

RoMEO Studies 8: Self-archiving – when Yellow and Blue make Green: the logic behind the colour-coding used in the Copyright Knowledge Bank

Celia Jenkins, Steve Proberts and Charles Oppenheim, B. Hubbard

Authors: Celia Jenkins, Researcher, Steve Proberts, Lecturer and Charles Oppenheim, Professor of Information Science at the Department of Information Science, Loughborough University, UK. E-mail: C.D.Jenkins@lboro.ac.uk, S.G.Proberts@lboro.ac.uk, and C.Oppenheim@lboro.ac.uk

Bill Hubbard, SHERPA Manager, Greenfield Medical Library, University of Nottingham, UK. E-mail: Bill.Hubbard@nottingham.ac.uk

Abstract

Purpose – The purpose of this research is to show how the self-archiving of journal papers is a major step towards providing open access to research. However, copyright transfer agreements (CTAs) that are signed by an author prior to publication often indicate whether, and in what form, self-archiving is allowed. The SHERPA/RoMEO database enables easy access to publishers' policies in this area and uses a colour-coding scheme to classify publishers according to their self-archiving status. The database is currently being redeveloped and renamed the Copyright Knowledge Bank. However, it will still assign a colour to individual publishers indicating whether pre-prints can be self-archived (yellow), post-prints can be self-archived (blue), both pre-print and post-print can be archived (green) or neither (white). The nature of CTAs means that these decisions are rarely as straightforward as they may seem, and this paper describes the thinking and considerations that were used in assigning these colours in the light of the underlying principles and definitions of open access.

Design/methodology/approach – Detailed analysis of a large number of CTAs led to the development of controlled vocabulary of terms which was carefully analysed to determine how these terms equate to the definition and “spirit” of open access.

Findings – The paper reports on how conditions outlined by publishers in their CTAs, such as how or where a paper can be self-archived, affect the assignment of a self-archiving colour to the publisher.

Originality/value – The colour assignment is widely used by authors and repository administrators in determining whether academic papers can be self-archived. This paper provides a starting-point for further discussion and development of publisher classification in the open access environment.

Keywords

Self-archiving, Publishers, Copyright Knowledge Bank, Institutional Repositories, SHERPA/RoMEO

Word length: 4393

1. Introduction

The Copyright Knowledge Bank database (CKB) was created as part of the programme 'Partnering on Copyright' jointly funded by the UK's Joint Information Systems' Committee (JISC) and SURF - the higher education and research partnership organisation for network services and ICT in the Netherlands. The CKB database is a development of information held by the existing SHERPA/RoMEO database of journal publishers' self-archiving policies (<http://www.sherpa.ac.uk/romeo.php>). The SHERPA (originally standing for Securing a Hybrid Environment for Research Preservation and Access) and its RoMEO (Rights Metadata for Open Archiving) database provides academic authors, institutional repository (IR) administrators and publishers the ability to check the conditions and restrictions that publishers place on self-archived scholarly articles (see Gadd *et al.*, 2004). Such conditions are specified in publishers' copyright transfer agreements (CTAs) – these are the legal documents that are signed by authors assigning copyright or a copyright licence in the work to the publisher. These agreements often grant some rights back to the author, which may enable them to self-archive the work in some way. The new CKB information offers improved coverage and functionality providing more comprehensive information on whether, and under what circumstances, self-archiving is allowed.

As part of its user interface, and in order to highlight the self-archiving status of a publisher, i.e., to what degree a publisher permits self-archiving, the CKB provides a classification system, assigning self-archiving colours of 'green', 'blue', 'yellow' and 'white' to publishers dependent on their self-archiving policy. In order to assign these colours, a logic needs to be applied to the conditions and restrictions specified in the publishers' CTAs. Details of how these publisher policies have been reduced to terms in a controlled vocabulary can be found in RoMEO Studies 7 (Jenkins *et al.*, 2006). This paper describes this logic and how it is used to assign a self-archiving category (colour) to each publisher.

1.1 The self-archiving colours

Colours are used to categorise publishers according to the degree to which they allow self-archiving. This provides a simple way of identifying the general self-archiving policy of a particular publisher.

The system of self-archiving colours white, yellow and green was developed during the original RoMEO Project (Gadd *et al.*, 2004) and is heavily used in open access circles. It was developed for use in the SHERPA/RoMEO database with the addition of 'blue'. Publishers are assigned a particular self-archiving colour, depending on what they allow to be self-archived.

The self-archiving colours are:

- white (neither pre-prints nor post-prints can be self-archived);
- yellow (only pre-prints can be self-archived);
- blue (only post-prints can be self-archived);
- green (both pre-prints and post-prints can be self-archived).

Another colour is 'gold', although this is not used in the CKB as it is independent of self-archiving rights, relating to publishers which publish Open Access journals, i.e., journals that are always free at the point of use and for which no subscription fee is payable (Brody *et al.*, 2004).

For each publisher entry in the CKB, details are maintained on whether a type of work can be self-archived and, if so, under what circumstances it is allowed. It is this information that is used to generate the self-archiving colour of the publisher. The colour is then prominently displayed in the CKB alongside that particular publisher's self-archiving details.

In order to explain how the self-archiving colours are generated for the CKB, it is necessary to look in more detail at the aims of the CKB and its composition.

2. Aim of the Copyright Knowledge Bank

The CKB builds on the SHERPA/RoMEO database, by providing improved coverage and functionality. It was important that at the start of the CKB's development, its aims were clearly identified and defined, as these would have a major effect on how the CTAs would be analysed and represented in the CKB. These aims were identified as being:

- To report on what publishers' CTAs allow authors to do with their Work, once they enter into a relationship with the publisher. The focus is firmly on describing the self-archiving policies of publishers.
- To act as a tool to provide academic authors and librarians/IR managers with the information on which publishers allow self-archiving, and under what circumstances. To this end, the self-archiving colours are algorithmically generated based on fields maintained within the database.

It was therefore important to provide a clear and comprehensive way of describing journal publishers' self-archiving policies described in CTAs and supporting documentation. To this end, self-archiving terms and definitions were identified and put into categories, and a resulting controlled vocabulary developed. This was used to systematically analyse journal publishers' CTAs (Jenkins *et al.*, 2006).

3. Definition of self-archiving

As the concept of self-archiving is central to the CKB, it was vital to be clear as to what it actually means. There are various definitions given to self-archiving, so it was felt important to identify one which could then be used as a benchmark throughout the analyses of self-archiving policies. The decision was made to adopt the definition of Open Access provided in the Bethesda Statement on Open Access Publishing, which states that the authors retain the right to reuse their work without any restrictions, and so can self-archive their work straight away in open access online repositories. The Bethesda definition states:

“A complete version of the work and all supplementary materials... is deposited immediately upon initial publication in at least one online repository that is supported by an academic institution, scholarly society, government agency, or other well-established organization that seeks to enable open access, unrestricted distribution, interoperability, and long-term archiving.”
(Suber, 2003)

Self-archiving, therefore, depends on the ability to archive within independent non-commercial archives. This has an impact on the acceptability of archiving within, for example, a publisher's archive which has no guarantee of longevity, permanent free and open access and independence.

Although the CKB is primarily aimed at academic authors and IR administrators, the views of publishers must also be considered. The CKB is required to provide comprehensive, clear data on publishers' self-archiving policies. This involves simplification and a degree of interpretation of CTAs. However, any interpretation must also fairly represent publishers' self-archiving policies, and the analysis must take into account what actually appears in the CTAs, what publishers 'mean' by their statements and why publishers insist on these statements. Therefore consideration has to be given to how the CTAs should be analysed and the information presented to the public, with an attempt to balance the needs and views of both groups.

One very important example of this is the assignment of self-archiving colours, as this is the most prominent and easily identifiable way of categorising publishers. Publishers, as well as authors, take notice of the self-archiving colour that is assigned to them or their counterparts. It is therefore vital that considered decisions are made on how to categorise publisher's self-archiving policy, especially as Open Access and self-archiving are at such an important stage of development.

4. The controlled vocabulary of the CKB

The CKB is based on the original SHERPA/ROMEO database, and uses the same three basic categories to analyse CTAs. These are:

- 'What' can be self-archived
- 'Conditions' of self-archiving, including 'where' it can be self-archived
- 'Restrictions' of self-archiving, including 'when' it can be self-archived

4.1 'What can be self-archived'

The CKB contains information on three different 'types' of work:

- *Pre-print*. A pre-print has been defined by the CKB to be "The primary, draft version/form of the Work, *up to* and *during* the peer review process". In other words, a pre-print is any version of the work which is still being revised prior to, or during, the peer review process.
- *Post-print – author version*. A post-print has been defined by the CKB to be "The definitive version/form of the Work, *after* peer review, which has been accepted for publication, for which copyright has been assigned or a licence agreement has been signed". The 'author' version of a post-print is one which is produced by the author, with all peer-review comments and revisions integrated into the text, but which has not been typeset or edited by the journal publisher.
- *Post-print – publisher version*. The publisher version of the post-print is similarly defined by the CKB as one which has been peer reviewed, has been accepted for publication, and for which copyright has been assigned or a licence agreement has been signed. However, the publisher version of a post-print is one which has the publisher's copy-editing and formatting are in place, i.e., a publisher-produced PDF falls under this category.

When assigning a self-archiving colour, each of these types is considered, along with any conditions or restrictions that are attached to these.

4.2 ‘Conditions’ of self- archiving

The definition of a condition is that it is a requirement which publishers ask is met in order to self-archive, but which does *not prevent* an author from self-archiving their work immediately.

Examples of conditions are:

- “An electronic link must be provided to the journal’s homepage/Web site”.
- “The copyright holder of the Work must be acknowledged”.
- “The archived Work must not be used for any systematic external distribution by a third party”.

By carrying out the conditions, as specified by the publisher, authors are permitted to self-archive their work. Conditions therefore do not affect the colour categorisation of a publisher. For instance, a publisher indicating that a pre-print can be self-archived as long as it is linked to the publisher’s homepage would still be assigned a ‘yellow’ colour. If the same publisher allowed a post-print to be self-archived with a similar condition, then that publisher would be assigned a ‘blue’ colour. Publishers assigned both ‘yellow’ and ‘blue’ colours are re-designated as ‘green’ (yellow and blue make green). Anything considered a ‘condition’ therefore does not preclude publishers from being assigned a ‘yellow’, ‘blue’, or ‘green’ colour.

4.3 ‘Restrictions’ when self archiving

The definition of a restriction is that it is a requirement which publishers ask is met in order to self-archive, but which *prevents* an author from self-archiving *immediately*.

Examples of restrictions are:

- “Formal permission from the publisher must be sought if the Work is to be posted electronically”.
- “The previous version of the Work must be replaced with its abstract and full citation”.
- “The Work can only be self-archived 6 months after its publication in the journal”.

If a restriction(s) is specified by the publisher then self-archiving of the work is not allowed, at least straight away, thereby resulting in a publisher not being assigned the relevant colour. For instance, if a publisher allows the self-archiving of pre-prints, but also stipulates that the post-print can only be self-archived in a year’s time, i.e., after a one year embargo, then that publisher is awarded the ‘yellow’ colour, but not the ‘blue’ or resulting ‘green’ colour.

5. Application of the logic

However, this simple logic is not sufficient in all situations. The CKB contains many conditions and restrictions which are inter-related, leading to some complex algorithms for determining the colour of a publisher. The complexities of the CKB self-archiving status logic is particularly evident where publishers specify detailed circumstances applying to ‘restrictions’. Not all terms represented in the restrictions category are actually ‘colour-loss’ restrictions. When assigning colours, close consideration needs to be taken into the effects of specifying ‘what can be archived’ and any further ‘conditions’ and ‘restrictions’.

5.1. The effect of ‘types’ on self-archiving status

5.1.1 Pre-prints.

Although a work before submission to a publisher can be regarded as a pre-print, a publisher which requires the pre-print to be removed on submission to the journal is not assigned a ‘yellow’ colour – i.e., is considered ‘white’ rather than ‘yellow’. This is because the publisher is prohibiting the self-archiving of the pre-print during the peer review process. Moreover, it should be the case that before entering into a relationship with the publisher (i.e., before the CTA is signed), all authors are free to self-archive their work. Allowing the self-archiving of all other forms of pre-prints, from submission to the publisher to acceptance (before any copyright agreement is signed) will result in a publisher being assigned a ‘yellow’ colour (or ‘green’ if the post-print can also be self-archived).

5.1.2 Post-prints

For the purposes of the CKB logic, both author and publisher version of the post-print have equal weight, i.e., they have the same effect on the self-archiving status of a publisher. For example, if a publisher allows the self-archiving of a ‘post-print – author version’, but not the ‘post-print – publisher version’ it is still assigned a ‘blue’ colour. This is because the most important aspect of the post-print is the *content* of the work, rather than its *presentation*.

5.1.3 Parts of a work

All three definitions of pre-prints and post-prints refer to the whole work. If a publisher only allows the self-archiving of parts of a work, this is not considered to be self-archiving, and so cannot be awarded the related colour.

5.2 The effect of ‘Conditions’ on self-archiving status

As most conditions do not prevent an author from self-archiving straight away, these requirements do not have a negative effect on a publisher’s self-archiving status, i.e., they do not result in a loss of colour.

5.2.1 ‘Where’ conditions

Although, in most cases ‘where’ an author can self-archive is considered a condition, there are two instances in particular where it does have a negative effect on self-archiving status. These are discussed below.

- Some publishers only permit work to be mounted on a secure network, such as intranets and electronic reserves, which require an authentication mechanism to access the work. In this case, not only is an author not free to choose where to mount the work, but it also restricts its access to specific groups. Therefore, this does not constitute self-archiving, and so results in the publisher losing a colour and being classified as white.
- There are also some publishers who specify that a work must only be self-archived in a specific digital archive/repository, or only on their online version of the journal. In many cases, publishers insist on the author only mounting work on PubMed Central (www.pubmedcentral.nih.gov), so as to accommodate the US National Institutes of Health (NIH) self-archiving policy (US NIH, 2005), which at the moment also allows an embargo to be specified.

This policy, however, is currently under review. In other cases, publishers only offer an open access option using the ‘author-pays’ model, where authors, or funding bodies on their behalf, pay for the work to be openly accessible. When considering how to categorise these publishers it was decided that if the only allowable archive is a commercial site, such as a journal archive, then this does not constitute open access archiving as described by the Bethesda definition. Therefore, in such a case, the publisher loses the appropriate colour. If, however, the publisher specifies a digital repository, or online version of the journal, but still allows the author to self-archive elsewhere of their choosing then this does not result in a loss of colour.

Although both these instances result in a loss of colour, they are not technically considered ‘restrictions’ as they prevent self-archiving ever taking place. ‘Restrictions’, such as embargoes, only delay self-archiving.

5.3. The effect of ‘Restrictions’ on self-archiving status

Restrictions are more prohibitive as they prevent immediate self-archiving. They therefore often result in a loss of colour.

Some publishers state that an author has to formally ask for their permission in order to self-archive. Even though an author *may* be able to self-archive a work by receiving this formal permission, this is still regarded as a restriction and no colour is awarded. This is because the author would have to wait for the permission to be granted before they can self-archive, and, in some cases, they may not get the required permission.

5.3.1 ‘When’ terms/embargoes

Publishers which stipulate an embargo do not fully support self-archiving, as authors would have to wait before they can make their work openly accessible to the public. As embargoes describe *when* work can be self-archived, all ‘when’ terms are regarded as ‘restrictions’ and should lead to a loss of colour. However, this decision has major implications on the CKB logic. To illustrate this, consider a statement in a CTA that states “the work may be self-archived before its publication in the journal”. This is clearly not a restriction. However, if a publisher’s CTA specifies that “the work may be self-archived only on its publication in the journal” then this could be regarded as a restriction. It may not at first seem to be a restriction, and even the Bethesda definition of Open Access states that an article only needs to be “deposited immediately upon initial publication in at least one online repository”. This indicates that access prior to publication (e.g. after an article has been accepted, peer reviewed and copy edited but not yet published in print or made available on the publisher’s web site) is not a requirement of Open Access, but on reflection it is apparent that it should be considered an embargo. This is because authors would have to wait until their work is published before they can self-archive. In some cases, this may take a period of months following the acceptance of the work by the journal. However, this statement was not felt to be as restrictive as other embargoes that commence from the date of publication, therefore it was thought unhelpful to make it a colour-losing restriction for publishers. The CKB logic was therefore programmed to exclude this restriction. Statements of this kind are actually presented by the CKB as ‘conditions’ rather than ‘restrictions’, so as to inform, but not confuse, end users.

5.3.2 Work removal and replacement

Self-archiving implies that a work is made permanently available, so the removal or replacement of a work may result in restricting access for a certain length of time. For example if a publisher insists that a pre-print be removed on submission it loses the 'yellow' colour, meaning the publisher goes from 'yellow' to 'white'.

However, if a publisher requires an author to remove a pre-print but, on removal of the preprint, lets the post-print be archived, then the publisher would keep its 'green' status. This is because the publisher allows the author to self-archive both the pre-print and the post-print, and so there is uninterrupted access to the work - from pre-print to post-print. The post-print is in practice considered the most important version to make available to the public.

Some publishers allow pre-prints to be self archived but specify that they be removed on submission or during the peer review process, but that once the peer review process is finalised a post-print can be mounted. This effectively means that there may not be entirely uninterrupted access to the work. However, despite these circumstances, SHERPA/RoMEO still awards a 'green' colour, as there is close to uninterrupted access.

However, if the publisher only permits the self-archiving of pre-prints, but requires these to be removed on acceptance or replaced by a link and citation, for example, then the publisher would be awarded a 'white' status. This is because the pre-print will only be openly accessible for a short time, and would not be replaced with the definitive version of the work.

The logic regarding work removal and replacement is an example of balancing the requirements of publishers and of self-archivers in the open access environment. Publishers do not have to allow indefinite access to both pre-print AND post-print to acquire 'green' status, simply they have to allow predominantly uninterrupted access to pre-prints before publication and immediate and uninterrupted access to post-prints after publication.

6. Conclusions

The Appendix to this paper summarises the logic behind awarding a publisher 'white', 'yellow', 'blue' or 'green' status. The use of these self-archiving colours as a means of identifying self-archiving friendly publishers plays an important role in the open access movement. Not only does it provide both those who self-archive and publishers with the information they need on self-archiving policies of publishers, but also acts as an advocacy tool to encourage self-archiving practices and to send out a clear message about what open access self-archiving actually means.

As it has far-reaching consequences to the OA movement, considered decisions must be made when creating the logic which calculates the self-archiving status of publishers. Such decisions can be controversial, and this paper has discussed just some of the political decisions to be made when deciding what being a 'green' publisher actually means. The authors would welcome comments from stakeholders regarding the decisions we have taken.

ACKNOWLEDGEMENTS

We wish to thank JISC and SURF for funding this research. We also wish to thank Mike Gardner for his helpful technical assistance.

EDITOR'S NOTE

This paper follows on from a series of other papers on the RoMEO project – the most recent one being currently considered for publication in *Journal of Information Science* (Jenkins et al, 2006).

References

(All URLs were checked 4th January 2007)

Brody, T., Carr, L., Gringras, Y., Harnad, S., Hilf, E.R., Hitchcock, S., Oppenheim, C., Stamerjohanns, H. and Vallieres, F. (2004), "The access/impact problem and the green and gold roads to Open Access", *Serials Review*, Vol. 30 No. 4, pp. 310-314.

Gadd, E., Oppenheim, C. and Proberts, S. (2004), "RoMEO Studies 5: IPR issues facing OAI data and service providers", *The Electronic Library*, Vol. 22 No. 2, pp. 121-138.

Jenkins, C., Oppenheim, C., Proberts, S and Hubbard, B. (2006), "RoMEO studies 7: creation of a controlled vocabulary to analyse copyright transfer agreements", paper submitted to the *Journal of Information Science*.

Suber, P. (2003), *Bethesda Statement on Open Access Publishing*. Available at: <http://www.earlham.edu/~peters/fos/bethesda.htm>.

US National Institutes of Health (NIH) (2005), *Policy on Enhancing Public Access to Archived Publications Resulting from NIH-funded Research*. Available at: <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-05-022.html>.

APPENDIX

This Appendix summarises the self-archiving status logic found in the CKB, as applied to its controlled vocabulary. It gives the definitions of each self-archiving colour, along with the situations which make a publisher that colour.

What constitutes self-archiving?

True self-archiving means that the author has a choice of **where** they can place their Work and **when**.

All authors can self-archive their Work before submission to a publisher, but it is when they enter into a relationship with the publisher that any conditions and restrictions stipulated by a publisher have an effect on its self-archiving status.

What are the general effects of conditions and restrictions?

In general, conditions do not result in a loss of colour, whereas restrictions do.

1. WHITE – Self-archiving of the PRE-PRINT and POST-PRINT is not formally supported.

Example reasons for awarding “White”

- Cannot self-archive the pre-print and post-print.
- Can only self-archive the pre-print and post-print with formal permission from the publisher.
- Can only self-archive the pre-print and post-print in a specified commercial digital repository or online version of a journal.
- Can only self-archive the pre-print and post-print on a secure network, i.e., has an authentication mechanism in place.
- Can only self-archive the pre-print and post-print at a specified time after (not before or on) publication.

2. YELLOW – Only the PRE-PRINT (i.e., up to and during peer-review) can be self-archived.

Example reasons for *not* awarding Yellow

- It will not accept the submission of a paper that has already been self-archived.
- It requires the pre-print to be removed on submission to the journal (and does not allow archival of a post-print).
- It requires the pre-print to be self-archived in a specific place only, i.e., on a secure network, or in a particular digital archive and/or online version of the journal.
- Its formal permission is required to self-archive pre-prints.

3. BLUE – Only the POST-PRINT (i.e., after peer-review) can be self-archived.

Example reasons for *not* awarding Blue

- It requires the post-print to be self-archived in a specific place only , i.e., on a secure network, or in a particular digital archive and/or online version of the journal only.
- It requires the post-print to be self-archived at a specified time after (not before or on) publication.
- Its formal permission is required to self-archive post-prints.

4. GREEN – The PRE-PRINT and POST-PRINT can be self-archived.

A publisher is still considered GREEN even if the pre-print has to be replaced with a post-print on submission. This is because the author has been able to self-archive both the pre-print and post-print, even if not at the same time and there has been uninterrupted (or at least minimal interruption to) access to both types of work.

Example reasons for *not* awarding Green

- It requires the pre-print and post-print to be self-archived in a specific place, i.e., on a secure network, or in a particular digital archive and/or online version of the journal only.
- It requires the pre-print and post-print to be self-archived at a specified time after (not before or on) publication.
- Its formal permission is required to self-archive pre-prints and post-prints.