Self-archiving – the "right" thing? An introduction to the RoMEO Project
Elizabeth Gadd, Charles Oppenheim and Steve Probets

Introduction

In Higher Education today there are many calls for change in the scholarly communication process. A number of alternative publishing models are being advocated in the hope of counteracting ever-increasing journal price increases. One suggestion is that academics 'self-archive' their research papers either by making them available on their own web pages, or by submitting them to an institutional repository or a subject-based archive (e.g. ArXiv, CogPrints). It is argued that self-archiving will 'free' the research literature from expensive 'toll-gate' access, thus offering academics greater visibility and impact for their work.

The Scholarly Publishing and Academic Resources Coalition (SPARC) recently bought out a position paper making a case for the creation of institutional repositories[1]. They cite spiralling journal price increases, an increase in the overall volume of research, advances in digital publishing technologies, and uncertainty over who will handle preservation archiving, as four good reasons why institutional archiving of research output is a good idea.

The Open Archives Initiative (OAI)[2] has eased the passage of institutional repositories by developing a protocol by which information about resources such as research papers can be disclosed by Data Providers (in this case, institutional repositories) and harvested by Service Providers (such as cross-archive search engines). The Joint Information Systems Committee (JISC), inspired by the vision of the OAI, has recently funded a number of projects under its FAIR (Focus on Access to Institutional Resources) programme[3] with the hope of establishing a network of institutional repositories in the UK.

Essentially, the process is as follows:
1 An academic writes a research paper
2 The academic 'self-archives' a preprint of that paper on their institutional repository (or similar)
3 The academic submits the paper to a journal for peer reviewing
4 The journal accepts and publishes the paper
5 The academic archives the 'postprint' on the institutional repository
6 The repository (Data Provider) discloses information about the article
7 Service providers harvest that information
8 Users access the work (either the 'vanilla' version of that paper via the repository or the value-added version of the paper via a journal)

It may be that either one or the other of steps 2 and 5 take place – not both. So, an academic might just choose to archive a preprint of her paper and not archive the postprint (say it doesn't get accepted for publication). Alternatively, she may just choose to archive the postprint if there isn't a strong preprint culture in her discipline.

Seems straight forward? Unfortunately not. In fact, the whole process is encumbered with rights issues that could hamper the success of the movement. It is these issues that the RoMEO (Rights metadata for open archiving) project[4] has been funded by the JISC's FAIR programme to identify and address.

Rights issues relating to self-archiving

1 – Academic writes a research paper
The success of the self-archiving movement will depend to a great extent on the will of the academic and her relationship with her institution regarding the copyright in her journal articles. Who owns the copyright in journal articles written by academics? Is it the academics themselves, their institution, or an outside sponsor? How is that decided? Is it governed by contract, law or custom and practice?[5] Most importantly of all, does the academic know who owns copyright? If there is third-party material in the paper the copyright issue gets even more complicated, as it does if the article is written by more than one author. One of the first tasks of the RoMEO project is to survey academic authors on these and related issues. The relationship between the academic and the Higher Education Institution will also be examined.

2 – The academic 'self-archives' a preprint of that paper on their institutional repository
Institutional repositories are a relatively new thing in the UK. A handful of institutions have established fledgling services and a larger number are part of JISC-funded FAIR projects which will support the development of repositories. One of the problems identified by existing repositories is that of assessing whether the academic has the right to self-archive a particular paper. The fear is that many academics sign journal publishers' copyright assignment forms without reading them, and will therefore be ignorant of the rights status of their own work. Academics may have to warrant to the repository that they have the right to self-archive. The danger is that the academic will take the same laissez-faire attitude to signing that warrant as they did signing the original copyright assignment form.

Once the paper is freely available on the web, other issues arise. How might that work be legally used? How does the author want the work to be protected? It may be that the author would be happy with less protection than copyright law allows, but still wants some protection, so shouldn't simply waive all rights. Can the author specify the copyright status of their work? If so how? Could the author use an existing 'alternative' copyright regime (e.g. copyleft[6])? One of the specific tasks of the RoMEO project is to develop a simple set of rights metadata elements that authors, Data Providers and Service Providers may use to specify the rights status of a work to an end user.

3 – Academic submits paper to journal for peer review
The big question here is whether a journal will review a paper (let alone accept it) if it has been self-archived. Many journals ask authors to sign copyright assignment forms prior to peer review that warrant that the paper has not been previously published. Therein begin all the arguments as to what constitutes "publication". (Incidentally, some of those forms do not return copyright to the author if the paper is rejected). To address some of these issues, the project will be performing a survey of publishers' attitudes towards self-archiving, and an analysis of their current copyright assignment forms.

Of course there is a bigger issue here, and that is the influence of the UK Research Assessment Exercise (RAE) on the scholarly communication process. The story is an

old one: every four years the RAE rates the quality of a University's research. One measure of quality (though one that is not used by RAE panels) is the importance of the journals (often measured by impact factor) in which research is published. So important is it to be published in a prestigious refereed journal that many academics will sign any copyright assignment form put in front of them in order to get that publication. Thus, the long-term benefits of open-access research literature come second to the importance of a high RAE rating and the personal kudos of being published in a "high-impact" journal.

4 – The journal accepts the paper
Once a journal accepts a paper, those publishers that haven't already asked for copyright assignation will do so. Some publishers are now offering authors an alternative in the form of an exclusive licence. However, an exclusive licence may be just as restrictive as copyright assignment, so authors need to read them carefully. How will the academic respond to a restrictive licence? Some may amend the licence and return it, some institutions may provide their staff with preset clauses or alternative licences to present to the publisher. Others may choose to adapt an existing model licence such as the one provided by the ALPSP. If a publisher does allow author self-archiving, it may be on certain conditions. If so, what are they and do they check if authors comply?

5 The academic archives the postprint on the web
For the academic to self-archive their refereed postprint, they must have retained the right to do so. If the publisher only allows self-archiving of *pre*prints, the Harnad-Oppenheim copyright proposal[7] may be brought into play. This proposes that a preprint, plus a list of amendments or errata based on reviewers comments, will not infringe the copyright in the postprint. Assuming the publisher does allow self-archiving of the refereed postprint, another question arises. Are academics allowed to self-archive the publisher pdf (provided by many publishers to their authors)? Some academics believe they have the right to do so, but it may not be the case.

By this point, there may be a number of different versions of the paper available: the unrefereed preprint, the refereed postprint in plain text or HTML on an institutional repository, the refereed and typeset version in the printed journal, and a series of electronic versions (realpage, catchword, pdf) provided by the publisher's electronic journal. There are important version labelling issues to be resolved here, not least because each version may have a different copyright status.

6 The Data Provider discloses information about the article
If you mention 'rights issues' to OAI Data Providers, they will automatically assume you are talking about the copyright status of their metadata, not the rights of the resources themselves. This is a key issue for Data Providers particularly if they are investing considerable resources creating metadata. There have already been cases where commercial services have harvested freely available metadata and incorporated it into "toll-access" services[8]. The questions are legion: Are there any rights in an individual metadata record? If so, who owns them? Does it depend on the intellectual effort gone into the metadata creation? If so, how is such effort measured? Do data providers wish to assert any rights over either individual metadata records, or whole data collections? If so, what do they want to protect, and how

might this be done? RoMEO's survey of OAI Data Providers should provide the answers to some of these questions.

7 Service providers harvest that information

Service Providers are at the other end of the metadata rights status questions. How do they ascertain the rights status of the metadata they're harvesting? What happens if Service Providers add any value to the metadata? It may be that one metadata record has multiple copyright owners – just like a multimedia work. The other questions Service Providers may have to ask themselves in the future is how they should make use of any rights or permissions information relating to the documents themselves? Do they have an obligation to provide that to the end user? If so, how? Again, the RoMEO project's survey of Service Providers should provide some of the answers.

8 - Users access the work

At the end of the chain, there is an end-user (often another academic) who wants to make use of the research paper. As mentioned before, there may be a number of different versions of the work available. How can they ensure they have access to version with the least copyright restrictions? How do they know what use they can make of the paper (print, forward to colleagues, create a derivative work)? If an 'alternative' copyright regime is used (e.g. copyleft) will they know what that means? If the rights status is displayed to them, will they read it, or should access be controlled via 'technological measures' to use EU Directive terminology[9]?

Conclusions

The aim of the RoMEO project is to ensure that copyright issues do not hinder the development of author self-archiving via institutional repositories. It hopes to do this by assessing the key rights issues for each stakeholder group through a series of surveys, and by making recommendations that address those issues. The project team would be grateful if readers would advertise the surveys (available from http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html) as widely as possible to relevant parties in their institution.

One important deliverable of the project is the development of a series of simple rights metadata elements that authors may assign to their works. It is hoped that this will have a number of positive effects. Authors may be more willing to self-archive if they are reassured that they can protect the rights in their work that are important to them. At the same time, a system that is less restrictive than traditional copyright law will promote the free sharing of research and improve an authors impact and visibility – precisely the motivation of the open access movement itself.

The outcomes of the one-year project will be reported in the professional and scholarly literature. Self-archived versions of all research outcomes will, of course, be made available.

References

[1] Crow, R. (2002). The case for institutional repositories: a SPARC position paper. Washington DC, The Scholarly Publishing and Academic Resources Coalition. http://www.arl.org/sparc/IR/ir.html
[2] The Open Archives Initiative. URL: http://www.openarchives.org/ Accessed 7 October 2002.

[3] The FAIR programme. http://www.jisc.ac.uk/dner/development/programmes/fair.html Accessed 7 October 2002.

[4] The RoMEO Project. URL: http://www.lboro.ac.uk/departments/ls/disresearch/romeo/index.html Accessed 7 October 2002.

[5] Oppenheim, C. (1996). Copyright in HEIs: a discussion paper. Strathclyde, Department of Information Science. URL: http://www.jisc.ac.uk/pub/copyright/charles1.htm Accessed 15 October. See also: Weedon, R. (2000). Policy approaches to copyright in HEIs: a study for the JISC Committee for Awareness, Liaison and Training (JCALT). Glasgow, The Centre for Educational Systems.

[6] What is copyleft?, Free Software Foundation Inc., 2002. URL: http://www.gnu.org/copyleft/copyleft.html Accessed 11 October 2002.

[7] For the Harnad/Oppenheim self-archiving copyright strategy see section 6.1 of: Harnad, S. (2001) For whom the gate tolls: How and Why to Free the Refereed Research Literature Online Through Author/Institution Self-Archiving, Now. URL: http://www.ecs.soton.ac.uk/~harnad/Tp/resolution.htm#Harnad/Oppenheim Accessed 14 October 2002.

[8] Krichel, T. (2002). Re: OAI and intellectual property issues. Email to discussion list: OAI-general@oaisrv.nsdl.cornell.edu.

[9] (2001). "DIRECTIVE 2001/29/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society." Official Journal of the European Communities **L**(167): 10-19.