

This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

The RoMEO Project: protecting metadata in an open access environment

PLEASE CITE THE PUBLISHED VERSION

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Gadd, Elizabeth A., Steve Proberts, and Charles Oppenheim. 2019. "The Romeo Project: Protecting Metadata in an Open Access Environment". figshare. <https://hdl.handle.net/2134/278>.

[This article has been submitted to Loughborough University's Institutional Repository by the author.](#)

The RoMEO Project: Protecting metadata in an open access environment

Elizabeth Gadd, Charles Oppenheim and Steve Proberts

OAI-PMH

The Open Archives Initiative's Protocol for Metadata Harvesting (OAI-PMH) enables the 'disclosure' of metadata by Data Providers and the harvesting of that metadata by Service Providers. Although there is nothing to stop commercial providers from utilising this open-source protocol [¹], it has its roots in the open access community and as such is used by many open archives. These include subject-based archives such as ArXiv [²], CogPrints [³], and the increasing number of Institutional Repositories, many of which have been established as a result of funding via the UK JISC FAIR (Focus on Access to Institutional Repositories) programme [⁴].

The RoMEO Project

The RoMEO Project (Rights Metadata for Open archiving) [⁵] was also funded under the FAIR programme. It is investigating all the IPR issues relating to the self-archiving of research papers via institutional repositories. One key issue is how best to protect such research papers, and the metadata describing those papers, once in an open access environment. The investigations have taken the form of online surveys of academic authors [^{6, 7}], journal publishers, Data Providers and Service Providers [⁸], as well as an interesting analysis of 80 journal publishers' copyright transfer agreements [⁹]. There were two principal aims of the data gathered through these surveys. The first was to inform the development of some simple rights metadata by which academics could protect their open access research papers. The second was to inform the creation of a means of protecting all the freely available metadata that will soon be circulating as the OAI-PMH is more widely adopted.

The development of the rights metadata solution will be documented fully in the sixth and final study in the RoMEO Studies Series [¹⁰]. This paper concentrates on the second aim: how to protect freely available metadata disclosed and harvested under the OAI-PMH.

Survey of Data and Service Providers

A full report on the online surveys of Data and Service Providers has been written up in RoMEO Studies 5 [9]. However, in summary, responses were received from 22 Data Providers (DPs) and 13 Service Providers (SPs) and some interesting discoveries

were made with regards to the protection required by Data and Service Providers over their metadata.

Are there rights in metadata and if so, who owns them?

Perhaps the first question requiring an answer when considering protecting the rights in individual and collections of metadata is this: do such rights exist and if so, who owns them? This issue is debated fully in RoMEO Studies 5, but we conclude that individual metadata records probably qualify for copyright protection, the owner of which would be either the record's creator, or the employer of that creator.

Collections of metadata records would certainly qualify for Database Right in the EU, the owner of which would be 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 [¹¹]" or their employer if employed to create the database.

Rights owned by Data Providers

Assuming that both individual and collections of metadata qualify for either copyright or database right, we found that 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 five cases where the authors alone created the metadata disclosed by the DP, the author would be the sole 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.

Rights owned by Service Providers

The majority of SPs (75%) enhanced the metadata that they harvested. The important question is, do their enhancements merit copyright protection? The UK Copyright Designs and Patents Act describes works of joint ownership 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 [¹²]". 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" effort by the creator, it would seem logical that the enhancements would also need to demonstrate such effort in order to qualify. Thus, enhancements such as normalising field values and adding in domain addresses to URLs that lack them may not involve sufficient effort to qualify the resulting enhancements for copyright protection, but subject classification and the addition of name authority might.

Does open-access metadata need protecting?

Views of Data Providers

It was clear that most DP's had not really thought about whether their open-access metadata needed protecting. 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 they felt this right 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.

Views of 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, but legally incorrect.

How should it be protected?

Data Providers

Although the majority of DPs did not initially see the point of open access metadata protection, a subsequent question about the acceptable use of metadata appeared to raise awareness amongst DPs as to the benefits of it. Indeed, 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 DP’s 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.

Service Providers

Again, 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 through the OAI provenance schema [¹³], 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.

Would a standard means of protection be useful?

Only two DPs had experienced unacceptable use of their metadata; nonetheless, 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 also 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 initially held by many DPs, until they considered the potential for abuse of their metadata.

Creative Commons

One initiative aiming to support open access by providing a 'public domain plus' level of copyright protection (that is, more protection than donating a work to the public domain, but less restrictive protection than that provided by copyright law) is Creative Commons (CC) [¹⁴]. CC has designed a series of licences by which creators may make their works available on open access whilst retaining some measure of control over them. The licences allow display, public performance, reproduction, and distribution of a work whilst providing creators with four optional restrictions: attribution, non-commercial use, no derivative works, or permitting derivative works under a "sharealike" condition (meaning that subsequent works have to be made available under the same terms as the original.) Creators select the restrictions they wish to apply. In total there are a possible 11 alternative licences.

<Insert all images from <http://www.creativecommons.org/learn/licenses/> here.

Caption: Figure 1: Creative Commons Licence Options>

Each licence consists of a brief "human-readable" statement called the Commons Deed to communicate the terms quickly to end-users; a full licence document describing the conditions in legal code; and some machine-readable rights metadata in RDF/XML.

Despite the fact that the ROMEO Project proposal specified the development of a 'perfect fit' rights metadata solution for protecting academic research papers using ODRL (Open Digital Rights Language) [¹⁵], the increasing momentum of the Creative Commons Initiative led us to consider it as an alternative rights metadata option. We concluded that the CC solution provided a good enough fit with our survey findings to meet the majority of the needs of academic authors' open access papers, and that the ongoing support CC would provide beyond the life of the RoMEO Project (which ends in September 2003) would be an additional benefit. When it came to considering the *metadata protection* solution, CC also seemed to meet the requirements as laid down by the Data and Service Provider surveys, namely, the attribution requirement, prohibiting non-commercial uses and either allowing derivative works under a 'sharealike' condition, or prohibiting derivative works completely.

Using CC to protect metadata under the OAI-PMH

We then set about considering how best to disclose CC rights information under the OAI-PMH.

Individual metadata records

The OAI-PMH specification allows each metadata record to have an optional <about> container. One of the suggested purposes for this container is to hold rights information about the metadata record itself. Although the protocol does not suggest how this might be done, it does state that the contents of all <about> containers must conform to an XML schema. A problem initially arose here in that although CC provides machine-readable rights metadata as part of the licence 'package', that metadata is supplied in RDF/XML which as yet does not have an XML schema. Fortunately, negotiations with CC have proved fruitful and they have kindly agreed to write such an XML schema for their RDF. This work should be completed by September 2003 and will then be published. The alternative proposed by us, to create ODRL versions of the 11 CC licences, taking the form of XML instances which would conform to the pre-existing ODRL XML schema has therefore not been pursued

Metadata collections

Describing the rights status of an entire collection of metadata records depends on all records adhering to a single rights statement. This is fairly straightforward if all records are owned by a single rights holder (e.g. the Data Provider), however, it is not so straightforward if a number of rights holders are involved (e.g. authors). In response to the OAI-PMH Identify verb, DPs may optionally provide a <description> of their repository. Again, the contents of such a container must conform to an XML schema. A schema has been written to describe the contents of an eprints repository (*XML Schema to describe content and policies of repositories in the e-print community* [¹⁶]), which allows for an optional <metadataPolicy> element. This element may in turn contain <text> and/or <URI> elements. We propose that the <text> element contains a statement to the effect that all the records within this (named) repository adhere to the chosen CC Licence, and that the <URI> element contains either the URI of the appropriate CC Commons Deed (which in turn links to the legal code), or the URI of a generic RDF/XML instance of the chosen licence.

Future work

One issue, not addressed by the RoMEO proposal is how to manage a joint rights ownership situation between Data and Service Provider. Thus, if a Data Provider allows Service Providers to harvest, enhance and re-disclose their metadata under a *different licence* to their original one, how will the joint copyright ownership arrangement be specified? The situation could become increasingly complicated as more Service Providers harvest, enhance and re-disclose the metadata. Once the proposals resulting from our research are put into circulation, it may well be that other issues arise. However, what is important at this stage is that some means of protection is made available for others to debate and build upon.

In this vein, the RoMEO Project is currently in negotiations with the OAI with regards to developing a general specification for disclosing rights information (about both metadata and resources) under the PMH. Should this go ahead, it is hoped that the specification would be complete by February 2004.

Acknowledgement

The RoMEO Project should like to thank the UK Joint Information Systems Committee for funding this research and development project.

References

- ¹ Anon. "The Institute of Physics adopts OAI technical framework." Institute of Physics News 2002(25 November 2002) URL: <http://www.iop.org/Physics/News/0467j>.
- ² ArXiv. URL: <http://www.arXiv.org/>
- ³ CogPrints. URL: <http://cogprints.ecs.soton.ac.uk/>
- ⁴ JISC. Focus on Access to Institutional Resources (FAIR) Programme. URL: http://www.jisc.ac.uk/index.cfm?name=programme_fair
- ⁵ RoMEO Project. URL: <http://www.lboro.ac.uk/departments/lis/disresearch/romeo/index.html>
- ⁶ Gadd, E., Charles Oppenheim and Steve Proberts (2003). "RoMEO Studies 3: How academics expect to use open-access research papers." Journal of Librarianship and Information Science 35(3): In Press.
- ⁷ Gadd, E., Charles Oppenheim and Steve Proberts (2003). "RoMEO Studies 2: How academics want to protect their open-access research papers." Journal of Information Science 29(5): [In Press].
- ⁸ Gadd, E., Charles Oppenheim and Steve Proberts (2003). "RoMEO Studies 5: The IPR issues facing OAI Data and Service Providers." Submitted to Journal of Information Law and Technology.
- ⁹ Gadd, E., Charles Oppenheim and Steve Proberts (2003). "RoMEO Studies 4: The author-publisher bargain: an analysis of journal publisher copyright transfer agreements." Submitted to Learned Publishing.
- ¹⁰ Gadd, E., Charles Oppenheim and Steve Proberts (2003). "RoMEO Studies 6: Rights metadata for Open Archiving - the RoMEO Solutions." To be submitted to Program.
- ¹¹ DIRECTIVE 96/9/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 11 March 1996 on the legal protection of databases
- ¹² Great Britain. (1988). Copyright, designs and patents Act 1988, Ch. 1. 10(1). Elizabeth II. London : HMSO
- ¹³ Open Archives Initiative. "XML schema to hold provenance information in the "about" part of a record". URL: <http://www.openarchives.org/OAI/2.0/guidelines-provenance.htm>
- ¹⁴ Creative Commons. URL: <http://www.creativecommons.org/>
- ¹⁵ Open Digital Rights Language Initiative. URL: <http://odrl.net/>
- ¹⁶ Open Archives Initiative. "XML Schema to describe content and policies of repositories in the e-print community". URL: <http://www.openarchives.org/OAI/2.0/guidelines-eprints.htm>