

Books, Computers, and the Law

James Grimmelman

Archivists Round Table of Metropolitan New York

12 January 2010

The road here

The *first* book scanning revolution

- Consider the printing press
- First use: printization
- Transforms our relationship to text
- Transforms our preservation strategies

Copyright 101

- Original expression is copyrightable
- Specific exclusive rights
 - incl. “to reproduce the work in copies”
- Fair use defense
- Life + 70 = 1923

Copyright 102: orphan works

- Some books are in print and being actively managed: yay!
- Some books are now out of copyright: yay!
- But many millions are in copyright but out of print ...
 - And many of those have unfindable copyright owners
- These books aren't doing anybody any good
 - And their status is a failure of the copyright system

Google

- “To organize the world’s information and make it universally accessible and useful.”
- “Don’t be evil.”

Google's scanners

- Partner with libraries
- Nondestructive photography
- OCR
- Full-text search, text+image display
- Public-domain books

The Rubicon

- Google starts scanning in-© books
 - Full-text search
 - Shows “snippets,” links to bookstore
 - Eventually, lets © owners opt out
- Looks a lot like the Web, where Google has been winning its fair-use fights

Outbreak of hostilities

- Late 2005, authors and publishers sue
 - Opt-out isn't acceptable to them
- Potential damages astronomical
- Very important: they sue in a class action

The settlement

Deal of the century

- Google pays \$60/book to settle past claims
- Going forward:
 - Scanning and search still on
 - Opt-out requests still honored
 - Snippets off (by default) for in-© books

Revenue models

- Preview up to 20% online, with advertising
- Consumer Purchase of online e-books
- Institutional Subscription for libraries, schools, etc.
 - Public Access: one free terminal
- Research Corpus for the machines

Handling the money

- © owners can set their own prices for sales
- Google keeps 37% of all the revenues
- The 63% goes to a new Registry, which pays © owners and coordinates with Google
 - Authors and publishers split the 63%
 - Unclaimed funds eventually redistributed

Claiming books

- Online database for © owners to sign up
- Only as good as its metadata
 - And libraries have a lot of bad metadata
- Isn't this just another opt-out system?
 - And there are some © owners we expect not to show up

The settlement game

- Original opt-out deadline: May 5, 2009
 - Pushed back to September 4
 - Fairness hearing scheduled for October 7
- DOJ files a skeptical but optimistic “statement of interest”
 - Parties ask for a chance to renegotiate
 - Amended settlement proffered late on November 13

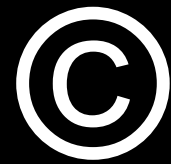
Current timeline

- Opt-out / objection / amicus deadline: January 28
- DoJ deadline: February 4
- Parties file responses: February 11
- Fairness hearing: February 18
- Judge Chin rules: ?
- Case finally resolved: ???

Objections

Civil procedure

- How good was the notice?
 - Especially outside of the U.S.
- Who are these authors and publishers?
 - Do they speak for, e.g. academic authors, who want open access rather than \$\$\$?
- What kind of a class action is this ...



- “© is opt-in” is dogma in most of the world
 - (Perhaps not the best rule, but ...)
 - Technically, © owners are “agreeing”
 - But U.S. class action law is *sui generis*
- Lots of controversy over publishing-industry details

Information policy

- Huge centralization of books in Google
- Libraries burnt by journal pricing
- Reader privacy
- Equitable access, pro and con
- Is Google a library?

Antitrust

- Coordinated algorithmic pricing (but constrained to look like competition)
- Blanket pricing for subscriptions (but looks a lot like BMI/ASCAP)
- What about them orphans?
 - Good luck trying to compete!

The heart of the deal

- The settlement makes many orphaned books available again *because it's opt-out*
 - If you like books, that's good
 - But is this a legitimate use of a class action?

Closing thought

- 0 and ∞ make sense, but 1?
 - If we think orphan works are valueless, then they should be public-domain
 - If we think these rights are worth respecting, then they should be respected
- I'm still looking for the limiting principle

Questions?