Project RoMEO Final Report

Project Director: Prof. Charles Oppenheim
Technical Director: Dr Steve Probets
Research Associate: Elizabeth Gadd

1 Background to project

The JISC FAIR programme aimed to "evaluate and explore different mechanisms for the disclosure and sharing of content (and the related challenges) to fulfil the vision of a web of resources built by groups with a long term stake in the future of those resources, but made available to the whole community of learning." Many of the projects funded under the programme are exploring the establishment of Institutional Repositories (IR) of academic research output, using the Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH) to disclose and harvest metadata about those resources (often referred to as eprints). However, some of the main barriers to the success of such repositories are not technical, but legal and cultural. In particular, the IR model raises a wide range of IPR issues.

For example, if academics sign away their right to self-archive through journal publisher Copyright Transfer Agreements (CTAs), the whole process may collapse at the first hurdle. Once a paper has been self-archived, how can academics ensure that the rights they want asserting (say the right to be named as author, and to stop the text being altered) are asserted? Conversely, how can they ensure that other rights given them by copyright law that they may not care for (e.g. to prohibit copying) are waived? There are also rights issues for Data Providers (DPs) and Service Providers (SPs). DPs may be pleased that their metadata is being harvested, but what if someone then starts selling access to it? SPs may add value to the metadata they harvest, but who owns the rights in that enhanced metadata and how can it be protected?

The RoMEO Project was funded for one year (1 August 2002 – 31 July 2003) to investigate the IPR issues relating to the self-archiving of eprints by academic authors via Institutional Repositories. It aimed to develop some simple rights metadata by which such papers may be protected in an open-access

environment. It also aimed to investigate the issues relating to the IPR protection of metadata disclosed by Data Providers and harvested by Service Providers, with a view to developing a means by which the rights of such freely-available metadata might be protected under the OAI-PMH.

2   Methodology

The project was divided into two broad phases. Phase One collected data from relevant stakeholders (academic authors, journal publishers and OAI Data and Service Providers) on appropriate IPR issues which was fed into the development of the rights metadata and metadata protection solutions in Phase Two.

2.1   Phase One

2.1.1 Literature Review

The first activity of Phase One was a review of the literature in 5 main areas: copyright ownership and assignment issues from both the HE perspective and journal publishers' perspective; the required protection and usage of academic research papers; general digital rights management issues and emerging standards; rights issues already raised by the OAI and other open access communities; and the impact of relevant legislation. The literature review informed the creation of a map of rights issues which was later put out to the community for discussion and comment.

2.1.2 Stakeholder surveys

To gain data from the relevant stakeholder groups, four online surveys were performed. Each survey was mounted on the project web pages and advertised via a number of relevant email discussion lists.

a) Academic author survey

The academic author survey set about to determine exactly how academics wanted to protect their own freely available research papers and use others' freely available papers. The survey was based around a series of permissions, restrictions and conditions taken from the Open Digital Rights Language (ODRL). Academics could select those they wished to apply. More information on the survey methodology can be found in RoMEO Studies 1, 2 and 3.

b)Journal Publisher Survey

This survey solicited journal publishers' views on the self-archiving movement, as well as asking for views on the copyright status of publisher-generated metadata, and on the management of rights and permissions information.

c) OAI Data Providers Survey

This survey asked Data Providers about their agreements with authors; any copyright protection they afforded the resources contained within their

archives; and their views on the rights status of any metadata they create. More details on this survey can be found in RoMEO Studies 5.

d) OAI Service Providers Survey
The Service Providers survey asked about a range of rights issues they faced when harvesting, enhancing and disclosing metadata records. More details on this survey can be found in RoMEO Studies 5.

2.1.3 Journal Publisher Copyright Transfer Agreement (CTA) analysis
In addition to the data collected from journal publishers via the online survey, an analysis was performed on a large number journal publishers' copyright transfer agreements. The aim was to map the contractual relationship between publisher and author, particularly with regards to authors rights (e.g. to self-archive) and warranties under the agreements.

2.2   Phase Two
Phase Two of the project took the results of the author survey showing how academics wished to protect their open-access research papers and fed them in to the development of an appropriate rights metadata solution. It also took the results of the Data and Service Provider surveys and fed them into the development of the metadata protection solution.   Both development activities took into account emerging standards and issues raised by the literature in Phase One. They were also informed by ongoing consultation with the community through conferences, email discussion lists, and by direct consultation with relevant experts and Advisory Board members.

3   Activities

3.1 Phase One

3.1.1 Academic author survey findings
The academic author survey met with an excellent response with 542 authors completing the online questionnaire. Respondents were from 57 countries with the largest group (one-third) being based in the UK. The data allowed for comparisons between self-archiving and non-archiving respondents, (an almost equal number of each completed the survey)and between the protection required of academics-as-authors and academics-as-users of open-access research papers. These analyses are written up in RoMEO Studies 2 and 3. In summary, over 60% of academics were happy for their works to be displayed, given away, printed out, excerpted from, and saved freely. The majority wanted sales to be prohibited. The only restrictions required were that all copies should be exact replicas of the original work, and between 50-60% wanted them to be used for non-commercial purposes only. The only condition demanded was that the author(s) should always be attributed as such (see Figure 1).

Permissions	Restrictions	Conditions
Display	Exact replicas	Attribution
Give	For non-commercial purposes (optional)	
Print		
Excerpt		
Save		
Sell (prohibit)		

Figure 1: Protection required of academic open-access research papers

3.1.2 OAI Data and Service Provider Findings
Twenty-two DPs and thirteen SPs responded to their respective online surveys. The full results of the survey are documented in RoMEO Studies 5. However, in summary, the majority of DPs (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 (see Figure 2).
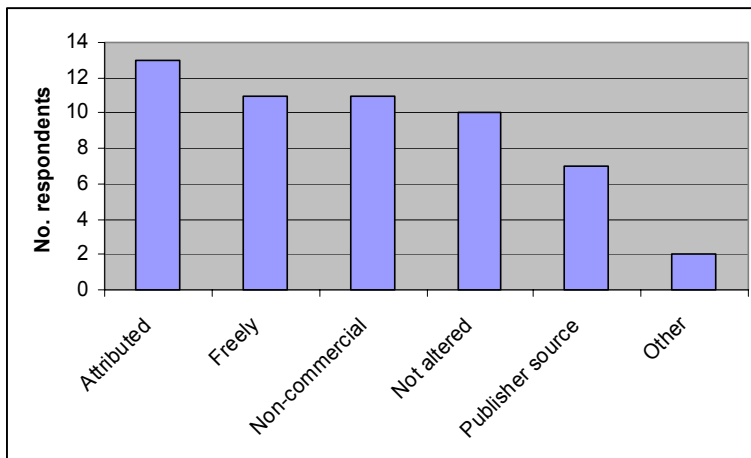

Figure 2 Conditions required by DPs on the use of metadata

Just three main conditions of use were listed by SPs: one was 'by prior agreement', another was attribution of the Provider, and the third was that subsequent harvesters disclose the metadata under the same conditions as it was harvested.

3.1.3 Journal publisher findings

Only 18 journal publishers responded to the questionnaire. A high-level analysis was performed on the data, in case there were strong correlations between the data, however none were to be found. As no conclusions could be drawn from such a small sample, it was felt that a detailed analysis would not be worth performing. The reason for the poor response rate was probably two-fold. Firstly, the survey was quite wide-ranging in its scope covering general opinions on self-archiving as well as detailed operational permissions matters. Thus the survey would probably have had to have been completed by a number of different staff within each publishing house. Publishers may have been more wary about giving opinions on behalf of their organisation than the other stakeholder groups. Secondly, the job titles of staff able to answer the questions would have been different at each publisher, thus where we had no named contact, the email advertising the questionnaire may have gone astray.

However, 80 journal publishers' CTAs were obtained, making for a fascinating analysis. The full analysis has been written up in RoMEO Studies 4. However, two of the key findings were that just under 50% of journal titles allowed authors to self-archive, but under widely varying conditions, and about 30% of agreements did not allow authors the right to do *anything* with their own research paper.

The results of the analysis were of such interest that the RoMEO Publisher Policy Directory was created showing which publishers support author self-archiving and under what conditions.

3.2 Phase Two
Developing a means of expressing the rights we had identified through metadata was not a straightforward process.

3.2.1 Dismissing Creative Commons
Being aware of the Creative Commons (CC) initiative from the outset, our first step was to compare authors' protection requirements with the licences offered by CC. However, we initially felt that the CC licences would not meet those requirements. Our main concerns were that:

* 67% of author respondents wanted to limit or prohibit aggregation, however all CC licences allow aggregation ('incorporating the work into another work')
* All CC Commons Deeds use the term 'Copy', however, our survey showed great confusion about the use of this term. Many seemed to think it meant 'plagiarise'.
* 55% of respondents wanted to limit use to 'certain purposes', one example being for educational purposes. The CC licences do not provide such a restriction.

To keep the RoMEO solutions simple, we wanted to adopt the same approach for both the rights metadata and the metadata protection solution. Thus, having dismissed CC for the rights metadata, we would need to look to our main alternative, namely, developing an application profile of an existing Digital Rights Expression Language (DREL) (See 3.2.2), not only for the rights metadata but for the metadata protection as well. It was felt that this would not be a compromise for the metadata protection solution, as there were only three main conditions required by D&SPs (see 3.1.2). It was felt that these conditions could be met in a single XML instance of rights metadata and would be fairly straightforward to implement through an existing DREL.

3.2.2 Choosing ODRL

The project had two choices when considering how best to describe rights and permissions over content. It could either have developed it's own language for doing so, or it could utilise an existing language. The benefits of using an existing language in terms of pursuing standardisation, saving time, and encouraging interoperability, were significant. A full description of the selection process is outlined in RoMEO Studies 6. In summary, the commercial XrML (Extensible Rights Markup Language) was dismissed in favour of the open source ODRL (Open Digital Rights Language).

An application profile of ODRL relating to the protection requirements of academic authors was therefore developed. For metadata protection, a single XML instance of ODRL describing the three necessary conditions was produced.

3.2.3 Returning to Creative Commons

Having developed the draft ODRL solution, a number of circumstances forced us to reconsider that choice. Firstly, although 57% of respondents wished to limit aggregation (and 10% wished to prohibit) it was felt that if the CC licences could provide the required limitations, then those academics should be satisfied. As the CC licences did provide most of the core limitations as specified by RoMEO academic author respondents, the importance of this discrepancy was diminished. Secondly, although the Commons Deeds used the term 'Copy', the term used in the CC 'legal code' documents was 'reproduce'. As the legal code would be considered the Master version of any CC licence, it was decided that this problem was less significant than originally thought.

In addition to these reconsiderations, there was an increasing interest in the CC licences from other members of the open access community including the OAI, the DCMI, and DSpace (who had decided to use the CC licences as a standard component of their open-source repository software). Finally, an analysis of a number of existing archives' metadata policy statements suggested that the "one-size-fits-all" approach proposed by RoMEO may actually result in "one-size-fitting-nobody". The analysis showed that whilst the

CC licences would not meet every archives' needs, they would provide more flexibility than the single ODRL XML instance approach.

It was generally agreed that whilst a bespoke RoMEO solution may accurately meet the perceived needs of a slightly wider cross-section of the community, if it led to a proliferation of different rights solutions then it would be likely to hinder rather than promote the communication of rights in an open access environment. It was therefore decided that it would be in the community's best interests to align RoMEO's work with what looked set to become an emerging standard: Creative Commons.

3.2.4 Expressing CC Licences through OAI-PMH
Having decided to work with CC licences, the next step was to consider how best to disclose the CC metadata under the OAI-PMH. The biggest problem here was that CC's metadata was expressed in RDF/XML and did not have an XML schema – something essential for metadata exposed under the PMH. Two options were proposed: 1) negotiate with CC regarding the development of an XML schema, or 2) write ODRL versions of the CC licences which could map to the ODRL XML schema. In fact, both options were pursued. CC kindly agreed to write an XML schema for their metadata, and we produced ODRL versions of the CC licences.

We proposed that the permissions status of both individual and collections of both metadata records and resources were described. Individual resources would be protected by the use of <dc:rights> as well as a separate rights metadata record, referred to in <dc:rights>. Individual metadata records would be protected using a record's optional <about> container. Whole collections of metadata and resources would be protected by the optional <description> response to the Identify verb. This would provide both a default statement on <metadataPolicy> describing the permissions status of metadata, and <dataPolicy> describing the permissions status of resources. The default data policy would in most cases have to be a simple copyright statement, unless the repository only accepts resources meeting a minimum set of CC licence terms.

Full details of the proposed solution may be found in RoMEO Studies Rights Solution document (see appendices).

3.2.5 Next steps: working with the OAI
During the development process, the OAI, CC and ODRL staff were regularly consulted. The OAI, in particular, have taken a keen interest in our work and recently proposed a collaboration between RoMEO and themselves in order to develop general guidelines for expressing rights under the OAI. These guidelines will be extensible, although in the short-term CC licences will be used as an exemplar. This will benefit the e-prints community as our research has shown that CC licences provide a 'good fit' for academic research papers.

The RoMEO Rights Solution document should form the basis for a White Paper which would be discussed by a specially-selected OAI Rights Technical Committee. Once refined through discussion, the paper will be put out to the community for comment and a final version should be completed by Spring 2004.

3.3 Advocacy and promotional activities
Sections 4.1 and 4.3 below outline the fruits of a wide range of project advocacy and promotional activities. In addition, project staff attended and contributed to a number of internal conferences and seminars including two Open Archive Forum events and the Paris Open Access to Scientific Information conference.   The RoMEO Project term was also extended by one month (1-31 August 2003) in order to undertaken additional advocacy, promotion and consultation work. During this time, the Research Associate advertised her availability for advocacy talks, and advocacy materials (leaflets and powerpoint presentations) were developed based on the findings of the project. The team also focussed on their work with the OAI Rights Technical Committee.

4   Outputs

4.1 RoMEO Studies Series
The work of the RoMEO Project has been written up in six peer-reviewed papers all of which either have or will soon be self-archived on the project's web pages at
http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html
Copies are attached as appendices to this report.

4.1.1 RoMEO Studies 1: The impact of copyright ownership on academic author self-archiving
This paper considers the claims for copyright ownership in research papers by universities, academics, and publishers by drawing on the literature, the author survey and the CTA analysis. It concludes that self-archiving is not best supported by copyright transfer to publishers even though some agreements allow self-archiving.

4.1.2 RoMEO Studies 2: How academics want to protect their open-access research papers

This study considers the protection for research papers afforded by UK copyright law, and by e-journal licences. It compares this with the protection required by academic authors for open-access research papers as discovered by the RoMEO academic author survey. It concludes that most academic authors are primarily interested in preserving their moral rights (right to be attributed as author and to prevent derogatory treatment), and that the

protection offered research papers by copyright law is way in excess of that required by most academics.

4.1.3 RoMEO Studies 3: How academics expect to use open-access research papers
This paper considers previous studies of the usage of electronic journal articles through a literature survey. It then reports on how academic authors expected to use open-access research papers and compares this data with how academics wished to protect their open-access research papers. It concludes that academics-as-authors are prepared to allow far more usage than academics-as-users have so far come to expect.

4.1.4 RoMEO Studies 4: An analysis of Journal publishers author agreements
This study reports in full on our analysis of 80 journal CTAs. It considers when rights are assigned; what rights are retained; author and publisher warranties; and what authors are permitted to do with their research papers once copyright has been assigned. It concludes that the community would benefit from a jointly-designed model author-publisher agreement that provides a balanced solution that meets the needs of both parties.

4.1.5 RoMEO Studies 5: IPR issues for OAI Data and Service Providers
This paper reports the results of the OAI D&SP surveys. 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. It shows that 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.

4.1.6 RoMEO Studies 6: Rights metadata for open-archiving
The final study reports on the rights metadata solution proposed by the RoMEO project which will afford protection to research papers made freely available on the web. It shows how such metadata can be disclosed and harvested under the OAI-PMH

4.2 RoMEO Publisher Policy Listing
The RoMEO Publisher self-archiving policy directory was not an output promised by the original proposal, but was written into the project plan once the project had commenced due to the perceived need. This has been extremely well received (see 4.6.2 below).

4.3 Conference papers, popular articles and advocacy talks
Members of the project team have given two peer-reviewed papers at international conferences on the work of RoMEO, as well speaking at three seminars aimed at information and publishing professionals, and two

advocacy events aimed at academics. In addition they have written three popular articles communicating their work to practitioners.

4.4 RoMEO Rights Solution
The project has produced a paper describing their proposed rights metadata and metadata protection solutions. It is expected that this will form the basis of the white paper that will inform the work of the OAI Rights Technical Committee outlined above.

4.5 Proposed OAI Rights Technical Committee
Should the above committee be formed, the resulting Guidelines for disclosing rights under the OAI-PMH will be of considerable benefit to the open access community and a real achievement for the project and JISC as a whole.

4.6 Impacts

4.6.1 General
There has been a great deal of interest in the RoMEO Project illustrated by numerous invitations to speak, write, and advertise the work of the project more widely on mailing lists. Virtually all the relevant conferences attended by project members during the project have made reference to the project's work. Some examples of feedback received on presentations and written work include:

- "just a quick email to tell you how impressed I was with your presentation. I thought it was very interesting, very well structured and - most of all - very clear, so even a non-copyright expert like myself could understand a lot."
- "I read your article "RoMEO Studies 1: The impact of copyright ownership on academic author self-archiving"(JDOC) as part of our research and was very impressed with it."
- "This survey is…a felt need. The paper is well written and no attempt is made to play to the gallery. [Often the open access and open source can lend themselves to be causes for crusades!]"

4.6.2 RoMEO Web pages & Publisher Policy Directory
Perhaps the aspect of RoMEO's work that has had the most impact has been the Publisher self-archiving policy directory. In the nine months since the Directory web page was mounted there have been 7,623 hits on the RoMEO home page, of which 3,363 (44.1 %) went on to view the Directory. Figure 3 demonstrates the number of hits over time.
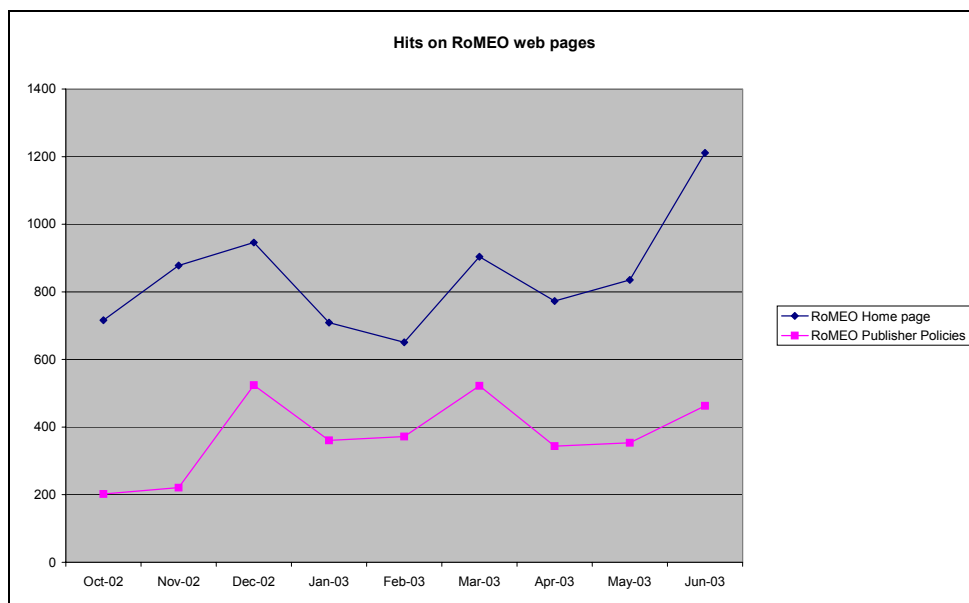
Hits on RoMEO web pages

Figure 3 Hits on the RoMEO home page and Publisher Policy page

We have received a number of comments about the web pages in general ("fascinating and useful") and the "very useful" Directory in particular. For example:

- "Just wanted to let you know how incredibly useful this work is. I'm just writing a paper for our Information Strategy Committee about e-print archives and I shall reference and incorporate this info. Keep up good work!
- "Thank you very much for your mail on OAI and for your valuable work your doing with ROMEO. We are facing exactly the same problem in Germany with our eDoc-Server of the Max Planck Society and are therefore very grateful that someone is conducting these surveys systematically."
- "I found the "copyright transfer agreement listing"…an excellent source for finding out whether we can archive a particular journal article."

4.6.3 Impact on the OAI community
At the time of writing the full RoMEO Rights Solution has not been launched in the community. This is because we do not want to pre-empt further discussions with the OAI-Rights Committee. However, the perceived importance or our work by the OAI, combined with our finding that the vast majority of D&SPs would welcome a standard solution to the problem of disclosing rights under the PMH, indicates that the work will have a great impact.

4.6.4 Impact on the Scholarly Communication System
It is too early to prove the impact of the project's findings on the current system of scholarly communication, however, the provision of empirical data to support a number of important premises should be of great benefit. For example: the fact that most academics are happy to allow far more liberal

usage of their research papers than either copyright law or e-journal licences allow. Also, the fact that academics-as-users are likely to make less liberal usage of such papers than academics-as-authors are prepared to allow. The findings of the Journal CTA analysis should also promote great discussion. Particularly, the fact that over 80% of journals do not allow authors to assert their moral rights; that just under 60% of journals are prepared to publish public domain US Government-owned works (showing they do not need copyright in order to publish); and that where employers own copyright they often get a better deal from CTAs than individual authors. See RoMEO Studies 1, 2, 3 and 4 for more details.

5   Future priorities for research area
Our research has found that there are many more issues that need addressing, and proposes further research in the following five areas:

5.1   The University to Academic relationship.
A co-ordinated effort is required to encourage Universities to assert their rights in research output to stop the copyright drain out of HE to publishers. The SURF collective of Dutch Universities has run two successful "Copyright and Universities" conferences (the Zwolle conferences) at which a wide range of stakeholder groups have agreed on a set of principles supporting the objective: "To assist stakeholders—including authors, publishers, librarians, and universities—to achieve maximum access to scholarship without compromising quality or academic freedom." Funding is needed to manage negotiations with stakeholders to develop and implement such policies, contracts and tools to effect a culture change in UK HE.

5.2   The Academic to University relationship.
Assuming that recommendation 5.1 above will not happen overnight, institutional repositories will still need to consider the rights and warranties they require of academics in order to mount and disseminate their work electronically. It is recommended that a model author/HEI licence is designed by which academics may licence IR's the appropriate rights. This may take the form of a click-thru licence asking authors, in simple terms, what the copyright status of their work is, and providing the IR with the security they need to 'publish'. It is also recommended that a set of standard "notice and takedown" procedures are developed by which illegally mounted materials may be removed without prejudice pending investigation of a complaint from a third-party. The development of standard licences and procedures such as these should greatly enhance the take-up of IR's, particularly if the results can be incorporated into the standard IR software packages such as eprints.org and DSpace.

5.3   The Author to Publisher relationship.
Our research has shown that 41% of authors "freely" assign copyright to publishers, and the majority of those that do not, are happy to sign exclusive licence agreements instead, many of which can be equally as restrictive as

copyright assignment forms. The RoMEO Project developed advocacy materials by which universities can encourage academics to retain their rights and self-archive legally. However, effecting a culture change on this scale will take a consolidated and centralised effort. It is recommended that a "Know your rights" campaign is established across all universities, that may feed into the Zwolle processes. This will inform academics,

     a) That under law, their employer is arguably the legal owner of copyright in their research outputs, although the employer may choose to ignore these rights;
     b) That exclusive licences are often equivalent to copyright assignment;
     c) Of the things they should check for in copyright assignment forms;

It is also recommended that a series of tools are developed to assist academics in their discussions with publishers. These may include a negotiation checklist and sample clauses by which academics may amend existing publisher licences. Academics also need to be informed about publishers' self-archiving policies. Project RoMEO has developed a list of such policies, but this will need to be maintained and updated. To be of maximum benefit, it should be developed to include other activities allowed or prohibited by publishers including the ability to use their work in PhD theses, to distribute to colleagues, students, etc,. This would give academics a fuller picture of the author-publisher bargain. Making the database searchable by journal title would also make it far more user-friendly.

5.4   The Publisher to Author relationship
In response to academics' increasing desire to retain copyright, the ALPSP have developed a Model Exclusive Licence by which academics can licence their rights rather than assign them. However, we have shown that exclusive licences can be equally as restrictive as copyright assignment. We have also shown that about one-third of such documents do not allow authors any rights at all to do anything with their own papers once copyright is assigned or licensed. It is recommended therefore, that representatives of UK HE join with representatives of publisher and other stakeholder groups, to develop a more author-friendly model licence, preferably on a non-exclusive basis. This sort of collaboration has been performed successfully before with the development of the JISC/PA Model Licence for electronic resources, and more recently through the PALS (Publisher and Library Solutions) COUNTER project that has developed a set of standardised usage statistics.

The work of RoMEO in understanding how academics wish to allow end-users to use their research papers should be used to persuade journal publishers to make their own end-user copyright policies more liberal. Our research showed that just under half of e-journal licence agreements prohibit the forwarding of a single electronic article to an end-user, something that publishers and libraries have agreed is actually permitted by law. It is recommended

therefore, that discussions with JISC negotiating teams and publishers are undertaken to consider these results.

5.5   *IPR in non-eprint media*
The work of the RoMEO Project focussed only on IPR issues relating to eprints, however, institutional repositories may contain a far wider range of materials, including images, learning objects, and theses. It is recommended that further research is performed to investigate the extensibility of the RoMEO rights metadata to other forms of media and document type.

5.6   *Conclusions*
Despite the numerous achievements of the RoMEO project in a very short time-frame, there is still much work to be done in the area of IPR and institutional repositories. The RoMEO Project team would be pleased to develop any of the enclosed recommendations for future study into a full proposal.

List of appendices
1) RoMEO Studies 1
2) RoMEO Studies 2
3) RoMEO Studies 3
4) RoMEO Studies 4
5) RoMEO Studies 5
6) RoMEO Studies 6
7) RoMEO Rights solution report