It was clear from responses to the two questions on the copyright status of metadata records and collections that most DP's did not perceive there to be any real IPR issues here. The largest group of respondents believed that individual metadata records were facts "and there is no copyright in a fact". Sixty-eight per cent acknowledged that their collections theoretically qualified for database right, but that this was implicitly waived in the OAI community. Not surprisingly then, when asked whether they asserted the copyright status of their individual or collections of metadata records, the largest group of respondents in each case answered, "No, never thought about it". Slightly more had developed means of protecting their metadata collection than individual records, but twice as many stated that they would like to be able to protect individual records than whole collections.

However, the subsequent question about the acceptable use of metadata appeared to raise awareness amongst DPs as to the benefits of metadata protection. Although the majority had not thought about metadata protection before, 90% listed conditions under which they expected their metadata to be used. Over half of these wanted metadata to be attributed to their DP, to continue to be freely available once disclosed, to remain unaltered and to be used for non-commercial purposes. These results were corroborated by the list of "unacceptable uses" the respondents came up with. One issue of concern was some DPs' desire that metadata should remain unaltered. Were this to be implemented, it would inhibit the function of Service Providers, many of whom need to enhance the metadata (e.g., provide subject indexing or authority control) in order to provide services.

10.6.2 Service Providers

Twice as many Service Providers disclosed both their own metadata *and* harvested others' data as those that only harvested others' data. This may have influenced their views on the rights status of metadata as they were not only end-users, but creators of such metadata. However, twice as many did not check the rights status of others' metadata before harvesting, compared to those that did. Half of those that did not check held the view that "Metadata is implicitly free in the OAI", and may have assumed that because they allow their own metadata to be freely harvested, that they had the same right to harvest others' data. It is a logical assumption. As one respondent wrote in the final comments section, "there are a multitude of methods for preventing indexing by Web services" for those that wish to prevent it. However, as we have shown, this is legally incorrect.

Despite their general view that metadata was implicitly free under the OAI, the majority of SPs (54.5%) said that they would only be happy for other SPs to harvest *their* enhanced metadata under certain conditions. Half of these stated that the condition was "with prior agreement", thus taking any automation out of the process. A slightly larger majority (63.6%) said that they would be happy for other SPs to then further enhance their enhanced metadata, again on certain conditions. None said they were happy for unconditional harvesting and/or further enhancing.

The two conditions of importance to SPs were i) attribution (perhaps through the OAI provenance schema), and ii) that freely available enhanced metadata remained freely available once harvested by another SP. These conditions were also stipulated by the DPs. However, many DPs also stipulated that metadata should be used for non-commercial purposes and that it should not be altered. As the business model of some SPs may depend on commercial viability, and on the need to enhance the metadata to provide a service, it is not surprising that such conditions did not appear on their list.

10.7 Ownership of rights in metadata

10.7.1 Data Providers

Assuming that both individual and collections of metadata qualify for either copyright or database right, in just over three-quarters of cases DPs are at least the joint rights owner, if not the sole owner, of those rights. However, in the five cases where authors created the metadata disclosed by the DP, the author is the rights holder. Of course, the rights owner has the power to decide how that metadata may be used by third-parties. As it is unlikely that authors will be interested in how their metadata is used by others (although they would certainly benefit from wide dissemination), DPs may wish to include a statement in their agreement with authors asking for a non-exclusive royalty-free licence to use the metadata in whichever ways they see fit.

10.7.2 Service Providers

The majority of SPs (75%) enhanced the metadata that they harvested. The important question is, do their enhancements merit copyright protection? As mentioned above, works of joint ownership are described as "a work produced by the collaboration of two or more authors in which the contribution of each author is not distinct from that of the other author or authors". Thus, arguably the enhancements made to a metadata record by the SP would qualify them for joint copyright ownership, because the contribution of one cataloguer is not distinct from that of the other.

However, as one of the original qualifications for copyright ownership is the demonstration of "sweat-of-brow" by the creator, 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. Despite this, seven of the eight SPs that enhanced metadata believed that their enhancements qualified for copyright protection – even though three of these were not interested in asserting those rights.

10.8 How to manage the distribution of rights between Data and Service Provider

It is difficult to know how to manage this situation. On the one hand, it would probably be simpler to assume that all enhancements qualify for copyright protection, and thus all metadata-enhancing SPs become joint copyright owner with the originating DP. However, DPs may be unhappy with this if the only "enhancement" the SP is making is simply to map the metadata to their local data model. On the other hand, it would be unfair to assume that no enhancements qualify for copyright protection, particularly if there has been financial investment on the part of the SP.

What would seem to be an equitable solution is that DPs retain copyright in their original records, but that SPs become joint copyright owners in the enhanced records. SPs may or may not wish to disclose those records for further harvesting. This will depend on whether the harvesting terms and conditions stipulated by the DP insisted that the SP continued to make the metadata "freely available" and whether freely available means for searching by end-users or for harvesting by other SPs. However, one of the conditions a DP may wish to stipulate is that as joint copyright owners in the enhanced metadata, they should receive a copy of the enhanced metadata for their own purposes.

10.9 Usefulness of standard metadata rights solution

Despite the fact that only two DPs had experienced unacceptable use of their metadata, 77.2% agreed that a standard way of describing how their metadata may be used would be helpful. They felt that such a solution should be simple, flexible, and machine-readable, and recognised that a generalised solution, although not satisfying everyone's needs, would certainly be a step in the right direction.

As with the DPs, the overwhelming majority of SPs thought that having a standardised way of describing the rights status of metadata would be useful. Only one respondent felt the developing of standardised metadata rights information went against the spirit of open access - a view held by many DPs, until they considered the potential for abuse of their metadata.

In discussions about the protection of material available on open-access, it should be remembered that just because something is publicly available does not mean that it is in the public domain. It is possible to assert reasonable conditions of use on publicly available works without hindering the spirit of open-access.

11 Conclusions

The general lack of awareness of IPR issues relating to the work of DPs was a cause for concern. The majority of SPs had given more consideration to the issues, but this may have been a natural caution stemming from the fact that they were building services on third-party data. Despite the evident need for awareness-raising activities, there is no obvious candidate to undertake such campaigns. There is a general feeling that IPR issues belong in the domain of closed-access, not open-access works. Thus, many open-access organisations do not wish to become embroiled in the debate. The Open Archives Initiative deliberately avoided addressing rights issues within the protocol itself for fear of alienating individual groups (Bird, 2001). Instead, by providing containers which may be used for rights information, they hope that communities will generate their own solutions. It may be, therefore, that such issues have to be addressed on a community-by-community basis. Alternatively, standards could be adopted, developed and/or promoted by organisations developing DP software such as eprints.org or DSpace.

It is certain that the IPR issues take on an additional layer of complication when considering the role of SPs in the harvesting, enhancing and further disclosing of DP's data. A metadata protection solution will need to build in a means for multiple copyright owners to be added as data is further harvested and enhanced. Utilising the provenance verb for this purpose as suggested by respondents would certainly seem to be a way forward. Whatever the final details are, it is clear that a standard way of describing the rights status of both documents and metadata would be welcomed by the OAI eprints community. The sixth study in the RoMEO series will describe the solutions in full (Gadd, Oppenheim, and Probets, 2003e).

References

Bass, M., et al (2002). DSpace: a sustainable solution for institution digital asset services: spanning the information asset value chain: ingest, manage, preserve, disseminate: internal reference specification: functionality. Cambridge MA, Hewlett Packard Company: 10. URL: http://dspace.org/technology/functionality.pdf

Bird, S. (2001). White paper on machine-readable rights information. Pennsylvania, University of Pennsylvania: http://www.ldc.upenn.edu/sb/oai/rights.html.

Caltech Library System. (2001) Author permission agreement URL: http://caltechlib.library.caltech.edu/archive/00000006/

Crow, R. (2002). SPARC Institutional Repository Checklist & Resource Guide. Washington, SPARC: http://www.arl.org/sparc/IR/IR_Guide.html.

DSpace. (2002). Non-Exclusive Distribution License URL: http://dspace.org/mit/policies/license.html

Dublin Core Metadata Initiative. (2003). URL: http://dublincore.org/

Eprints.org (2003) URL: http://www.eprints.org/

The FAIR Programme. (2002) URL: http://www.jisc.ac.uk/index.cfm?name=programme_fair

Gadd, Elizabeth., Charles Oppenheim and Steve Probets. (2003a) RoMEO Studies 2: How academics want to protect their open-access research papers. Journal of Information Science, [In Press].

Gadd, Elizabeth., Charles Oppenheim and Steve Probets. (2003b) RoMEO Studies 3: How academics expect to use open-access research papers. Submitted to Journal of Library and Information Science.

Gadd, Elizabeth., Charles Oppenheim and Steve Probets. (2003c) RoMEO Studies 1: The impact of copyright ownership on academic author self-archiving. Journal of Documentation. 59(3), 243-277

Gadd, Elizabeth., Charles Oppenheim and Steve Probets. (2003d) RoMEO Studies 4: The author-publisher bargain: an analysis of Journal Publishers Copyright Transfer Agreements. To be submitted to Learned Publishing.

Gadd, E., Charles Oppenheim and Steve Probets (2003e). "RoMEO Studies 6: Rights metadata for Open Archiving - the RoMEO Solutions." To be submitted to Program.

Giavarra, Emmanuella, and Charles Oppenheim. (1997) EU Database Directive: some clarifications. SCONUL Newsletter 12, 6-7.

Hunt, C. (2002). Archive User Survey. Electronics and Computer Science Department. Southampton, Southampton University: http://www.eprints.org/results/.

Liu, X., Kurt Maly and Mohammad Zubair (2001). "Arc - an OAI service provider for digital library federation." D-Lib Magazine 7(4): http://www.dlib.org/dlib/april01/liu/04liu.html.

McClelland, M., D. McArthur, et al. (2002). Challenges for service providers when importing metadata in digital libraries. D-Lib Magazine 8 (4) Apr 2002. URL: http://www.dlib.org/dlib/april02/mcclelland /04mcclelland.html.

Nixon, W. (2002). "The evolution of an institutional e-prints archive at the University of Glasgow." Ariadne 32 URL: http://www.ariadne.ac.uk/issue32/eprint-archives/

Nixon, W. (2003). "DAEDALUS: Freeing Scholarly Communication at the University of Glasgow." Ariadne 34 URL:http://www.ariadne.ac.uk /issue34/nixon/intro.html

Open Archives initiative. (2003a) URL: http://www.openarchives.org/

Open Archives initiative. (2003b) Registered Data Providers URL: http://www.openarchives.org/Register/BrowseSites.pl

Open Language Archives Community (2003). URL: http://www.language-archives.org/

Oppenheim, Charles, Heather Rosenblatt, Raymond Wall (Eds). (2001) The Aslib guide to copyright. London : Aslib Information Management.

Pinfield, S. (2001). "How Do Physicists Use an E-Print Archive? Implications for Institutional E-Print Services." D-Lib Magazine 7(12): http://www.dlib.org/dlib/december01/pinfield/12pinfield.html.

Pinfield, S., and John McColl (2002). "Climbing the scholarly publication mountain with SHERPA." Ariadne 33 URL: http://ariadne.ac.uk/issue33/sherpa.

Project RoMEO. (2003a). OAI Data Provider Survey URL: http://www-staff.lboro.ac.uk/~lbeag/Data-provider-questionnaire.htm

Project RoMEO. (2003b). OAI Service Provider Survey URL: http://www-staff.lboro.ac.uk/~lbeag/Service-provider-questionnaire.htm