Managing Copyright in the Digital Repository: Beyond "Undue Diligence"

Paper presented at the ACA Conference, Victoria, June 2014

By Richard Dancy

Introduction

My topic is copyright in the context of a digital repository that acquires, preserves and provides access to born-digital records. Do copyright issues play out any differently here than in the traditional paper-based or analog archives? And if so, what the policy and technical requirement for managing them? The crux of my argument is that an overly cautious, conservative approach to copyright that evolved with analog records and makes sense there provides an inadequate basis for managing born-digital records and we need an alternative. So I will say why I think that and try to sketch out the elements for an alternative, which can be characterized as a risk-management approach.

It seems to me as archivists we instinctively hate copyright: it's something that just gets in our way, holds us back, it's an obstacle. To tie this back to the conference theme: in any self-respecting archival utopia copyright does not exist. Or, if it does exist, it doesn't apply to archives. Or, if it does apply to archives, is reducible to a few simple rules that we can apply with no fuss no muss. All of this is "utopian" in the bad sense of the word: it's not going to happen. But the real value of utopia is not whether it exists or not – by definition, it doesn't exist, it is no-place – it lies rather with the desire for utopia, because this works as a spur to critical and experimental action in the present: we know more or less what we don't like about the status quo, we have any idea of where we'd rather be, let's start building things that close the gap. So that is the spirit in which this is offered.

The title of my paper, the bit about "undue diligence" I've cribbed from the name of a seminar that was held in 2010 under the auspices of the Online Computer Library Center. This looked at low-risk strategies for making unpublished materials accessible online. The basic idea of "undue diligence" is that in relation to copyright

archivists are more conservative than the law itself requires us to be – we complain about the law, but we work with a very narrow interpretation of it that fails to take advantage of existing elements that we can use to our advantage. The boundaries in which we operate may in fact be more flexible than we make them out to be and there may be space for us to push them further. So like all good utopians, the first place to start is by changing ourselves and the way we perceive, in this case, the way we perceive our copyright situation.

What I will do here is talk about what I mean by "digital repository" and why I think copyright does play out a bit differently here; identify the traditional conservative approach and why it is problematic in this context; say a bit about some recent initiatives to go beyond it and why developments in the Canadian copyright landscape over the last few years make this a good time to try something similar in Canada, using the concept of fair dealing to stake out the ground for a new approach; lastly I'll try to sketch out the elements of what that alternative policy approach might be.

A couple of caveats. I work at SFU Archives and we are now in Year 2 of a three-year project to build up our capacity to preserve accessible digital archival materials. Copyright is one of the policy areas we wanted to look at – in part I'm reporting on what we are doing or proposing to do and so we are hoping to get some feedback on this. Earlier this year, we prepared an internal discussion paper looking at these issues, and this presentation is a much-condensed version of that report. The paper is available on our website, I'll give the url at the end of the presentation. It is a long and long-winded document, but it does lay out actual arguments and more detailed analysis than I can give here in 20 minutes.

It seems to me customary in these kind of talks for the speaker to say "I'm not a copyright lawyer but ..." and the effect of the but is that at least they are an expert, they've followed the case law, they've studied the issue for a long time and so on. So, I have to say, I'm not an expert but ... And the old adage that a little knowledge can be a dangerous thing might be true here. But we know that there

are experts in our field on this area and we know also that most institutions are facing the same problems that we are.

There is an opening today I think in the Canadian copyright landscape for archivists to develop a new approach to copyright. This would achieve in the archives sector the broad public policy goal that the Canadian courts have repeatedly said is the purpose of the *Copyright Act*; namely, to strike a balance between the public interest in dissemination and advancement of knowledge on the one hand and the interest of copyright owners in securing their economic benefits on the other. If the Canadian archives community can reach broad agreement on what constitutes such an approach, we can develop common technical and policy tools to support it. All repositories stand to benefit. The Canadian Supreme Court has shown a willingness to consider "the custom or practice in a particular trade or industry" as a relevant consideration in assessing the fairness of particular practices. The ability to ground our activities in a broader, shared community of reasoned, professional practice within the Canadian archives system is thus vital to the development of a defensible institutional policy.

Digital repository

So, digital repository, here is my very crude representation. The scope of the project does include some digitization of analog materials, but the primary motivation driving it is to develop the capacity to capture born-digital records into our archival systems. We want to be able to do in the digital world what we've long done in the analog one: namely, acquire, preserve and provide access to records of enduring archival value.

We're working with Archivematica and Atom software, which many of you will be familiar with, while our own university IT Services are provisioning the technical infrastructure in terms of storage space and VMs needed to run the software. Basically we are following the OAIS model (Open Archival Information System). Digital objects are transferred to the system and turned into Submission

Information Packages (SIPs). These are run through or ingested into Archivematica software which performs a whole range of operations on them: virus checks are run, unique IDs assigned to each object, metadata extracted, file formats identified and copies made in designated preservation and access formats, and so on. At the end, Archivematica generates two basic outputs: the Archival Information Package (AIP) comprising the original objects + preservation copies + metadata packaged up into a single object and sent to the storage system for long-term preservation; and the Dissemination Information Package (DIP) representing the access copies + minimal descriptive data sent off to the access system for further description and public dissemination. In our configuration, the access system runs on Atom software and the normal method of providing access is via the internet and a web-based finding aid system that links the archival description directly to the digital access copy.

It's worth noting that copying activities take place throughout this whole set-up, e.g. copying on transfer, ingest, archival storage, data management, repository administration. It would be interesting to see how copyright applies in all these functions. But I'm not doing that, my focus here is only on the dissemination function and the provision of access. Even so, it does not cover all the policy issues relating to dissemination, only those pertaining to copyright. The other huge issue here is protection of personal information and compliance with privacy protection laws. We recognize that any comprehensive solution must integrate both intellectual property rights management and privacy protection, but I'm only going to discuss the copyright side here.

The copyright problem

So what exactly is the copyright problem here? Basically, it springs from the collision of three facts: the nature of copyright, the nature of archives, and the nature of the internet. When you add them together you get a problem.

Copyright: archival records are typically protected materials under the Canada Copyright Act. There's an interesting discussion of the meaning of originality in the 2004 Supreme Court case CCH Canada Ltd v. Law Society of Upper Canada. In order to attract copyright protection, a work must be an expression of ideas involving the exercise of skill and judgment. This still sets the bar pretty low. There may well be some records that do not qualify, but most narrative archival documents will be protected under copyright. Ownership derives from authorship: the copyright owner is typically the author of the record; in a business context where the author is an employee, this will typically be the author's employer.

The nature of archives on the other hand is to be an aggregate of records made **or received**. It is "records received" that are most problematic, the incoming correspondence, the reports and documents received as attachments, the various documents that the fonds creator collected from elsewhere and included among his or her files. The fonds creators owns these documents as physical property, but the third parties who authored them remain the copyright holders. This means that any given archival fonds or series will typically include a large volume of third-party copyright-protected works. In addition, archival documents are not typically publications, rather they are records, practical by-products of day-to-day activities. In many cases there is often simply not enough information to identify the copyright owner(s) or if you can, to locate them (so-called orphan works) in order to obtain permission.

Finally, the problem by nature the internet, because it means that online dissemination of an archival record is a form of copying that resembles an act of publishing; if done without the copyright owner's permission, it is liable to claims of copyright infringement.

So, to summarize: we have fonds that contain a large portion of third-party protected materials, many of which themselves are orphan works. Disseminating these without permission exposes the Archives to liability of infringement.

The conservative approach

Given the difficulties, SFU Archives – like many institutions – has up till now adopted a cautious approach to copyright, above all seeking to avoid anything that smacks of infringement. This is what I am calling the "conservative approach". Material will only be disseminated if one of the following three conditions is met:

- The material is already in the public domain; or
- Copyright is owned by the repository, either as first owner or through assignment of rights; or
- Permission of the copyright owner has been obtained.

This leaves the archivist with the daunting task of identifying all third-party copyrighted material within a fonds through item-level review, locating all owners, and obtaining explicit permissions before disseminating. Given the volume of documents in a typical fonds, no archives will have the resources for routine item-level review. Given the extent of orphan works among the protected materials, moreover, large portions of a typical fonds or series will not meet the conditions required for dissemination.

In the traditional paper-based or analog repository, the conservative approach is not especially problematic because it does not hinder the Archives' ability to perform our core function of providing access. Here there is a clear distinction between providing access and making a copy. In the physical reading room researchers are able to consult third-party protected materials, and the Archives can make single copies for individual researchers under the Copyright Act's fair dealing and LAM (libraries, archives and museums) exemptions.

Problems with the conservative approach

For the digital repository the situation is different in at least three respects:

1. The digital repository relies on an online access system and thereby blurs the distinction between access and reproduction. Online dissemination is directed to the public at large rather than a point-to-point transmission to a specific individual; as such, it resembles an act of publishing. Providing access via dissemination is itself a copying activity and online researchers cannot as such consult third-party protected material without putting the Archives at risk of infringement.

2. With digitization, the Archives can choose which material to upload to the digital repository and which to exclude. Anything that poses any complex copyright questions can simply be left out. But with born-digital records, simple exclusion is not an option. Routine transfers of born-digital records to the repository will include third-party copyrighted material: it cannot be excluded from the repository, it must be actively managed within the system. Accordingly, the Archives will need the means and resources to identify copyright-protected materials, a policy for handling different copyright scenarios, and technical tools that translate policy decisions into "actionable" system rules. In the physical reading room, item-level analysis for copyright clearance is an exceptional case that occurs from time to time. But for the digital repository wanting to implement the conservative approach, it becomes a routine requirement for normal provision of access. These are time- and resource-heavy demands that are not feasible to implement.

3. But even if the Archives has the will to apply the conservative approach and even if it has the resources to be able to do so, the result is still bad because it leads to a kind of patchwork access that threatens the archival integrity of the records themselves and their reliability for research. Within an aggregate of related records (e.g. a file) protected and unprotected materials are often inextricably intermixed. Simply excluding the protected works obscures the archival bond linking the various records and may result in a misleading presentation of those that are disseminated.

For the digital repository, then, copyright touches on access no less than reproduction. Adhering to the conservative approach here severely limits – if not effectively paralyzes – the Archives' ability to provide access because it requires time-intensive item-level analysis as a condition for making materials available. In many cases (orphan works) it leads to denial of online access; in other situations, it results in a misleading patchwork of access.

One solution to the impasse is simply to channel all access to digital materials through specially configured terminals in the Archives reading room that allow users to consult documents while preventing copying or downloading. This effectively reduces the "online reading room" to an adjunct of the physical one. Aside from the cost of providing such terminals, this does not effectively utilize the potential of online technology to expand access to our holdings, and it disadvantages researchers who are not able to physically travel to the repository. It may be the only solution "if all else fails." But is it the case that all else must in fact fail?

Recent US initiatives

Dissatisfaction with this state of affairs has motivated the search for an alternative approach. Two recent initiatives in the United States are the *Well-intentioned practice for putting digitized collections of unpublished materials online* (W-iP guidelines) developed by the Online Computer Library Center (OCLC) in 2010 and subsequently endorsed by the Society of American Archivists (SAA) in 2011; and the *Code of Best Practices in Fair Use for Academic and Research Libraries* devised by the Association of Research Libraries (ARL) in 2012.

W-iP "promotes a well-intentioned, practical approach to identifying and resolving rights issues that is in line with professional and ethical standards," and it offers specific recommendations for assessing and minimizing risk. The ARL code identifies a number of situations – including "creating digital collections of archival and special collections material" – in which librarians and archivists can and should rely on the fair use exception in the US *Copyright Act* (section 107). For each scenario,

the code articulates a fair use principle, specifies the limitations that should be observed, and identifies enhancements that would strengthen the fair use case. Both documents make the point that archivists have typically been more conservative than the law in fact requires, and both make room for online dissemination of protected material without the permission of copyright owners in certain circumstances. They do not say that dissemination without permission is always fair use, only that it **can be** fair use, and they offer guidelines for assessing circumstances so that archivists can make informed, professional judgments in particular cases.

Changes in the Canadian copyright landscape

Can Canadian archivists translate these ideas and practices to our own legal context? There are differences between Canadian and US copyright law in general and between Canadian fair dealing and US fair use in particular. But there are grounds for thinking the Canadian copyright landscape does offer archivists space for a similar shift in approach.

In 2012 Parliament passed the *Canada Copyright Modernization Act* and, just weeks before, the Supreme Court issued five decisions relating to copyright (the so-called "copyright pentalogy"). With these decisions, the Court affirmed and elaborated the "user rights" approach to copyright it had signaled earlier in Théberge v. Gallerie d'art du Petit Champlain Inc. (2002) and in the landmark CCH Canadian Ltd. v. Law Society of Upper Canada (2004). This approach emphasizes that the purpose of the Act is not simply to protect the economic interests of copyright owners, but to balance those against the public interest in the dissemination of works of art and culture and the advancement of knowledge; and it treats fair dealing not simply as a defence against charges of infringement, but as a fundamental user right central to the very purpose of the copyright.

Fair dealing allows use of protected works without permission in certain circumstances. The 2012 revision of the *Copyright Act* expanded the list of

allowable purposes under fair dealing to include education, satire and parody, adding to the existing purposes of research, private study, criticism, news reporting and review. CCH insisted that fair dealing "must not be interpreted restrictively" (CCH, para 48). CCH also established a test of six factors by which to assess fairness: purpose of the dealing; character of the dealing; amount of the dealing; alternatives to the dealing; nature of the work; and effect of dealing on the work. These now play a role in Canada analogous to that played in the US by the four factors set out in the American Copyright Act, i.e. they provide the basic framework for analyzing the fairness or unfairness of any given use. A recent comparison of fair dealing / fair use in Canada, UK, and the US found that "Canada's regime is the most user-centred." If American archivists are able to make headway with US fair use, it seems reasonable to think that Canadians can do the same within fair dealing.

Fair dealing

In the discussion paper mentioned earlier that is available on our website, we tried to apply the CCH tests for fair dealing to the practice of disseminating archival materials online without permission. Our conclusion was that there is no simple yes or no, i.e. we cannot say that online dissemination of archival materials as such "is fair" or "is unfair." Instead, under each factor there are a number of considerations that lead to fairness in some cases and to unfairness in others. I've summarized our findings here in table form, I won't go through it in detail, just highlight a couple of general points.

The basic idea here is that if we can identify the factors that tends to fairness, we can build the policy framework around those.

The factors tending towards fairness or unfairness relate both to the materials themselves, but also the manner under which they are disseminated. Dissemination tends to fairness the more the materials are unique, unpublished documents not created with a created a commercial intent and not commercially available

elsewhere, exhibiting a low degree of originality (exercise of skill and judgment); and the more access decisions are grounded in a documented policy and the access system is able to accommodate different types of access for different types of material or purpose (eg. in some case, view-only, in others full download). Conversely, dissemination tends to unfairness the more the materials include non-record publications or other documents commercially available elsewhere or exhibiting a high degree of creative originality; and the more access decisions are ad hoc and the access system can only provide access by routinely creating durable new copies outside the repository (download).

On balance, the typical archival scenarios fall closer to fairness than unfairness. It is the *character of the dealing test* (factor 2) that is most difficult to meet because dissemination over the Internet fails the "point-to-point" transmission test and looks more like a publishing activity that tends to unfairness. To mitigate this unfairness, we need access systems that recover for the digital realm the traditional distinct between access (view only) vs reproduction (download).

There is no certainty, then, that a fair dealing defence of archival dissemination will succeed. But it remains the most promising (and perhaps the only) starting point for any advance beyond the conservative approach. The fair dealing tests provide a framework for developing "well-intentioned" standards of judgment and consistent practice, and fair dealing gives Canadian archivists a basis for assessing and incorporating the recommendations of US best practice guidelines. In CCH, the Court allowed that "the custom or practice in a particular trade or industry" could be a relevant consideration in assessing what counts as fair dealing in that sector (CCH, para 55). If the archival community uniformly adheres to a conservative approach, there is the danger that this will become entrenched as the "custom" of our industry; conversely, if we can ground a more open practice in a reasoned policy, this could have a positive impact on how the courts assess archival practices in the future.

Elements of a risk management approach

What are the outlines of a more open policy? The conservative approach applies a series of yes/no dichotomies: either a work is protected or it is not; if it is protected, either the repository has obtained permission to disseminate it or it has not; if it has not, it cannot disseminate. In place of these dichotomies, a risk management approach sees a continuum spanning varying degrees of risk, and it proposes to deal differently with different materials according to where they fall in the risk spectrum. Put simply the proposal is: high-risk works will be managed according to the conservative requirements, but low-work material will be disseminated without permission.

For this approach to work, four key elements are needed:

1. A high-level policy statement that explicitly aligns the archival mission with the broader purpose of copyright to achieve a balance between the public interest in the advancement of knowledge and copyright owners' economic interests.

2. Criteria for assessing risk that are clear and grounded in considerations drawn from the *Copyright Act* and the application of the CCH fair dealing tests; criteria should be capable of application at aggregate levels.

3. An online access system ("virtual reading room") designed to recover for the digital realm the traditional distinction between access and reproduction.

4. Procedures for complaints and dispute resolution ("notice and takedown").

So basically over the coming year, we want to work on each of these. The key for #1 and #4 is that they be highly visible from within the access system.

Risk assessment

I'll just focus here on the second element, risk assessment. What are the criteria for assessing risk?

The Court has characterized the purpose of copyright in terms of "balance," and this concept provides a basis for articulating risk. In the narrow sense, risk is just the liability of being sued. In a broader sense, the Archives takes copyright risks when it tilts the balance too heavily in favour of the public interest in dissemination against the owner's legitimate interests. But what are the criteria for assessing a "reasonable balance"?

Copyright protects creators' moral rights and owners' economic rights. In the discussion paper, we argued that dissemination only marginally touches on moral rights – the threat to moral rights comes not from dissemination as such but rather from users and what they subsequently do with the materials that have been disseminated. That leaves owners' economic rights as the main source for criteria to assess risk arising from dissemination as such.

The basic principle proposed is:

- Does the record – considered as a self-contained work circulating as an "article of commerce" – have a commercial value from which its owner could reasonably expect to accrue economic benefits?

The greater the potential commercial value, the greater the risk in disseminating the record without permission.

Archival records are typically utilitarian documents created as by-products of activities aimed at some other practical goal. In creating a record, the author is not typically creating an "article of commerce" but transacting some business, which requires the production of documentation (a form, correspondence, minutes, a report). As such, few archival records have an independent economic value for their

copyright owners. But some do, and the detailed analysis of the CCH tests provided several criteria for identifying records with potential commercial value:

- The document is a published work included in a fonds (likely to be commercially available elsewhere).

- The document is unpublished but created with commercial intent in mind.

- The document is unpublished, but is creatively original (embodies a high degree of literary, artistic or professional skill and judgment).

- The document is unpublished, but was authored by an individual who is (or has become) prominent in a certain field or has attained a certain fame, celebrity or notoriety (even documents of low originality may have a commercial value if the author is a "famous" figure).

At the other end of the spectrum, there may be archival records that do not even meet the threshold of "originality" required to attract copyright protection:

Applying these criteria will admittedly be somewhat subjective, but it should be possible to group material into three distinct "risk" categories that correspond to different dissemination outcomes. [No risk / Low risk / High risk].

The expectation is that within any given fonds, the bulk of records will be low risk; there will be a very small proportion that are not protected (not original) and another small volume that are high risk. We also expect to be able to apply the criteria at aggregate levels: the aggregate-level review should act as a kind of "triage" identifying high-risk materials that require more comprehensive item-level analysis.

The next step for the Archives is to test these expectations by actually applying the criteria to a number of test cases. As the Archives gains experience applying the

criteria, we will look to refine them and to develop standardized tools (e.g. checklists, guidelines) to support consistent decision-making.

Let me just say a few words about this notion of a virtual reading room. The basic idea is to apply the concept of "technological neutrality" to the provision of archival reference services. The researcher's experience should be the same whether they visit the physical reading room or view material online. They should be able to consult copyright-protected materials and request single copies under fair dealing. Ideally the access system should be able to automate different access outcomes depending on the type of request and the risk status of the material requested. This is easy to say, but I think there are really quite a lot of challenges from a technical point view to implement it.

Conclusion

So, let me sum up. The risk assessment approach we are proposing centres on the concept of reasonable potential commercial value. It aims to identify high-risk materials where dissemination without permission would unfairly tip the balance against owners and thereby deprive them of economic rewards. It proposes to deal with these works conservatively, seeking owners' permission before disseminating.

If, however, the records disseminated have no discernible commercial value, then it will be deemed fair to allow the public interest in dissemination to override the owner's right to authorize copying because here the copyright owner's economic interests are not harmed. In these cases, dissemination does not deprive the owner of economic benefits that realistically do not exist and so does not tilt the balance between public interest and owners' rights unfairly against the owner. But if that claim is to be more than a mere subjective assertion, there need to be some standards of judgment for assessing particular cases. Fair dealing provides the basic framework for articulating those standards.