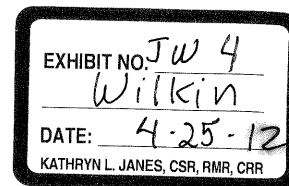# EXHIBIT 97

blyberg.net

A library-geek blog

John Wilkin talks Google Print & Digitization

One of the many good things about being a PL in Ann Arbor, is the close proximity to the University of Michigan. Occasionally, we are lucky enough to benefit from their enormous brain trust and perhaps experience some intellectual osmosis.

AADL shut down today for staff day, an annual staff training day that sometimes involves talks by guest speakers. We had the opportunity to hear from John Wilkin this year. John's official title is Interim Associate Director for Digital Library Services at the University of Michigan. We works with Google as head of the Humanities Text Initiative at the university library. He spoke for about 45 minutes on the UM-Google digitization deal. He shared some insights on Google Print from UM's point of view, how UM benefits from Google Print, and what they plan on doing with their access to the data. It seems that much of the attention focused on GP is directed at Google, so its interesting to hear about GP's implications from someone at UM.

UM-Google Digitization Deal
What was UM's deal with Google, Why and how did it happen? (not his presentation title)

How did we get there?
UM started putting text online during the 80's, so there was some precedent. Also, the fact that Larry Page is an alumni of the College of Engineering doesn't hurt. The social networks were in place at UM to pursue the project.

What will be digitized?
Any bound material at the UM libraries (with some abstaining libraries) will be digitized. This amounts to about 7 million volumes. The process of digitization must not damage the material, it must be high quality, and during the process, the materials must be very well cared for.

Legal issues? (No! Why would there be?)
There are many legal issues to be worked out, but as part of the deal, UM is indemnified by Google against any legal action.
UM has agreed to be responsible for moving the material to Google. Material is in Google's possession for about 5-7 days. It is kept on carts and they don't really worry about the order in which material is digitized. UM gets a copy of every image scanned (and I assume, all associated data, metadata, etc)

Where UM and Google part
UM believes that Google should not be tasked with the long-term care of digital material. Google is ultimately a business, whereas UM is an educational institution, they have completely different motives to possess the data. Therefore, UM wants to make sure that there are no constraints on the material and that the services that are associated with it are appropriate.
UM is also putting pressure on Google to make the material accessible to the visually impaired.
UM believes in "ubiquitous access".

Why would google do all this?
Google benefits by gaining user-base, name recognition — more power!

About the files
600 dpi ITU G4 (Bitonal)
300 dpi JPEG-2000
Standardized files / standard formats
OCR
Checksums
Production notes

What does the technology look like?
He talked a little bit about the technology. He stressed that while it's highly confidential, new technology they've developed, it really doesn't matter what it looks like (he flashed a picture of an old Tandy I think it was). What is important is that the technology does a very good job. He said that items are very well cared for: they are not disbound and the process of digitizing the material does not damage it. He also said that the process produces quality images and the product adheres to appropriate rights. It's also very fast.
Here's something interesting though. They group items by size before they scan them, the scan order is determined by those sizes. I guess we can infer that whatever the machine is, it needs to be adjusted to material size.

What sort of differences will UM's system have?
Even though the digitized content will be the same as GP, UM will offer different types of services because the UM audience is different. UM plans to offer more flexible displays, more powerful citation tools, power searches. UM is definitely not trying to compete with Google. They envision a set of data mining tools that will allow them to do things like automatic translation and content extraction. UM's system will be geared toward research. Google will not show the OCR stuff, UM will.

IMPORTANT! UM will be responsible for the permanent repository of the content.

What are the transformative implications?
Broad, efficient, democratizing access. UM envisions access as a driver for basic infrastructure development in other parts of the world, especially third world countries. They see profound implications in science, medicine, educational networks.
John sees digitization as a catalyst to exaggerate and resolve intellectual property issues. He pointed to "orphan copyright" issues as a prime example. He said that current copyright law is keeping material locked by dead copyright holders (hence orphan copyright). He hopes that the impetus behind GP will help legislation next year that will protect libraries from orphan copyright.
He also pointed out that many publishers are now taking the stance that the mere act of copying something is a violation of CR, despite what you do with it. He stated that projects like GP may help stop publishers from pushing their licensing business model, because if they start licensing material, it could be devastating to libraries.
He hopes that digitization could lead to a cooperative, "universal", digital library. Some items are unique to UM, such as it is at any academic institution. It would be well worth it to provide those materials ubiquitously. A digital library would exacerbate the paradox of "library as place". Libraries with digital collections, or access to digital collections can move a lot of unused materials out, transform physical library space to adapt to new emerging roles.
Libraries can facilitate "specialization" (ceding "generalist" role to google).

Some mentions from Q&A

Government documents important. The Tulane Gvt. document collection was badly damaged after Katrina. If it had been done digitalized, we'd still have access to those materials.

OCR being used to drive txt searching? How is OCR being corrected? It isn't. Google is concentrating on

improving their engineering. The best way to improve it is to use voting systems. GP uses 6 voting engines. Can't do anything manually given the volume of material.

how will Google make money?
Selling ads — long term strategy is eyeballs.. more people to site the more popular/powerful they become. UM stipulated that google could not ever charge for the service.

Why has so little been digitized at other libraries? Harvard said to UM: (paraphrase) You haven't done much deep thinking about CR issues. UM decided to stay focused on principals. They don't want to break any laws or hurt any books. UM had signed contract months before. Other institutions did days before the public announcement.

Yes, it will be able to link to catalogs (such as AADL).

OCLC to digital record? Consensus is to represent a new digitization as a new work. UM believes thats a bad way to do it.. OCLC wants to do it that way.

Yes, they do scan foreign language material.

Google's technology doesn't handle maps or oversized inserts.

It was a low key talk, given the venue, so John abridged a fair amount of his presentation. Regardless, I thought it was an intriguing insight into what's happening just a few blocks up the road. I really appreciate that he took the time to come over for a few hours and talk to us after bearing witness to the morning's earlier antics.

Google Print's implications really are so broad and so deep that we're not going to know the full extent its impact will have until we're looking back on it with a fair amount of history to put it in context. What a fabulous endeavor. I wish them the best of luck.

About this entry

You're currently reading "John Wilkin talks Google Print & Digitization," an entry on blyberg.net

Published:
      11.11.05 / 7pm

Category:
      Libraries

---

2 Comments

Jump to comment form | comments rss [?] | trackback uri [?]

1. *eli 11.12.05 / 9am*

   John Wilkin was a fantastic presenter, and did an excellent job adapting his presentation for a (primarily) non-technical audience. What's most amazing to me is that an organization of the size (and therefore inertia) of U of M had the cojones to essentially give the finger to the publishing industry and