

Dan Cohen's Digital Humanities Blog » Blog Archive » Clifford Lynch And Jonathan Band On Google Book Search

The topic for the November 2005 Washington DC Area Forum on Technology and the Humanities focused on “Massive Digitization Programs and Their Long-Term Implications: Google Print, the Open Content Alliance, and Related Developments.” The two speakers at the forum, Clifford Lynch and Jonathan Band, are among the most intelligent and thought-provoking commentators on the significance of Google's Book Search project (formerly known as Google Print, with the Google Print Library Project being the company's attempt to digitize millions of books at the University of Michigan, Stanford, Harvard, Oxford, and the New York Public Library). These are my notes from the forum, highlighting not the basics of the project, which have been covered well in the mainstream media, but angles and points that may interest the readers of this blog.

Clifford Lynch has been the Director of the Coalition for Networked Information (CNI) since July 1997. CNI, jointly sponsored by the Association of Research Libraries and Educause, includes about 200 member organizations concerned with the use of information technology and networked information to enhance scholarship and intellectual productivity. Prior to joining CNI, Lynch spent 18 years at the University of California Office of the President, the last 10 as Director of Library Automation. Lynch, who holds a Ph.D. in Computer Science from the University of California, Berkeley, is an adjunct professor at Berkeley's School of Information Management and Systems.

Jonathan Band is a Washington-based attorney who helps shape the laws governing intellectual property and the Internet through a combination of legislative and appellate advocacy. He has represented library and technology clients with respect to the drafting of the Digital

Millennium Copyright Act (DMCA), database protection legislation, and other statutes relating to copyrights, spam, cybersecurity, and indecency. He received his BA from Harvard College and his JD from Yale Law School. He worked in the Washington, D.C. office of Morrison & Foerster for nearly 20 years before opening his own law firm earlier this year.

Clifford Lynch

- one of things that have made conversion of back runs of journals easy is the concentration of copyright in the journal owners, rather than the writers of articles
- contrast this with books, where copyrights are much more elusive
- strange that the university presses of these same univs. in the google print library project were among the first complainers about the project
- there's a lot more to the availability of out of copyright material than copyright law—for instance, look at the policies of museums, which don't let you take photographs of their out of copyright paintings
- same thing will likely happen with google print
- while there has been a lot of press about the dynamic action plan for european digitization, it is probably a plan w/o a budget
- important to remember that there has been a string of visionary literature—e.g., H.G. Wells's "worldbrain"—promoting making the world's knowledge accessible to everyone—knowledge's power to make people's lives better—not a commercial view—this feeling was also there at the beginning of the Internet
- legal justifications have been made for policy decisions that are really bad
- large scale open access corpora are now showing great value, using data mining applications: see the work of the intelligence community, pharmaceutical industry—will the humanities follow with these large digitization projects
- we are entering an era that will give new value to ontologies, gazetteers, etc., to aid in searching large corpora

- if google loses this case, search engines might be outlawed [Lawrence Lessig makes this point on his blog too —DC]
- because of insane copyright law like sonny bono act there might be a bifurcation of the world into the digitized world of pre-1923 and the copyrighted, gated post-1923 world

Jonathan Band

- fair use is at base about economics and morality—thus the cases (authors, publishers) against google are interesting cases in a broad social sense, not just pure law
- only 20% of the books being digitized are out of copyright (approx.)
- for certain works, like a dictionary, where even a snippet would have an economic impact on the copyright holder, google will probably not make even a snippet available
- copyright owners say copyright is opt-in, not opt-out (as Google is making it in their program)—it seems dumb, but this is a big legal issue for these cases
- owners are correct that copyright is normally an opt-in experience—the owner must be contacted first before you make a use of their work, *except* when it's fair use—then you don't need to ask
- thus the case will really be about fair use
- key precedent: *kelly vs. arribasoft*: image search, found in favor of the search engine; kelly was a cantankerous photographer of the West who posted his photos on his website but didn't want them copied by arribasoft (2 years ago; ended in 9th circuit); court found that search engine was a transformative use and useful for the public, even though it's commercial use; court couldn't find any negative economic impact on the market for kelly's work [this case is covered in [chapter 7 of *Digital History*](#)^[1] —DC]
- google's case compares very favorably with arribasoft
- publishers have weaker case because they are now saying that putting something on the web means that you're giving an implied license to copy (no implied license for books)—but they've argued before that copyright applies just as strongly on the web

- bot exclusion headers (robots.txt)—respected by search engines but that sounds like opt-out, not opt-in—so publishers also probably shouldn't be pointing to that in their case
- publishers are also pointing to the google program for publishers, in which publishers allow google to scan their books and then they share in revenues—publishers are saying that the google library program is undermining this market, where publishers license their material; transaction costs of setting up a similar program for library books would be enormous—indeed it can't be done: google is probably spending \$750 million to scan 30 mil. books (at \$25/bk); it would probably cost \$1000/bk if you had to clear rights for scanning; no one would ever be able to pay for clearing rights like this, so what google is doing is broad and shallow vs. deep but narrow, which is what you could do if you cleared rights—many of these other digitization projects (e.g., Microsoft) are only doing 100K books at most
- if google doesn't succeed at this project, no one else will be able to do it—so if we agree that this book search project is a useful thing, then as a social matter Google should be allowed to do it under fair use
- what's the cost to the authors other than a little loss of control?

This entry was posted on Monday, November 28th, 2005 at 10:38 pm and is filed under [Books](#)^[2], [Copyright](#)^[3], [Digitization](#)^[4], [Google](#)^[5], [Search](#)^[6]. You can follow any responses to this entry through the [RSS 2.0](#)^[7] feed. You can [leave a response](#)^[8], or [trackback](#)^[9] from your own site.

References

1. [^ chapter 7 of Digital History](#) (chnm.gmu.edu)
2. [^ View all posts in Books](#) (www.dancohen.org)
3. [^ View all posts in Copyright](#) (www.dancohen.org)
4. [^ View all posts in Digitization](#) (www.dancohen.org)

5. [^ View all posts in Google](#) (www.dancohen.org)
6. [^ View all posts in Search](#) (www.dancohen.org)
7. [^ RSS 2.0](#) (www.dancohen.org)
8. [^ leave a response](#) (www.dancohen.org)
9. [^ trackback](#) (www.dancohen.org)

Excerpted from *Dan Cohen's Digital Humanities Blog » Blog Archive » Clifford Lynch and Jonathan Band on Google Book Search*

<http://www.dancohen.org/2005/11/28/clifford-lynch-and-jonathan-band-on-google-book-search/>

READABILITY — An Arc90 Laboratory Experiment

<http://lab.arc90.com/experiments/readability>