Mass-digital copyright?

Maurizio Borghi

Centre for Intellectual Property Policy & Management
Bournemouth University

1. How digital copyright rules the (online) world

2. From digital to mass-digital

3. A new paradigm in copyright law (and why we should care)
Digital copyright

Legal paradigm resulting from the encounter of copyright norms with the digital networked technology

(WIPO Treaties / DMCA / InfoSoc + e-commerce Directives)

- Reproduction: ‘in any manner or form’ / permanent + temporary (subject to limitation)
- Distribution: no online exhaustion
- Technological measures: anti-circumvention provisions
- Safe harbour provisions for service providers

Digital copyright – criticisms

*The copyright discourse of the last decade has been largely dominated by few recurrent themes...*

- DC transplants to the digital world the proprietary model of the offline world
- DC is not user-friendly
- DC jeopardizes the availability of exceptions and limitations
- DC threatens civil rights and liberties
- Copyright industries must adopt ‘new business models’...

*Most of the troubles with DC seem to depend on the uncertainty about the status of the ‘copy’ in the digital world*
Why is copying an infringement?

Copying as ‘predictor’ of intention to infringe (Litman, Miller & Feigenbaum)

...or as a wrong to the author’s autonomy as a speaking being (Drassinower)

Is copying per se (irrespective of the purpose of copying) a wrong?
  e.g. copying for private use
  copying to enable legitimate uses of a work (reading, listening,...)

but also
  copying for uses that are unrelated with the normal destination of the work

Mass-copying technologies / businesses

Services that operate on the basis of routine, automatic and indiscriminate copying of works en masse

- Indexing and search (search engines, digital libraries)
- Mass digitization projects (Google Books, HathiTrust, Internet Archive, Europeana...)
- Web scraping / crawling (Expedia, Ciao...)
- Media monitoring (Google News, Meltwater, Infopaq...)
- Development of content-transformation tools (e.g. statistical machine translation, voice recognition algorithms, automated text reading)
- Text mining services (iParadigms)
Copyright and mass-copying

*How does digital copyright accommodate mass-copying technologies?*

- Fair use defence in the US (successfully applied in a number of cases, e.g. *Kelly v Arriba*, *Perfect 10*, *iParadigms*, *HathiTrust*, …)

*In Europe?*

- Temporary reproduction exception – *Infopaq* 1 & 2 (CJEU), *Meltwater* (UKSC)
- Implied licence – *Vorschaubilder I & II* (BGH)
- *Ius usus innocui* – Spanish Supreme Court, 3.10.2012

Private ordering systems

*Large companies and institutions operating in mass-digital activities have figured out systems to legalize mass-copying within the current copyright regime*

- Regulatory solutions (not successful: e.g. Google Books Settlement Agreement; HathiTrust Orphan Works Project)

- Content-usage licence agreements – covering bulks of in-copyright works for certain uses (increasingly popular – e.g. agreements between Google/ Amazon and publishers associations, etc.)
Legislative initiatives

New exceptions to legalize wholesale copying in specific instances

- Proposed exception for text mining and data analytics (UK)

Licensing schemes covering mass-uses of certain categories of works (compulsory licensing with opt-out)

- French law on the digital exploitation of out-of-commerce books (22.3.2012)
- Extended collective licensing on orphan works (e.g. ERR 2013, UK)

How mass-copying is changing copyright

By effect of judicial interpretation, private ordering solutions and changes in legislation, a number of uses involving mass-copying are (de facto or de iure) permitted

- Tolerated uses (with opt-outs)
- Uses (partially) covered by existing exceptions or limitations (upon judicial examination)
- Uses covered by broad content-usage agreements (and possibly by other regulatory mechanisms)
- Uses (to be) regulated by new legislation

...Evils?
Minimal human intervention in selecting, organising, indexing and displaying content

Scanning ‘all books’ – irrespective of copyright status (but displaying only PD or licensed content)

Works increasingly used for automated processing (extraction of information, ‘data mining’, experimental uses)

“We’re not scanning all those books to be read by people. We’re scanning them to be read by Artificial Intelligence”. (Anonymous Google Engineer)

Books to form part of the ‘Big Data’ environment

“A single liquid fabric of interconnected words and ideas” (Kevin Kelly)

---

Digitizing all the world’s books

Total number of books in bibliographic records in the world = 174m

*Mass digitisation projects (examples):*

- **1990s:** Project Gutenberg (30k texts, mainly public domain books)
  [http://www.gutenberg.org](http://www.gutenberg.org)
- **1997:** Internet Archive (extensive collections of digital media content; about 3.5M books in PD or licensed by authors;)
  [http://www.archive.org](http://www.archive.org)
- **2004:** Google Print (then Google Book Search, then Google Books): planning to scan and make available for search “all world’s books” (currently 20M)
  [http://www.books.google.com](http://www.books.google.com)
- **2006:** Live Search Books (December 2006 – discontinued on May 2008; 750K books scanned – now available from Internet Archive)
- **2008:** Europeana [http://europeana.eu](http://europeana.eu): the “portal of European cultural heritage” – 2.200 partner institutions (currently comprises 23M ‘digital objects’)
- **2008:** HathiTrust [http://www.hathitrust.org](http://www.hathitrust.org) partnership of 70 US university libraries; 10M volumes, 3.2M in PD
Evil in the details

DIY copyright

*Digital landscape increasingly shaped by private agreements between large companies / rightholders associations. Who represents the ‘public interest’?*

Integrity of the cultural heritage

*Becoming part of the Big Data environment, works and collections may risk being used in forms and ways that do not respect their intimate destination and purpose.*

Digital monopolies

*In the current regime, mass-digital technology may enable the creation of new (not readily transparent) monopolies on information and knowledge*

---

Need for a new regulatory framework?

- Works no longer used ‘as works’ but as data-containers
- Value no longer lies (only) in the use of individual works, but in the use of the aggregate of works *qua* data-containers
- New (and unpredictable) forms of dealing with large quantities of ‘digital objects’
- Cultural heritage integrity and accessibility: digital libraries as ‘cultural genome’ – fully-fledged public goods

*Subject-matter of regulation* = The ‘corpus’ of digital objects

*Activity to be regulated* = Automated processing
Analogical application of DP

Definition of automated processing (Art 2(b), Directive 95/46/EC)

...any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction;

- Can principles and concepts of data protection law be applied by analogy to digital repositories of copyright works?

What © can learn from DP

Regulation of population genetic databases

_Informed consent on all (future) uses not possible. General consent legitimizes uses of personal information that are consistent with public utility goals (under the doctrine of social solidarity). Data subjects have a right to opt-out_

Generic consent and purpose limitation: the Source informatics case

_When personal information is released there is a legitimate expectation that the information is not used for purposes that are unrelated with those for which it was released in the first place (under the doctrine of informational self-determination)_
In conclusion

✓ Copying as such can no longer be the benchmark of infringement
✓ Wholesale copying for automated processing (without public communication) is progressively gaining legitimacy in copyright
✓ Copying for uses unrelated to public communication can be permitted

However

✓ Digital repositories can also be ‘public goods’!
✓ Automated processing of digitized works must be 1. subject to purpose limitation and (abstent authorial consent) 2. consistent with public utility goals.

More on ‘mass-digital copyright’:

M Borghi & S Karapapa Copyright and Mass Digitization (OUP 2013)

Thank you!