

# COPYRIGHT ISSUES POSED BY MASS DIGITIZATION

Pamela Samuelson, Berkeley Law

IVIR Summer Course

July 11, 2012

# OVERVIEW

- Why mass digitize?
- Is mass digitization a fair use under US law?
  - *Authors Guild v. Google*
  - *Authors Guild v. Hathitrust*
- Various approaches to addressing orphan work problem for those wishing to digitize in-© works
  - Proposed EU Directive
  - French legislation
  - Other approaches
- Other mass digitization © & legal issues

# WHY LIBRARIES & ARCHIVES WANT TO MASS-DIGITIZE

- To preserve collections, cultural heritage
- To manage collections more efficiently
- To justify “deaccessioning”
- To provide as broad public access to works in their collections as possible
- To have a corpus on which to do data-mining research (e.g., trace thinker’s influence over time)
- To improve access for print-disabled folks

# GOOGLE'S MOTIVATIONS

- To make indexes of book contents
- To make computational uses (e.g., improve search technologies, automated translation tools) of corpus
  - Documents as bags of words, as databases
- To provide snippets in response to search queries, attract users because of better ability to provide answers to “tail queries”
- To build services for processing texts
- To develop corpus of books that could be licensed?

# MAIN IMPEDIMENTS TO MASS DIGITIZATION

- Most institutions with interests in mass-digitizing their collections (e.g., libraries, museums, archives) have neither the financial resources or technological expertise to do this
- Copyright is biggest obstacle
  - To be safe in EU, cannot assume works are in public domain (PD) unless predate 1870
  - To be safe in US, only pre-1923 works are in PD
  - Extremely expensive to do rights clearances per work
  - Not possible to clear rights to “orphan works”
- Also expensive to sort collections by what is in PD & what is not

# G'S DEAL WITH LIBRARIES

- Google had the vision for GBS, the technology & the financial resources to digitize books
- Major research libraries had the books & the desire to digitize the books, but not the resources to do this
  - They were also more cautious than G about ©
  - 11<sup>th</sup> A immunity for state universities (no \$ damages), so no wonder state universities offered Ms of books first
- Google was willing to indemnify libraries & give them Library Digital Copies (LDCs) of books from their collections that G scanned
- Several major research libraries signed up for this

# # OF MASS DIGITIZATIONS

- US Library of Congress “American Memory” project (9M items online)
- Europeana
- French Gallica project
- National Digital Library of China
- British Digital Library
- National Library of Sweden
- Japanese Diet Library
- Most of these digitize only public domain works

# GBS CORPUS

- GBS Project began in 2002
- @2M books in GBS corpus were scanned with authorization from publishers under the Google Partner Program (GPP)
- Corpus now includes about 18 million books scanned from university research library collections
  - @ 3M in public domain
- Varying estimates of eventual size of GBS corpus
  - Ranging from @20M to 174M



# GBS ACCESS

- For @3M books in the public domain, G makes whole book available for download in pdf (with G's watermark)
- As to books in ©, G now makes “snippets” available
  - It has not run ads vs. the snippets so far, but wants to
  - It provides links so users can buy pertinent books from Amazon or find them in libraries
- G says it is willing to remove book of GBS corpus if © owner so requests
- For @2M in-print books in GPP, © owners can negotiate with G about how much of their books to make available, with revenue-sharing arrangements

# AUTHORS GUILD v. GOOGLE

- In Sept. 2005, AG + 3 members sued G for © infringement for scanning books, storing and processing the scanned books, & displaying snippets
- AG brought lawsuit as a class action on behalf of all rights holders whose books were scanned from U Michigan library, now all libraries
  - AG had theory that authors owned © in e-books
- G's main defense has been fair use
- Similar lawsuit brought by 5 trade publishers soon thereafter, not initially a class action

# MOTIVATIONS TO SETTLE

- Litigation is expensive, takes years to resolve definitively
- Outcome in doubt because dispute over fair use
- Also unclear whether class could be certified
  - if class not certified, G would take objecting authors' books out of the repository; exposure much smaller than with class action
- G faced big damage exposure, possible injunction vs. scanning, & order to destroy its database of in-© works
- G had better technology & ideas about how to create new markets for books in digital environment than AAP, AG
- Settlement created an opportunity for a “win-win” if G willing to share revenue streams with authors/pubrs
  - Oh, and incidentally would give G a license to all books in © that none of its competitors could get, cartel pricing for pubrs?

# CORE OF GBSS

- G to get license to scan, make non-display uses of all in-© books; libraries can keep LDCs
- G able to commercialize all out-of-commerce works:
  - Institutional subscription database (ISD)
  - Sale of books “in the cloud”
  - Up to 20% of contents viewable in response to search query; OK to run ads alongside
  - 63% of revenues would go to Book Rights Registry for distribution to rights holders (RH)
  - RH can opt-out of commercialization
- Quite similar to extended collective license

# BENEFITS OF GBSS

- Would have removed a dark cloud of liability from the heads of G and cooperating libraries
- Would have vastly enhanced public access to books
- Revenues would have flowed to authors and publishers who registered with the BRR or join GPP
- Those authors and publishers who did not want their books in GBS could ask for removal
- New business models, choices for consumers
- GBS books would be accessible on multiple platforms (unlike the Kindle & Nook)
- In-the-cloud so can access from anywhere

# CLASS ACTIONS

- US law allows small # of individual plaintiffs to sue a firm that has committed the same wrongful acts vs. them and others through class action lawsuit
- Courts must certify that the class (find that class reps have typical claims, common fact issues)
  - If court perceives conflicts among class members (some want X, others want Y; some think X is unlawful, others think o/w), this can lead to disapproval of the class, in which case the individual plaintiffs can only sue on their own behalf
- Typically, class action lawyers get 1/3 or more of any \$ award if the class action is successful
  - Threat of aggregated damages on behalf of the class gives class action lawyers a lot of leverage to press for a settlement
  - Jon Band estimated that G's potential liability is \$3.6 trillion

# CLASS ACTION SETTLEMENTS

- Litigants cannot settle class action lawsuits without judicial oversight
- Once a settlement has been announced, US judges will typically approve the settlement class for purposes of allowing the litigants to give notice of the settlement to class members
- Class members then have 6 months to opt-out, object, or comment on the settlement
- Settling parties have an opportunity to respond to objections
- Judge holds a “fairness” hearing to determine whether the settlement is “fair, reasonable, & adequate” to class members, whether notice was adequate, etc.

# GBSS REJECTED MARCH 2011

Judge Chin found merit in 6 types of objections:

- Scope of the settlement cf. issue in litigation (“bridge too far”)
- Adequacy of representation problems
- Antitrust concerns
- User privacy concerns
- Copyright issues
- International treaty concerns



# “BRIDGE TOO FAR”

- DOJ: Class counsel has obligation to litigate the claims they brought vs. G or to settle THOSE claims
- Complaint alleged infringement for scanning for purposes of snippet-providing
  - GBSS went far beyond this to address issues that were not in litigation (e.g., no plausible fair use defense for selling books)
  - Would give G a benefit that it could get neither from winning the litigation nor from private negotiations
- DOJ’s conclusion: judge lacks the power to approve this settlement because it is “a bridge too far”; Chin agreed

# ORPHAN WORKS

- Millions of books in GBS corpus would likely to be “orphans”
  - RHs cannot be found after reasonably diligent search
  - Likely to make up substantial part of ISD
  - G planned to charge profit-maximizing prices for OWs to end of © terms, even though no RH to pay out
  - Unclaimed work fiduciary under GBSS to act for orphan owners
- Chin: Congress, not private parties, should decide who should have guardianship over OWs

# REPRESENTATION

- Authors Guild hired a lawyer to represent some of its members & class of RHs whose books G had scanned or was planning to scan
  - Class reps & counsel have duty to represent interests of *all* class members, not just the interests of some
- Chin agreed with me that academic authors have different interests than Guild members
  - Academics are more likely to want OOP books available on open access basis; not profit-maximizers like Guild members
  - AAP, Guild brief: interests of open access advocates are “plainly inimical” to the interests of the class
  - But far more books in GBS are scholarly books than are Guild member books, far more academic authors than Guild members

# ANTITRUST ISSUES

- GBSS would have given G a de facto monopoly over commercialization of OOP books
  - This would allow it to offer an ISD of OOP books that no competitor could match
  - Creates risk of excessive pricing
- GBSS would arguably entrench G's monopoly in the search market
  - GBS will help G better respond to “tail queries,” MS & Yahoo! at disadvantage
  - Implication: should G have to give MS & Yahoo! access to GBS to improve their search technologies?

# PRICE GOUGING RISK

- Prices of ISD to be set based on # of books in the corpus, # services provided, & prices of comparable products & services (+ type of institution)
  - More books + more services = higher prices
  - No comparable products or services
  - G arguably planning to scan all 120M+ books in the world
- Prices might be modest at first to get institutions to subscribe, but history & logic suggest prices will rise over time to excessive levels because G would have a de facto monopoly on ISD (cf. journal prices)
- Only check on price hikes was complicated arbitration process in Michigan side agreement
  - Libraries could complain to UM that prices are excessive
  - UM could decide to initiate arbitration, but will it?

# PRIVACY

- GBSS would have required G to collect extensive amounts of information about users' reading habits
- Almost nothing in the GBSS to protect user privacy interests or to limit G's reuses of user data
- G said it would apply usual privacy policy, but is this enough?
- Chin: not by itself reason to disapprove GBSS, but troubled by this
  - Hinted that revised settlement should address this

# © ISSUES

- “Fundamental” to © that reusers have to get RH’s consent
- Many opt-outs and objectors were upset about GBSS because of shift in © default from opt-in (ask me first) to opt-out (I have to come forward to tell you to stop)
- France, Germany, many foreign RHs complained that the settlement violated US treaty obligations
- Chin: not deciding int’l treaty objections are sound, but this disturbed him also; why Congress should deal with

# INTERNATIONAL OBJECTIONS

- Many publishers & author groups from other countries objected to initial class definition
  - Would have given G a license to all in-© books in the world
  - If books were not commercially available in the US, GBSS 1.0 would have treated them as OOP, so G could commercialize
- Class narrowed to books published in Canada, UK, Australia + those registered with the U.S. Cop Office
  - But many foreign books are still within the settlement because publishers or authors registered with the Cop Office
  - Non-US publishers often do not keep good records re this
  - G will make judgment about national origin based on sites listed in front of book
- Berne Convention, TRIPs violation asserted



# AG v. HATHITRUST

- Authors Guild + several EU collecting societies brought suit vs. U Michigan, UC, & HathiTrust in Sept. 2011
- For direct infringement because of their possession and use of library digital copies (LDCs) that Google provided to them (10M books, 70% in-©)
- Also challenging orphan work project under which HT planned to display books it believed to be orphan works
- As with AG v G, main defense is fair use
- Case is pending before different judge
- Not focused on contributory infringement although UM & UC materially contributed to G's infringement by providing books, knowing G would scan them without © permission

# FAIR USE

- Is not © infringement in the US
- 4 factors typically considered:
  - Purpose of D's use
  - Nature of the ©'d work
  - Amount and substantiality of the taking
  - Harm or potential harm to the market
- G & HT are relying heavily on *Kelly v. Arriba Soft* decision to support FU
- AG & publishers analogize GBS to *UMG v. MP3.com* & *AGPU v. Texaco*

# KELLY v. ARRIBA SOFT

- Kelly, a photographer, sued AS for © infringement for making “thumbnail” images of photographs from his website
- Kelly argued not fair use because:
  - AS had a commercial purpose & did not transform the photos
  - The photos are highly creative works that deserve strong protection
  - Copies were being made systematically of whole photos
  - Kelly wanted license revenues for this use, so harm to the market existed

# 9<sup>th</sup> CIR IN *KELLY*

- “Transformative” because thumbnails were used for different purpose than the original, also smaller
- Kelly made his work available on the Internet, and knew ‘bots spider the web
- Whole works copied, but copies were incidental to facilitating better access to works
- No harm to Kelly’s market, indeed thumbnails may help customers find Kelly to license images
- Indexing is important to proper functioning of the Internet; thumbnails promote public access to content on the Internet

G argues that GBS snippets are like thumbnails in *Kelly*

# OTHER CASES FAVORING G

- *Field v. Google*: fair use for G to make copies of web content, cache copies, serve up snippets in response to search query
  - easy way for F to stop spidering: use robots.txt file on website
- *Perfect10 v. Amazon.com*: fair use for G to display thumbnails of infringing images and link to images (of which not aware when displayed)
- *AV v. iParadigms*: fair use to digitize student papers to detect plagiarism
- *Dorling-Kindersley v. Bill Graham Archives*: fair use to reproduce Grateful Dead posters for book about this culture
- *Cambridge U Press v. Becker*: Ga State professors made fair uses of book chapters, etc for course reserves

# UMG v. MP3.COM (SDNY)

- RIAA firms sued MP3.com for © infringement for “ripping” music from CDs for database of sound recordings for new service to allow its customers to listen to digital copies of recordings they owned
- MP3.com argued this was fair use because it facilitated users’ access to their collections, didn’t harm market because customers already owned the CDs
- UMG prevailed in © infringement ruling
- Lawsuit was settled for \$53M
  - Judge indicated intent to award \$118M in statutory damages unless parties settled, even though no actual damages to plaintiffs and no profits made by MP3.com on service

# AGPU v. TEXACO (2d Cir)

- Publishers of journals sued Texaco for photocopies made by its researchers of articles in journals to which T subscribed
- Texaco argued fair use: for research purposes, sci-tech fact-intensive works have “thicker” fair use, only few articles copied, no harm to market because of Texaco’s subscriptions
- AGPU won (2-1), mainly because CCC had program for licensing of photocopying of articles
  - 2d Cir also emphasized commercial and archival nature of copies, whole works copied systematically

# FAIR USE IN AG v. G?

Authors & publishers argue:

- G has commercial purpose; is making non-transformative uses of the books
- Systematic copying of © works of all genres, creative works
- Whole thing copied, systematic, stored permanently
- Presume harm to the market; also harm because lack of control, risk of loss if security breaches occur; besides, we want to license such uses

Google:

- + transformative ala *Kelly*; promoting public access to information
- + necessary to copy to index, make snippets available; access to orphan books
- + whole thing, but only snippets available unless au/pubr agrees to more thru GPP
- + transactions costs problems with clearing rights = market failure; GBS enhances market for many books (we'll link to where you can buy them)



# FAIR USE IN AG v. HT?

## AG arguing:

- HT is making non-transformative uses of the books; got from G so comm'l
- Systematic copying of © works of all genres, creative works
- Whole thing copied, systematic, stored permanently
- Harm because lack of control, risk of loss if security breaches occur; besides, we want to license such uses
- Uses beyond library privileges not fair use

## HT arguing:

- + LDC is transformative because promotes public access to information; preservation
- + factual, out-of-print works
- + whole thing, but making only very limited non-infringing uses of the works
- + no harm to market for works; indeed, may enhance market
- + fair use is available to libraries beyond what library privilege allows

# MICHIGAN OW PROJECT

- For LDC books published between 1923-63, plan was search first to see if © renewed
  - If not, in PD & can make available to patrons
  - If renewed, grad student investigators would search for RHs with guidelines for conducting searches
  - If RH could not be located, search was documented
  - UM to announce 90 day cooling period: we think this book is an orphan; if no one shows up to claim it, we will make it available to our research community
  - UM would take down display if RH showed up later
  - 1st notice about putative orphans posted in Sept. 2011 (what prompted the AG lawsuit)
  - Once AG sued, UM stopped this project

# FAIR USE FOR ORPHANS?

AG arguing:

- Archival & non-transformative uses of the books
- Systematic copying of © works of all genres, creative works
- Whole works to be displayed on systematic basis
- Presume harm; harm because unlicensed; risk of loss if security breach
- Up to Congress to address OW problem

HT arguing:

- + transformative because promotes public access to information; research/study
- + orphan status is most important factor
- + whole thing, but necessary to make lawful uses
- + market failure because no existing or potential market for OWs; market cannot form because of orphan status
- + allowing display of orphans promotes “progress of science”

Jen Urban, *How Fair Use Can Help Solve the Orphan Work Problem*,  
[http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2089526](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2089526)

# LITIGATION SCHEDULE

- March 2011: Chin rejected settlement; urged parties to settle on new terms
- Fall 2011: Chin set schedule for preliminary motions, discovery, and summary judgment motions
- May 2012: certified class, rejected G's motion to dismiss AG as plaintiff
- Cross-motions for summary judgment in AG v. HT have already been filed
  - Reply briefs due later this summer; hearing in the fall
- G's summary judgment motion due this month; hearing to be held in the fall; AG has asked for jury trial
- Publisher lawsuit is on hold

# GOOGLE BOOKS IN EU

- G has been scanning public domain books from some European libraries
  - British Library joined others
- German decision finding no infringement for serving snippets
  - Scanning in US does not infringe German ©
- French decision finding infringement under French law for scanning of French RH books, even though done in US
  - Recently settled
- G reached agreement with Hachette, among other publishers, over GBS

# SIZE OF OW PROBLEM

- British Library estimate: @40% of in-© works in its collection are likely orphans
- JISC study: 13-50M OWs in UK
- European Commission: estimates 3M OW in Europe
- HathiTrust estimate: @50% of LDC in-© books in corpus likely orphans
- CMU study: 22% but much higher for older books
- Drummond Testimony to Congress: 20% of GBS books are likely orphans
- Michael Cairns: @660K books likely to be orphan
- Jonathan Band: predicting @75% of GBS books would be unclaimed & effectively orphans
- Authors Guild: orphan works are a “myth”

# EU DEVELOPMENTS

- Council of Europe recommendation to develop solutions for public broadcasters (1999)
- Europeana digital library initiative aims to preserve cultural heritage, not just PD works
- 2008 High Level Expert Group on Digital Libraries recognized OW problem
- i2010: Digital Libraries Initiative of EC recommending digitization & online access
- Comite des Sages report, *The New Renaissance* (2011) recommended changes
- Sept 2011 MOU on Digitization & Making Available of Out-of-Commerce Works
- Proposed EU Directive on OW (as amended)

# COMITE DES SAGES

Solution to OW problem necessary to facilitate “new renaissance” for Europe

8 points for proper solution:

1. Solution for OWs should be in place in all member states
2. Solution should cover all sectors: audiovisual, text, visual arts, sound
3. Solution should ensure cross-border recognition of OW status (need only look in one state)
4. Solution should ensure cross-border effects of OWs (if available online in one member state, should be online in all member states)



# COMITE DES SAGES

5. Solution should be compatible with the implementation of standards for digitization
6. If use is commercial, should be a right of remuneration if RHs later show up (\$ could be escrowed)
7. Search should be commensurate with commercial value & age to ensure reasonable transaction costs
8. Rights info databases such as ARROW should support OW info-sharing

# PROPOSED EU DIRECTIVE

- Objective: to create a legal framework to ensure lawful cross-border online access to orphan works in digital libraries or archives when acting on public interest missions
  - Desire to promote Europeana and similar digital library initiatives
- Some member states have OW laws, but only operable in those states; no cross-border uses
- Need for harmonization, yet respect for national traditions

# ARTICLE 1: SM, SCOPE

- SM: books, journals, newspapers, magazines, other writings, cinematographic, audiovisual, & phonograms contained in certain collections
  - Applies to embedded works also
- Scope: permitting certain uses of OWs from collections of publicly accessible libraries, educational institutions, museums, archives, audio or film heritage institutions, public broadcasting to achieve public interest missions

# OW DEFINED, SEARCH

- Art. 2: OW = if all RHs in the work or phonogram are not identified, or if identified, not located despite a diligent search
- Art. 3: Diligent search if consult the appropriate sources for type of work in ?
  - Member states to determine after consulting with RHs and users
  - Need to carry out search only in member state of first publication or broadcast
  - Member states to ensure search, use, & user are documented in single EU-wide publicly accessible database
  - Annex identifies types of sources to check

# OTHER PROVISIONS

Art. 4: mutual recognition of OWs within EU

Art. 5: if RH shows up, can put end to orphan status

Art. 6: Member states to provide for exception or limitation to permit org's to making OWs available, reproduce for purposes of digitization, making available, indexing, cataloguing, preservation or restoration

- Only pursuant to public interest mission
- But can generate revenues to cover digitization & making available expenses
- Fair compensation owed to RHs who end OW status

# ANNEX

- Need to search
  - legal deposit
  - library catalogues
  - publisher & author assns
  - ISBN, ISSN, other registries
  - collecting society databases
  - sources that integrate multiple databases & registries

# QUESTIONS RE OW DIRECTIVE

- What about OWs first published in non-EU jurisdictions? (might be about EU)
- If member states adopt very different OW regimes, will there be enough harmonization to make OW more accessible?
- What about derivative works? Can they be made? What happens if RH shows up later?
- What about nonconsumptive research on OW corpus?
- What responsibilities for preservation quality, ongoing maintenance of preservation copies?
- Display but not download?
- Free use? ECLs? Compulsory license? Escrow?

# OTHER APPROACHES TO OW

- Registries of rights management information
- Canada: licensing by © board (Japan similar)
- Limit on remedies approach in US © office study from 2006
- HathiTrust fair use argument
- Internet Archive approach
- French legislation
- Nordic country ECLs
  - Hargreaves Review endorsed ECLs too
- German Library project
- September 2011 MOU on books & journals
- Legislative alternative to GBSS



# REGISTRIES

- Hugenholtz & van Gompel: OW problem is an info problem
  - Lack of adequate © registries contributes to OW problem
  - Forward-looking solution would be to establish registries of RMI
  - Create incentives for voluntary participation in them
- Hargreaves Report on UK IP policy: need for registry (Digital © Exchange) to avert OW
- Comites des Sages: registration should be precondition to fullest level of rights; change Berne Convention if necessary

# CANADIAN APPROACH TO OW

- Sec. 77 of Canadian © law
- Prospective reuser must identify work he/she believes to be an orphan & why
- Then submits a petition to © Board of Canada asking for license to make a specific type of use of the work
- Board reviews petition, determines whether to issue license, may establish fee, set terms
- RH has 5 years to show up & claim \$ (if any) from collecting society
- 252 licenses issued for specific reuses, 8 denied

# US ORPHAN WORK STUDY

- © Office did a study in 2006 of in-© works whose RHs could not be located after reasonably diligent search
- Recommended legislation to allow reuses of OW
  - If RH later showed up, reuser of work have to stop further uses, but limits on damages (might well be \$0)
  - If reuser had prepared a derivative work, he/she could continue to exploit that derivative subject to compensation to RHs, but no injunction
- Maria Pallante has announced CO's intent to renew efforts to persuade Congress re OW
  - Oct 2011 Report on Legal Issues in Mass Digitiz'n

# INTERNET ARCHIVE

- More focused on out-of-print than orphan works as such, but still applies to OW
- If library owns a physical copy of an OOP book, it may (in IA's view) digitize the book & lend the contents as long as
  - DRM protects from further copying
  - DRM causes copy to be deleted after 2 weeks
  - No more copies lent out than library owns

# FRENCH LEGISLATION

- Bibliothèque Nationale de France to compile a freely accessible online database of all works published in France before 1 January 2001 that are not being commercially distributed by a publisher and are not currently published in print or digital form
- If RH does not opt-out within 6 months, CMO able to issue license to reproduce & display these works
- CMO responsible to pay compensation to RHs (50-50 split between authors & publishers)
- If RHs do not show up within 10 years, work deemed an orphan; CMO to allow public libraries to make copies freely available

# SEPT. 2011 MOU

10 library, creator, & publisher assn's  
agreed to 3 principles:

1. Voluntary agreements on out-of-commerce works
2. Practical implementation of collective agreements
3. Cross-border access to digital libraries

# PROS & CONS OF ECLs

- Pros:
  - Users can get license that covers all rights for the affected works & uses
  - Compensation collected; CMO responsible to distribute fairly
  - Diligent search for RHs not necessary
- Cons:
  - CMO may not always represent non-member interests fairly
  - Compensation to members may be uneven
  - Lack of transparency, good governance problems
  - In US, unclear who would be CMO/licensor

# GBSS AS OW SOLUTION

- GBSS was clever: let's generate \$ from commercializing OOP books, give 63% to BRR, & let BRR use part of this \$ to look for RHs to pay them what's due
  - When © owners located, they will likely sign up to get \$\$\$; no need to get advance permission
  - OW problem will disappear as BRR let RHs know about \$\$\$
  - Give money to literacy charities if true orphan



# MASS DIGITIZATION OF OOP

- One of the other innovations in GBSS was the quasi-ECL regime it would create for out-of-print works
- Like an ECL in that G would get license from the class to make works available beyond the RHs registered with BRR, BRR to escrow \$ for RHs till sign up
- Unlike ECL in that only G would have been able to get such a license, no gov't oversight, non-registrants with BRR could not get paid \$ due them from G
- Norway experimenting with ECL for OOP works; is this a good approach for dealing with digitizing non-orphans?

# LEGISLATIVE ALTERNATIVES

- Allow mass digitization of books with tiered access by qualified entities who are willing to commit to security measures
  - OK to digitize books for preservation purposes
  - OK to display snippets for in-© books (unless RH says no), with links to sources from which books can be lawfully acquired
  - Non-consumptive research privilege, at least for nonprofit researchers
  - Non-expressive uses privilege (e.g., to improve search tools)
  - Full text access for public domain and books known to be “orphans” (take down if RH shows up later)

# NON-DISPLAY USES

- GBSS defined as uses that do not display expression
  - e.g., display of bibliographic information, full-text indexing without display of Expression (such as listing the number or location of search matches), geographic indexing of Books, algorithmic listings of key terms for chapters of Books, and internal research and development
- *iParadigms* decision & search engine cases suggest this might be fair use in US
- Matthew Sag argues this in *Copyright & Copy Reliant Technologies*
- Hargraeves UK IP Report endorses exception to allow non-display uses

# AMICUS BRIEF

- Digital Humanities + IP profs submitting amicus curiae brief in AG cases:
- Digitizing texts for purposes of data-mining should be fair use:
  - Purpose is for research/study
  - Copy whole thing, but necessary to index
  - No harm to the market for the work
- Data-mining of digitized corpus does not infringe © because making only non-expressive uses of the text (i.e., extracting facts about the works)
- Important insights possible through data-mining corpus of research library texts

# OTHER ISSUES

- Public/private partnerships
- Interoperability, open standards
- Preservation quality scans
- Preservation as a constant process
- Digital deposits, works born digital
- High quality metadata: IP rights, CC, or public domain?
- Commercial & noncommercial services to add value to mass-digitized corpus of works
- Future of automated licensing
- User privacy protections (reading, viewing)
- Improved access for print-disabled persons

# BERKELEY DL © PROJECT

- Sloan Foundation grant to support research on how to overcome obstacles to digital library projects, such as Digital Public Library of America (DPLA)
- Series of white papers, briefs, op-eds:
  - <http://www.law.berkeley.edu/12115.htm>
- Conference on “Orphan Works and Mass Digitization: Obstacles & Opportunities”

# BERKELEY CONFERENCE

- Who wants to make use of orphan works and why?
- Who is concerned about broader access to orphans and why?
- What is the best approach to addressing the orphan works problem?
- What role should registries play in averting orphan work problems? What mechanisms will facilitate information sharing about which works are public domain, orphan, or open access?
- Who wants to do mass digitization and why?
- Should data mining and other non-consumptive uses of in-copyright digital works be permissible, and why?
- <http://www.law.berkeley.edu/11731.htm> (audio & slides)

# CONCLUSION

- Mass-digitization of cultural heritage is highly desirable
- Copyright is presently an obstacle to this more than financial constraints & technology limits
- Orphan works is the most pressing of the policy issues that needs to be resolved
  - # of solutions proposed
- Other changes may also be desirable (e.g., better registries, nonconsumptive research privileges, more access for print-disabled)