

Canaries in the **Text** Mine: Fair Use **Rights** and Text+Data Mining with **Licensed** Content

A SPARC Webinar
Brandon Butler
Director, Information Policy
University of Virginia Library
May 23, 2016

Roadmap

- Fair use and TDM: Introducing “non-consumptive” use
- The **limits** of fair use for TDM
- Licenses and TDM: **Three key questions**

Disclaimer!

- I am a lawyer, but I am not *your* lawyer. (I'm not even *UVA's* lawyer!)
- This *isn't* legal advice.
- There is *no single right answer* to these questions. Context matters, and libraries have many legitimate priorities that may lead to a diversity of equally legitimate strategies. What's key is making *informed decisions* grounded in *core values*.

Fair use and TDM:
Introducing “non-
consumptive” use

What is fair use?

- Flexible, open-ended user's right to do things with © works that implicate owner's rights (reproduction, distribution, public display, etc.)
- Statutory factors - purpose, nature of work, amount used, market effect
- Transformative turn - courts focus on new, non-superseding purpose as key to fair use

What is “non-consumptive” use?

- Term emerged from Google Books settlement
- Processing of copyrighted work by computers, “behind the curtain”
- “Non-consumptive” b/c work is never “consumed” by a human for its expressive content
- Only way to discover important, *unprotected* facts about works, corpora
 - United States “is” vs. “are,” e.g.

Why is “non-consumptive use” fair?

- New purpose
- Non-substitutional
- Access to unprotected facts
- Promotes purpose of © - access to information, increase of knowledge, “Progress”

How do we know it's fair?

- Embraced by **two leading copyright Circuit Courts** (9th Cir (“Hollywood Circuit”) and 2nd Cir (NYC Circuit)); **SCOTUS** denied cert. in Google Books
- **Cases** (a non-exhaustive list): Kelly v. ArribaSoft, Perfect10 v. Google, Field v. Google, A.V. v. iParadigms, Authors Guild v. HathiTrust, Authors Guild v. Google, Fox v. TV Eyes.

The **limits** of fair use for TDM

Look to the cases

- Search engine cases
 - Kelly v. ArribaSoft; Perfect 10; Field v. Google
- Book scanning cases
 - Google Books and HathiTrust
- Video scanning case - TVEye

Two takeaways

- **Tailor access** to purpose: how much of the original does a researcher reasonably need to see in order to conduct non-consumptive research?
 - Could be tricky; humanities scholars may need a mix of close and “distant” reading
- Avoid **substitution**
 - Does your TDM facility provide public access sufficient to satisfy the needs of an ordinary “consumer” of the article?

Licenses and TDM: Three Key Questions

Do you *need* a license to do TDM?

- All else equal, **NO!** TDM requires actions that implicate copyright (primarily copying works into a database for processing), but those actions are **protected by fair use**.
- DO NOT PAY for the **bare legal right** to do TDM; that is paying for **something you already have**.
- *Once you have lawful access/possession of a copy/corpus, TDM is a protected fair use.*
- **Analogy:** Public Performance Rights for classroom videos

Is your license a barrier to TDM?

- You can't do TDM without lawful access to the work; licensor can leverage access controls to limit lawful use.
- Limits on scale of access
- Limits on purpose of access
- Limits on downstream use
- Non-copyright legal claims: breach of contract, Computer Fraud and Abuse Act (CFAA), trespass to chattels (yes, really).
- Solutions: Get explicit fair use acknowledgment; get broad, general grants of access; push back on purpose or scale limitations; live within your license

Do you want specific provisions to facilitate TDM?

- Quality metadata
- Quality OCR/scans
- More download bandwidth, crawling directly on vendor's server, access to infrastructure, other services, etc.

Big picture, long term, political issues

- Who controls access, controls the world
 - E.g., Google agreements govern HTRC uses of PD works; publishers can thwart fair uses by restricting access
 - SSRN and Elsevier
- Who controls infrastructure “”
 - Much of the value of TDM is in services, technology
- Don't pay for bare rights
- Consider the value proposition of publisher-run TDM services

Thanks!

Now let's discuss!