Copyright provides rights over the distribution and use of intellectual works. The endeavor of populating institutional repositories with articles, images, films, and the like requires making and sharing copies of these items, so copyright issues permeate this activity. At first glance, copyright may seem to be a barrier to beneficial use. But there are exceptions and qualifications which make copyright a constraint that can be adapted to, or even used to encourage further sharing.

An understanding of copyright is crucial for library workers engaged with overseeing IRs. Some degree of expertise will also be helpful when a librarian educates and advocates to authors, copyright holders, and administrators on copyright matters. This chapter will review foundational concepts of copyright law and discuss what works are covered by copyright, what rights come with copyright, and what exceptions are available. The chapter will also assume that IRs are generally operated by nonprofit organizations pursuing academic, educational goals.

This information will help librarians formulate their institutional repository’s own copyright policies and practices. Dealing with copyright questions
is part compliance, part education, and part risk assessment. Many decisions involving copyright will depend heavily on the facts of the situation. Even when presented with the same question regarding copyright, different librarians will make different decisions, depending on their specific circumstances and professional judgments. This is reasonable and expected, but a solid grounding in copyright law will give one more confidence when a tricky question arises.

THE RIGHTS PROVIDED BY COPYRIGHT

Copyright provides a specific set of rights (often called a bundle of rights) that control how a work can be used, copied and shared.¹ The rights from this bundle that are most relevant to IRs are the reproduction and distribution rights. The reproduction right controls the making of copies of a work. Clearly, digitization or uploading a copy into an IR implicates this right. The distribution right involves giving, selling, or renting copies of a work to other people. An IR “distributes” copies when researchers download those copies from the repository. Two other rights, the display and performance rights, are mostly for visual works (art, photographs) and musical and theatrical works (songs, plays, movies). If an IR collects and offers access to images or audiovisual works, then display and performance rights will come into play.

All the rights mentioned thus far have involved sharing works in the form in which they were originally created or published. The “derivative works right” controls making a new form or version of a work, or making a new work that is substantially based on another work. For example, suppose an IR has uploaded a book in PDF format. If the IR converts the book into a MOBI e-book file, records an audiobook of the book, or translates the book into another language, the IR is making derivative works. Finally, the “digital audio transmission right” controls playing music over streaming websites. This right is usually relevant for commercial streaming services, and most IRs will not need to be concerned about this right.

This brief review of the rights provided by copyright can be supplemented by three more points. First, there are ways in which IRs can use works that do not involve any of these rights, like listing the metadata about a work, or linking to a copy located elsewhere. Second, these rights can be divided and transferred. This is why a book can be sold by one publisher in paper form and sold by another company as an audiobook. Third, when seeking permission from a copyright holder to share a work, IRs should be specific about which rights they need. For instance, an IR can ask for permission to reproduce, display, and distribute an article so that it can provide open access to it. The IR may also request permission to make derivative works to the extent needed to
convert the work into newer file formats. Now that we know what rights come
with owning the copyright in a work, let's look at what works can be covered
by copyright and who can own them.

**WHAT CAN BE COPYRIGHTED**

Copyright applies to "original works of authorship fixed in any tangible
medium of expression." This one sentence covers a vast amount of materials
that libraries routinely deal with, but it also excludes some important cate-
gories of works from copyright protection. First, in order to be protected by
copyright, the works must be original. The level of originality needed is very
low, but this requirement has been invoked to find that phone books, recipes,
and exact reproductions of artworks whose copyright has expired do not have copyright.

Second, the requirement that the work be fixed in a tangible medium
means that the work has to be "recorded" in some way. Simply thinking up a
great idea or humming a new tune to yourself does not create a copyrighted
work. These "creations" have to be set down in some tangible form to be eligible
for copyright. The second limitation also means that single words, and ideas,
systems, theories, and concepts cannot be copyrighted. One might articulate a
concept in writing and that specific expression of it would be copyrighted, but
another person could write their own articulation of the same concept, which
would also have its own copyright. So while copyright can cover a breathtaking
amount of material (in addition to most published works, think of every e-mail
you have ever sent, every photo on your phone, and most notes you have writ-
ten down), it excludes the basic ideas we all use to make new creative works.

Aside from the originality and fixation requirements, there are two major
ways a work is not copyrighted and is thus in the public domain. The term
public domain has a particular meaning in copyright; it means anything that
is not controlled by copyright. A work can be publicly available, but not in the
public domain. Works created by employees of the U.S. federal government,
for example, are ineligible for copyright. The federal government produces
publications, data, and legal materials, all of which can be freely copied. The
one caveat to this category is that federal government works are sometimes
created by private contractors, not federal employees. Those works can be
copyrighted, even if they are commissioned or purchased by the federal gov-
ernment. Also, sometimes a federal government publication will include some
elements of previously copyrighted material. For example, a government
report might include part of an academic paper written by private research-
ers. That copyrighted portion does not lose its copyright merely because the
federal government purchased or used it.
Copyright lasts a very long time, but like all things, it does eventually end, and works enter the public domain when their copyrights expire. Copyright typically expires 70 years after the death of the author (or last living author, if there is more than one). For works that do not have authors that can die as humans do (corporations, and anonymous or pseudonymous authors), copyright expires 95 years after publication (or 120 years after creation, whichever comes first). This means that virtually all works published before 1925 are in the public domain. There are nuances that apply to relatively scarce special cases, but there are several excellent resources that will help you analyze whether a given work still has copyright protection.

Under earlier versions of the U.S. Copyright Act (the last major reform of which was in 1976), works had to comply with other requirements, such as the creator filing for registration with the Copyright Office, renewing the work’s copyright at a certain time, or placing a copyright notice on each copy of the work. The 1976 reform removed these formalities, but works published before then are still governed by the earlier laws, so some works published after 1925 may have entered the public domain due to noncompliance with these formalities. The University of Michigan has published a guide on ascertaining whether a work has lost its copyright due to formalities. Frustratingly, sometimes there is insufficient information to determine if a work’s copyright has expired. An IR librarian will then have to look to her institution’s policies and practices for guidance and support, or try to find an exception to copyright.

COPYRIGHT OWNERS

Who owns the bundle of rights provided by copyright? By default, copyright vests with the author; that is, the person (or people) whose actions resulted in the creation of the work. This can lead to some surprising results. If one person takes a picture of another person, the photographer has the copyright to the image, and the subject does not. If two or more people contribute enough to a work to be authors, they are all equally joint authors for copyright purposes. This may not align with disciplinary norms or the authors’ expectations. They should make a written agreement if they do not want the default rules to apply.

If an author creates a work as part of their employment, then that work is a “work for hire” and the employer (rather than the author) is deemed to be that work’s creator and copyright owner. Most academic institutions have intellectual property policies which provide that authors own their scholarly works, even if those authors have created them as part of their employment. However, academic and other authors often transfer their copyright to publishers through written agreements (i.e., contracts). A complete transfer of
copyright or exclusive license to a work must be in writing, so if an author published a work but did not sign a document regarding its copyright, he or she still holds the copyright.

Authors often sign these agreements as a routine part of the business of getting published, but they would be wise to scrutinize the agreements to ensure they don’t transfer away the rights they will need for future uses of their own works. IR librarians should inquire if any agreements have been signed that would prevent an author from authorizing an IR to reproduce and distribute their works. If the agreement provides for a complete transfer, then the author no longer owns the copyright, and permission to share the work would have to be obtained from the publisher, or some other new owner.

There are three major strategies an institution can use to help authors reserve the rights needed so that adding their works to an IR is an option. For the best outcome, an IR will probably have to undertake all three of these strategies. First, librarians should educate authors on their rights and highlight especially important copyright provisions in publishing contracts. Many authors do not fully understand the rights provided by copyright or carefully examine contracts for provisions that will affect their long-term interests. Even if a publishing contract provides for a copyright transfer or exclusive license, some publishers, through their policies on authors’ rights, grant back specific rights to authors, such as the right to deposit one’s work in an IR. These policies tend to have additional conditions, like requiring specific attributions, embargoes delaying free distribution, and limitations on which version of the work can be shared by the IR. Librarians should, if possible, examine the applicable agreements or publishers’ policies to see if noncommercial distribution by academic institutions is permitted.

A second strategy is to ask authors to attempt to amend their agreements with publishers. This can be done by suggesting specific language changes or by providing an addendum that, when attached to an agreement, retains crucial rights for authors. Several universities have produced excellent examples of addenda which provide that, notwithstanding the signed contract, the author will retain rights to allow later uses and open access to the work. Most addenda preserve the author’s rights to use the work in later educational and scholarly activities and to provide open access via the author’s personal website or an IR.

A third and more systemic approach is for institutions and funders to adopt open access policies. Open access policies, which are typically adopted through faculty governance processes, require authors to grant a nonexclusive license in their works to the institution for the purposes of reproducing it in an IR. This license is granted as soon as the work is created, and thus before the authors can sign an agreement with a publisher. While they reduce the transaction costs required to obtain permissions or negotiate publication
agreements, institutional open-access policies are not a panacea. To be palatable to faculty members, most of these policies allow for waivers upon request, and librarians must work to collect and deposit the works in the IR.

**DETERMINING WHO OWNS A COPYRIGHT**

If someone other than the author submits a work to an IR, some effort may be needed to find the copyright owner, to whom a permission request should be directed. In many instances, however, identifying the author of a copyright is a relatively simple matter.

You should start with the initial owner, the author or perhaps the author's employer, and see if they have signed a document transferring the copyright. With time, though, trails of ownership can fade. Copyrights can be sold and transferred as commodities, and as publishers merge, split up, or go out of business, who owns the copyright in a work can become unclear.

There are several sources one can consult when researching copyright ownership. Registering a copyright with the Copyright Office is not necessary to own the copyright, but it is required to sue for copyright infringement. Thus, virtually all commercially valuable works will be registered with the Copyright Office. Of course, this is still a small portion of all the copyrighted material that exists, but checking Copyright Office records is worthwhile. There are three main types of copyright documents a librarian may encounter when researching copyright ownership. The first is a copyright registration. This is the initial claim over a work. A registration will indicate the creator of the work, the owner at the time of registration, and sometimes contact information for requesting permission from the owner.

The second is a recorded document. When an owner transfers copyright to another person or corporation, they will record the transfer with the Copyright Office. With registration and recorded transfer records, one can trace the ownership of a given work.

The third type of document is a copyright renewal. Before the 1976 Copyright Act, owners had to renew their copyright to obtain an additional term of copyright. If an owner did not renew, perhaps because the work was no longer making much money, the copyright expired and the work entered the public domain. Renewing has not been required since 1976 (with a small exception for works by non-U.S. citizens), so renewals are only relevant for works published between 1925 and 1963.

For registrations and documents filed since 1978, the Copyright Office has a digital catalog that can be searched by a work's title, author, and registration number. The catalog entries are not the certificates of registration, but they will often indicate who owned a work at a particular time and will provide contact information. For documents filed prior to 1978, the Copyright
Office is developing a Virtual Card Catalog that displays scans of the office’s card catalog summarizing registrations and other ownership records. At this writing, the Copyright Office describes this product as a proof of concept and as likely to change.

Until 1978, the Copyright Office published a serial, the Catalog of Copyright Entries (CCE), which summarized registrations, renewals, and other filings. Prior to the Virtual Card Catalog, the CCE was the primary way to research the copyright ownership of works created prior to 1978. The CCE has been digitized, but most instances of it online are scans of the print issues. Google has made a more searchable version, but it is helpful to acquaint oneself with the structure of the CCE. Different series within the CCE listed filings for different categories of works; for example, books were separate from periodicals, which were separate from musical works. Due to irregularities regarding when a work was registered and when that registration was listed in the CCE, if one has a target year to check for filings (say, 1941), one should also look in the prior and following years’ volumes (1940 and 1942).

When searching Copyright Office records, especially the CCE, you should consider possible variants of the title, the author’s name, and the publisher’s name. An author or publisher could easily have changed how they listed themselves on copyright filings over time. When working in the CCE, you should also keep in mind the copyright terms that were in effect at the time. For example, in 1941 copyright lasted 28 years, after which the owner could renew for another 28 years. A work published in 1941 would probably have had its copyright renewed before 1969, so a diligent researcher would check the 1968 and 1969 CCE volumes.

Registering copyrights and recording transfers with the Copyright Office creates useful records, but many works still covered by copyright did not receive such helpful attention. Like all property, copyrights can be sold, given away, and inherited. Aside from looking in Copyright Office records, librarians may need to research the biographies of authors and the histories of possible corporate owners, such as publishers and employers.

**ORPHAN WORKS**

Even with the most diligent searching, it can sometimes be virtually impossible to determine a work’s copyright with certainty. Suppose an IR wants to digitize decades-old issues of the school newspaper. Any paperwork regarding the copyright in each story and issue has long been lost, and contact information for authors is stale and incomplete, making it very difficult to obtain permission. Works that have indeterminate ownership are called “orphan works.” For these works, IRs have to consider their risk tolerance. On the one hand, without clear determinations of whether a work is still copyrighted or who
owns it, an IR cannot obtain express permission to reproduce or distribute that work. On the other hand, if a copyright owner has not enforced or made public notice of their ownership, it seems relatively unlikely they will appear and complain of infringement of their copyright.

When managing liability risk with orphan works, conducting a thorough search for the copyright owner is crucial. First, the search may be fruitful and locate the owner. Second, failure to find an owner provides more confidence that the IR will not harm anyone by reproducing or distributing the work. While the copyright statutes do not provide express rules on an adequate search for owners of orphan works, there is some indirect guidance from the Copyright Office in a report on orphan works.\textsuperscript{18} Also, in 2018 Congress enacted the Music Modernization Act.\textsuperscript{19} This law established a process for obtaining a copyright exemption for the noncommercial use of sound recordings made before 1972 which are not currently being commercially exploited. The not commercially exploited status is similar in some ways to orphan works, and part of establishing that status is performing a good faith, reasonable search for the copyright owner. The Copyright Office has issued regulations specifying what constitutes a good faith, reasonable search, and this framework could be instructive when an IR librarian is deciding on appropriate steps when searching for works with unclear ownership.\textsuperscript{20}

Another source of guidance is the organizations for cultural heritage institutions.\textsuperscript{21} The Society of American Archivists published a statement of best practices for orphan works that contains suggestions for researching the copyright ownership of materials in archival collections.\textsuperscript{22} Even if some of the specific research steps are not applicable to some materials considered for an IR, the principles articulated in the statement are relevant: multiple legal rationales may apply to a specific project or use; materials should be used, not left unused because of their obscure ownership status; and common sense should apply. The common sense principle here is a type of cost/benefit analysis; the more prominent a work or its use, the more effort an archivist should expend searching for the owner, while less effort is required for obscure works. IRs would be prudent to create a documented policy regarding orphan works.

There is no mandated form for documenting a search, but it should be detailed enough to demonstrate that the library made a diligent, reasonable search in good faith for the copyright owner.

**COPYRIGHT EXCEPTIONS**

For works an IR determines are still under copyright and for which it cannot obtain permission, the IR will need to assess if an exception to copyright applies.
Returning to the example of digitizing a school newspaper, suppose that despite diligent efforts, permission cannot be secured for some stories. If the IR includes those stories in its collections, it will have to rely on a copyright exception.

Two main avenues are available here: the fair use exception under Section 107 of the Copyright Act, and the library exception under Section 108. When libraries fulfill interlibrary loan requests or permit researchers to make copies of works for their own use, they are using Section 108. Some libraries have begun using Section 108(h) to make preservation copies of works that are not commercially available and are within the last twenty years of their copyright terms. While this is an important option, many more works do not fall under Section 108(h)’s scope, so this chapter will focus on the fair use exception, which is invoked much more frequently.

Some rhetoric about copyright suggests that using an exception like fair use is using an underhanded technicality to deprive authors of their valuable rights. However, these exceptions are enacted in the same statutes that created the rights in the first place. Congress, when writing the statutes, decided that these exceptions were a worthwhile part of the copyright regime.

FAIR USE

The fair use exception is a broad exception for uses of copyrighted works that are socially beneficial, such as teaching, news reporting, research and scholarship, commentary, and criticism. The fair use exception permits the limited use of copyright-protected works without having to obtain permission from the rights holder. In evaluating whether any particular use of a work is protected by fair use, four factors must be analyzed: the purpose and character of the use (and whether it is of a commercial nature), the nature of the copyrighted work, the amount and substantiality of the portion used in relation to the copyrighted work as a whole, and the effect of the use upon the potential market value of the copyrighted work. These fair use factors are abstract, and whether a use is protected by fair use depends on the specific facts in each situation, so it is very unlikely one can have certainty about each use. Thus, it is understandable if a librarian would prefer not to rely on fair use and instead require that any materials added to the IR either be clearly in the public domain, be owned by the university, or have documented permission from the copyright holder. This risk-averse approach, however, will seriously constrain the IR from socially beneficial projects that are likely protected by law.

Instead of forgoing fair use, IRs can make thoughtful decisions about fair use that, while not indisputable, are likely to fulfill the IR’s mission with a tolerable level of risk. IRs will probably have occasions to invoke the fair use
exception, so they should have clear criteria that correspond to the fair use factors. They can also take guidance from publications by library organizations. An excellent example is the Association of Research Libraries’ “Code of Best Practices in Fair Use in Academic and Research Libraries.” Both academic librarians and attorneys specializing in copyright vetted this publication. A documented process of considering the fair use factors has two major benefits. First, forcing the options and considerations to be written down increases the chances that your ultimate decision will be thorough and measured. Second, the documentation will be evidence of the IR’s good faith application of copyright law in case any dispute arises. As you evaluate each fair use factor, you should think about how you may apply it to an IR project and how you record that consideration.

This discussion of fair use relies heavily on two major copyright cases that involved the mass digitization of books. Both cases were rooted in the Google book-scanning initiative. In 2002, Google began partnering with major academic libraries to scan the text of tens of millions of books. The digital versions were searchable, and small snippets were accessible to the public. Google also gave digital copies to the contributing libraries, which joined to form the HathiTrust Digital Library (HDL). HDL made the books searchable and some books accessible.

The Authors Guild and some authors sued Google and HathiTrust for copyright infringement. In separate decisions, the U.S. Court of Appeals for the Second Circuit held that Google and HathiTrust’s uses were protected by fair use. Since the HathiTrust case was decided first and was closely followed by the Google case, and because HDL is more similar to an IR than Google, we will focus on the HathiTrust court’s fair use analysis. A simplified, rough, synthesis of the fair use factors is that a use should be socially beneficial, not unduly substitute for the work in the market, and should not use more of the work than reasonably necessary for the use.

**Purpose of the Use**

The first fair use factor is the purpose and character of the use (for example, copying, excerpting, distributing, or displaying) one wants to make of a copyrighted work. This factor considers whether the use of the work is worthwhile and whether the use is what courts have termed “transformative”: is the use different from what the work was originally produced for? For example, HDL took digital copies of books and made them searchable so that researchers could more efficiently identify books relevant to their research questions. The original books had not been designed to be searched as a massive corpus, and so this function was new. HDL also made books accessible for researchers who had documented challenges in reading print. If HDL had simply made all the
books accessible for reading, then that would have substituted for the books and the fair use argument would have been weaker.

Whether a use is educational or commercial can be considered under this factor, and a company simply copying works for profit is unlikely to have a strong fair use case, but a use that is educational or nonprofit may not be as decisive as one might think. The educational purpose of an IR may be socially beneficial, but it is still important to consider whether a use will substitute for the original work. Digitizing an assigned textbook, for instance, is likely to be considered replacing the work and is not a good candidate for fair use.

A common purpose of an IR is to preserve an institution's intellectual output for the long term. HathiTrust claimed preservation as a justifying purpose, but the court found fair use without specifically deciding whether a preservation purpose was sufficient. Nonetheless, an IR could argue that the current market does not provide an adequate assurance of long-term institutional ownership, so merely creating digital copies for retention is not supplanting the works. Note here that the creation of the copies for preservation, and providing public access to those copies, are separate uses that each require their own independent fair use analysis.

**Nature of the Work**

Closely related to purpose is the second fair use factor, the nature of the copyrighted work that is being used. A use that does not replace or supplant the main use of the work favors a strong fair use case. To understand the original purpose of a work, we must examine its nature. Often this factor is described as considering whether a work is primarily factual or creative. This suggests that uses of nonfiction works are more likely to be covered by fair use than uses of works of fiction. This may make a small marginal difference, but the HathiTrust court did not place much emphasis on this factor. Both HDL and IRs contain both fiction and nonfiction works, so the degree to which a work is factual is unlikely to have much effect on the fair use case for a given use.

**Extent of the Use**

The third factor is the amount and extent of the work being used. It makes some intuitive sense that the more copies that are made or the greater the portion of the work being used, the less likely fair use will be cover the use. Such a strictly quantitative approach is too simplistic, however, and does not appropriately weigh the beneficial purposes of the use.

Rather, the amount of the work used should be commensurate with the use. For example, in HathiTrust, the HDL made complete copies of every book and made multiple backup copies. The HathiTrust court held that having
complete copies was reasonably necessary for the full-text search function. Likewise, many backup copies were reasonable, given the need for redundancy in case of equipment failures and maintenance. As the ARL’s “Best Practices” notes, judges often ask if the material taken is appropriate in kind and amount, considering the nature of the copyrighted work and the use.

Many libraries use quantitative limits, especially when fulfilling interlibrary loan requests. Whatever their merits in that setting, such limits are unlikely to make much sense for IRs. While interlibrary loans make copies in response to discrete requests by individual researchers, IRs probably need complete copies for their purposes, even if those complete copies are not shared immediately. When it comes to full-text searching or building a corpus for textual analysis, complete copies will again be justified. IRs should thoughtfully consider and document how much of each work and how many copies will be needed for their purposes. Documentation of the decision-making process, and avoiding using more of a work than is needed, will help this factor weigh toward a finding of fair use.

**Effect on the Market**

The final factor examines the effect of the use on the market for and the value of the work. If the purpose of the use is transformative and adds something the original work did not have, and the amount used is reasonable for the use, then it is unlikely that the use will damage the potential market for the copyrighted work. This factor probably favors uses of works that are not being commercially exploited, since there is less of a market to damage. The *HathiTrust* court noted that since the HDL’s full-text search function did not substitute for the books in the market, it did not harm the value of the works.

**CONCLUSION**

Copyright must be a major consideration when depositing materials in an IR and providing access to them, but it need not be a great obstacle. With responsible efforts to ascertain copyright ownership, educate faculty authors, and consider the fair use factors, IRs can effectively preserve and provide valuable access to scholarly works. Written copyright policies and procedures for obtaining permission and invoking fair use will increase librarians’ confidence in decisions involving copyright.
NOTES


27. Authors Guild v. HathiTrust, 755 F.3d 87 (2nd Cir. 2014).

28. Authors Guild v. Google, 804 F.3d 202 (2nd Cir. 2015).