Copyright and the Big Data: regulating the power of knowledge

Dr Stavroula Karapapa, Associate Professor
Big Opportunity

Knowledge becomes accessible *en masse*

New knowledge can be generated

Big data unlocks innovation

Eg *new* services emerge

- 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)
Big Challenge

- Mass digitization
  - i.e. the conversion of works from physical support to digital format on an industrial scale
  - routine, automatic and indiscriminate *copying* of works

- Use and re-use of data
  - Eg new research tools, text and data mining, linking data and publications, extracting value from information, making data useful and *accessible*

- Licensing and rights clearance
  - Who *owns* information? What about the *orphans*?
Excessive copyright protection

How does copyright accommodate mass-copying technologies?

In the US

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

In Europe?

- Temporary reproduction exception – *Infopaq* 1 & 2 (CJEU), *Meltwater* (CJEU, UKSC)
- Library copying and e-lending - *Technische Universität Darmstadt* (CJEU)
- Implied licence – *Vorschaubilder I & II* (BGH)
- *ius usus innocui* – Spanish Supreme Court, 3.10.2012
Contracting out of liability

- Private ordering mechanisms to legalize mass-copying within the 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.)
Changing the law

New (but limited) © exceptions to legalize wholesale copying

- 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)

Ancillary copyright for press publishers

- German law of 14.03.2013 (“Google tax”)
Need for a *truly* new legislative framework?

- Laws do not take account of the aggregate impact of mass digitization.
- Value lies no longer (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’
  - Subject-matter of regulation = The ‘corpus’ of digital objects
  - Activity to be regulated = Automated processing

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

- Can the available (including the newly introduced) provisions accommodate the mass effect of the use and reuse of big data?

- Who represents the ‘public interest’?

  - Digital landscape increasingly shaped by private agreements between large companies / rightholder associations.

- The rise of digital monopolies – an elephant in the room?

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

  - e.g. Google’s ‘preferential uses’ on digitized books: monopoly over the ‘computational potential’ of the data aggregate
More on ‘mass-digital copyright’:

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

Thank you!