Learning to Live with the Legal Risks of Digital Collections

Peter Hirtle
Alumni Fellow, Berkam Klein Center for Internet & Society
Harvard University
peter.hirtle@cornell.edu
My name is Peter

I am not a lawyer

I am an archivist. I study copyright.

And digital collections.
What I have learned

• Digitizing is powerful
• But there are risks
  • Copyright risks
  • Privacy risks
  • Loss of control
• To understand why, you have to know the law
Copyright Basics

• Most fixed works with a modicum of creativity are protected by copyright

• Copyrights last for a long time

https://copyright.cornell.edu/publicdomain
# Copyright owner’s rights

<table>
<thead>
<tr>
<th>Exclusive rights to …</th>
<th>In plain English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reproduce</td>
<td>Make copies</td>
</tr>
<tr>
<td>Distribute</td>
<td>Sell; make available for downloading on your web site</td>
</tr>
<tr>
<td>Create derivative works</td>
<td>Make new work from an existing work; screenplay from novel; translation</td>
</tr>
<tr>
<td>Display the work publicly</td>
<td>Hang a painting in a gallery; display on the web</td>
</tr>
<tr>
<td>Perform the work publicly</td>
<td>Theatrical performance; show a movie; stream on the web</td>
</tr>
</tbody>
</table>
Privacy tort: publication of private facts

- **True information**
- about a person’s private life
- that would be highly offensive to a reasonable person, and
- not of legitimate public concern
  - Sexual activity
  - Health – including mental health, addiction
  - Economic status
Our problems

- Digitization can infringe on copyright owner rights
  - Reproduction
  - Distribution
  - Possibly public display
- Broad public access may impinge on privacy
  - No more “security through obscurity”
Locating copyright owners is hard

• Example: Thomas Watson papers
  • 8,434 pieces of correspondence; 3,304 names; 2,671 still ©
    • $8,000 to figure this out
  • 450 hours total, over 9 months
  • Result: dependable contact information for 4 authors

Maggie Dickson, *Due Diligence, Futile Effort: Copyright and Digitization of the Thomas E. Watson Papers*, 73 American Archivist 626 (2010)
http://archivists.metapress.com/content/16rh811120280434
Who owns the copyright?

The Children of (Joseph) Stephen Hirtle, ca. 1920
International liability: even for public domain

- Copyright rules vary by nation
- Internet distribution may make you liable abroad
- Example: Wikimedia Foundation v. Reiss Engelhorn Museum
Conclusion: Digitization = Legal Risk

• But we live with risk all the time:
  • Theft
  • Fire and other physical thefts
  • Accidents to staff and patrons
  • Institutional embarrassment

• Solution: policies and practices that limit the potential for harm
Elements that minimize risk

• Some material is in the public domain

• Some in-copyright works published between 1924 and 1944 may go online: 17 U.S.C. § 108(h)

• Potential monetary damages for unregistered material are low
  • Registered: up to $150,000 per work
  • Unregistered: actual economic harm

• Registered damages mostly waived for a “fair use”: 17 U.S.C. § 504(c)(2)
  • Need “reasonable grounds for believing that [the] use… was a fair use”

https://www.law.cornell.edu/uscode/text/17/504
Is digitizing a collection fair use?

• No one knows..

• Code of Best Practices in Fair Use for Academic and Research Libraries
  • Principle 4: Creating digital collections of archival and special collections materials
  • Digital collections may serve a different purpose than the originals, making them transformative
  • See the limitations

https://www.arl.org/focus-areas/copyright-ip/fair-use/code-of-best-practices
Also excellent: AAMD guidelines for museums (includes archives and other special collections)

More components for risk mitigation

• Takedown policy on request
• Example: Digital Commonwealth
• Copyright Holders
  “Digital Commonwealth and its members welcome feedback from any copyright owners who are not properly identified on this site so that the necessary corrections may be made. If you have any additional information about the images or would like to suggest a correction, please contact us.”

https://www.digitalcommonwealth.org/copyright
Clear statements about user responsibility

• Digital Commonwealth:
  “It is your obligation to determine and follow all copyright or use restrictions of the material presented here.”

• Library of Congress:
  “Responsibility for making an independent legal assessment of an item and securing any necessary permissions ultimately rests with persons desiring to use the item.”

https://www.digitalcommonwealth.org/copyright
https://www.loc.gov/collections/rosa-parks-papers/about-this-collection/rights-and-access/
Limitations on subsequent use

- **Digital Commonwealth:**
  “Digital Commonwealth provides access to copyrighted materials strictly for noncommercial educational and research purposes and places no restrictions on public domain material.”

- **Library of Congress:**
  “The Library of Congress provides access to manuscripts at the Library of Congress for educational and research purposes and makes no warranty with regard to their use for other purposes.”

Different collections = different risks

• Some content may present more risk
  • Newer versus older
  • Potentially highly commercial
  • Existence of potentially litigious owners

• Institutions will vary on the desired level of risk-avoidance
## Risk Factor Assessment for Copyrighted Tapes at PRA

This chart can be used to determine the risk factors for copyright status of tapes at the Pacifica Radio Archives. PRA staff must weigh the various factors together before determining whether it is appropriate to disseminate a tape for broadcast, re-transmission, etc.

<table>
<thead>
<tr>
<th>Risk factor</th>
<th>Low risk</th>
<th>Medium risk</th>
<th>High risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Copyright/distribution notice on tape?</td>
<td>No notation or tape belongs to &quot;Pacifica Tape Library&quot; or a variant thereof</td>
<td>Notice on label that says PRA has &quot;limited&quot; or &quot;restricted&quot; distribution rights to the program</td>
<td>Tape label explicitly states that we have no distribution rights to the program</td>
</tr>
<tr>
<td>Producer credit?</td>
<td>Credit exists for a producer who has consented for PRA to distribute his or her work</td>
<td>No producer credit or PRA cannot locate producer after good faith search</td>
<td>Credit exists for a producer who does not want PRA distributing their work</td>
</tr>
<tr>
<td>Music used in program?</td>
<td>No music used in program or recording falls in public domain</td>
<td>Contains small amount (less than half) of music not in public domain (fair use)</td>
<td>Contains large amount (more than half) of music not in public domain (not fair usage)</td>
</tr>
<tr>
<td>Literature in program?</td>
<td>No literature read or performed or excerpt used falls in public domain</td>
<td>Contains small amount (less than half) of a total work not in public domain (fair use)</td>
<td>Contains large amount (more than half) of an excerpt in the new recording (not fair usage)</td>
</tr>
<tr>
<td>Excerpts from other broadcasts/radio programs/TV shows?</td>
<td>No such excerpts in program, or it can be demonstrated that producer was granted usage of that excerpt by licensee</td>
<td>Contains small amount (less than half) of an older broadcast in the new recording (fair use)</td>
<td>Contains large amount (more than half) of an excerpt in the new recording (not fair usage)</td>
</tr>
<tr>
<td>Station ID announced on recording?</td>
<td>Pacifica ID (i.e. KPFA, KPKN, WBAI, KPFT, WPFH, Pacifica National Broadcast)</td>
<td>No ID announced</td>
<td>Non-Pacifica ID announced</td>
</tr>
<tr>
<td>Known litigious entities associated with tape?</td>
<td>None</td>
<td>Risk factor of personalities unknown</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Examples of risk minimization strategies

• One extreme: the Internet Archive
  • Regularly posts copyright content
  • Follows the Oakland Archive Policy (2002) for takedowns:
    http://groups.ischool.berkeley.edu/archive/aps/removal-policy
  • Redefining what is considered fair use
Example 1: The Pulp Magazine “Archive”

The Pulp Magazine Archive

Pulp magazines (often referred to as "the pulps"), also collectively known as pulp fiction, refers to inexpensive fiction magazines published from 1896 through the 1950s. The typical pulp magazine was seven inches wide by ten inches high, half an inch thick, and 128 pages.

About

Collection

2,074 results

Sort by: Views, Title, Date Published, Creator

Science fiction

Metadata

Text contents

Part of

Additional Collections

Media Type

Collections

Texts

Year

2016: 52

2013: 1
additional mailing offices. Copyright © 1969 by Universal Publishing & Distributing Corporation under International, Universal and Pan-American Copyright Conventions. All rights reserved. The publisher assumes no responsibility for unsolicited material. All stories printed in this magazine are fiction. Any similarity between characters and actual persons is coincidental. Title registered U.S. Patent Office. Printed in U.S.A. The Company also publishes Award Books, Novel Books, Teenage Books (United Kingdom), Vocational Guidance Manuals, Golf Magazines, Golfdom, Ski, Ski Business, Ski Area Management, Home Garden, The Family Handyman.
Example 2: The Great 78 Project

https://great78.archive.org/
Kissa Me Baby
by Ray Charles

Publication date 1952-02
Topics 78rpm, Popular Music
Publisher Swing Time
Digitizing sponsor Kahle/Austin Foundation
Contributor Internet Archive
Language English

508 Views
1 Review

Download Options

https://archive.org/details/78_kissa-me-baby_ray-charles_gbia0060584b
Example 3: Controlled Digital Lending (CDL)

- CDL: limited loan of digitized books to patrons
  - See https://controlleddigitallending.org/
- Implemented in IA’s Open Library

https://openlibrary.org/
Boston Public and CDL

https://archive.org/details/wastedtalesofgen00judg
Authors Guild and Society of Authors Alleged Copyright Infringement by the Internet Archive

In News by Porter Anderson / January 18, 2019 / 3 Comments

Should the Internet Archive not comply by February 1, the Society of Authors says in its letter, “we will have to consider legal action on behalf of our members to prevent this practice. If that becomes necessary we shall be seeking damages for copyright infringement and payment of all legal costs.”
Why the Internet Archive matters

• IA goes farther than most of us
• To date, it has not been challenged
• Takedown policy is effective
• No commercial exploitation is likely key
Example 2: Fair Use at the Library of Congress

https://www.loc.gov/collections/rosa-parks-papers/
6211 Mullen Rd. SE
Lacey, WA 98503

January 14, 2005

Dear Mrs. Rosa Parks,

I am happy to say you are a great woman. You are a role model to me and many other kids. I am sad that they hurt equal people just because they had different skin colors. I could help the world be more equal by helping people and stand up for everybody’s rights.

Sincerely,

Krystalyn Cox
3rd grade
Lakes Elementary School
Example 3: “Virtual Reading Room”
American Archive of Public Broadcasting

http://americanarchive.org

Harvard Book Store; WGBH Forum Network; Nick Bunker: The Mayflower Pilgrims and Their World

Date: 2010-06-26

Journalist and historian Nick Bunker steps into a long past world with his book, Making Haste from Babylon: The Mayflower Pilgrims and Their World, A New History. At the end of 1618, a blazing green star soared across the night sky over the northern hemisphere. From the Philippines to the Arctic, the comet became a sensation and a symbol, a
Welcome to the American Archive of Public Broadcasting Online Reading Room (ORR). To access media in the ORR, please review and agree to the ORR Rules of Use.

The American Archive of Public Broadcasting (the "AAPB") is a project managed collaboratively by the WGBH Educational Foundation and the Library of Congress ("we" or "us"). The AAPB seeks to preserve and make accessible significant historical content created by public media, and the AAPB’s Online Reading Room (the "ORR") plays a central role in our mission. In the ORR, you can view and listen to thousands of hours of digitized archival materials free of charge. We make the ORR and all materials available under terms that support and extend the educational/encyclopedia functions of the materials, as well as the mission of the AAPB and its partners. By accessing the ORR, you agree to abide by these terms.

I AGREE
UC Irvine: slightly more restrictive
Most restrictive: No online access
Salman Rushdie papers at Emory

• Locked down computers with emulated environment
• Pre-screened content
• No access outside of reading room
Summary: Don’t worry, act responsibly

- Remember your mission!
- Understand – and manage – the risks
- Commercial use increases the risk you face
- Chose the level of access appropriate for the content and your institution
- Act professionally and thoughtfully
- If in doubt, consult with your legal advisors