Mass-digitization and the 'dream' of universal access

Panel report by Kelly Breemen (IViR), July 2014

With: **Martin Kretschmer** *(CREATe, Glasgow University),* **Elisabeth Niggemann** *(Deutsche Nationalbibliothek),* **Pamela Samuelson** *(University of California, Berkeley) & moderator* **Martin Senftleben** *(VU University).*

The conundrum of online access. That is what the mass-digitization session of the Information Influx conference addressed. The panel focused on one of the central issues surrounding digitization by cultural heritage institutions: the ability to give access to the public. And what came to the fore was indeed a myriad of approaches and topics, all related to the dream of universal access.

Multiple approaches

The discussion started with presentations that gave an overview of three perspectives on the issue: academic, practical and juridical. Departing from these perspectives, the panelists outlined the following approaches: 1) a study of a practical, risk managed approach to rights clearance; 2) practical activities based on a legal approach (EU Orphan Works Directive and German Out-of-Commerce Law); and 3) the fair use approach as established in Section 107 of the US Copyright Act and US case law (Authors Guild v. HathiTrust and Authors Guild v. Google Books).

Risk-management: digitisation in practice and empirical realism

The academic perspective included the discussion of two studies that departed from the viewpoint that non-use of works is unacceptable. As this would add no value to owner, public or economy, works should be available accordingly. Right now, however, this is not the case. Books from the 1890s are for example more commercially available than books from the 1950s. In this sense, there is a 'hole', with a large part of works not being available. What could be solutions to take this material back out of the hole?

One of the studies[1] that were described empirically assessed the practical digitization approach of the Library of the Wellcome Trust for making material available.[2] This practical approach applies a strategy of risk management to enable digitization efforts. To this end, material is placed in a high, medium or low risk category. All the material outside the high or medium risk category is put online with a notice or disclaimer of a takedown policy. For the rest, it is attempted to track down the right holders. For 84 percent of the 14.000 items this succeeded, and 98 percent of the right holders gave permission for digitization and online access. Strikingly, for the 2 percent who did not give permission, it was not copyright that played a role in denying permission, but there were private issues involved. One of the conclusions of the presentation was that a large machinery of licenses seems unnecessary if

---

[1] V. Stobo, R. Deazley and I.G. Anderson, *Copyright & Risk: Scoping the Wellcome Digital Library Project*, CREATe Working Paper No. 10, December 2013, available via: http://www.create.ac.uk/wp-content/uploads/2013/12/CREATe-Working-Paper-No.10.pdf.

[2] See for more information: http://wellcomelibrary.org/about-us/projects/digitisation/.

everyone agrees on the access question. The study offers empirical realism and encouragement for low risk infringement.

Practice in Europe: ruled by a legal framework

The practical talk gave insight in the obstacles and developments for day to day library practice in Germany. This practice draws and depends on the legal approach to the issue on both the European and national level. It was illustrated that this approach is limited in success and still presents barriers.

The importance of digitization efforts was stressed with the motto of many people, especially younger generations: "if it's not online, it doesn't exist". However, this 19th, 20th and 21st century black hole is problematic, for a lot of material from that period (still) has many relevance today, comprising important events and topics that shouldn't be forgotten about.

At the European level, it was argued that the Orphan Works Directive[3] closes a gap. However, the required diligent search was considered (too) time-consuming. Still, for significant documents it is important that it is there, so it can be followed as an exception. At the German level, the new law on out-of-commerce works[4] was welcomed from a library point of view as a light way to obtain licenses for digitizing out of commerce print. All works printed in Germany before 1966 can be licensed by one of the collecting societies. This practice requires the building of a database and the use of metadata to check whether works are still in commerce. No right holders have to be found or negotiated with, works just have to be placed in the database. Furthermore, it will be a one-payment step and a license forever. Authors can claim it back if they wish. The license-fees are not yet known, but this legislation offers a light and easy way for digitization for at least this type of works.

Flexibility in the US: case law based on fair use

Finally, the talk on case law developments in the US and the fair use doctrine of Section 107 of the US Copyright Act discussed Google Books' and HathiTrust's digitization practices and the resulting court cases. The HathiTrust case revolved around three uses for the digitized copyrighted books: preservation of books, enabling full-text search through a database and access for print-disabled people. The United States Court of Appeals for the Second Circuit accepted fair use for almost every issue at stake in the case.

For fair use to be accepted, there are four criteria that are taken into account: 1. purpose and character of the taking (in this case the digitization); 2. the nature of the copyrighted work; 3. the amount used; and 4. the question of harm in the marketplace caused by the use.

In the HathiTrust case, it was decided that the purpose of the taking was different from the original: books are not created for the purpose of creating a full-text searchable database. For this point, the case was compared with thumbnails and search engines. As to the nature of the copyrighted works, it was considered that 93 percent of the books are non-fiction works, by scholars and for scholars. Then, as to the amount, it was concluded that although the

---

[3] Directive 2012/28/EU of the European Parliament and of the Council of 25 October 2012 on certain permitted uses of orphan works, available via: http://ec.europa.eu/internal_market/copyright/orphan_works/index_en.htm.
[4] See for an (unofficial) English translation:
https://www.vgwort.de/fileadmin/pdf/allgemeine_pdf/out_of_commerce_law_2013.pdf.

digitization concerned millions of books, this was reasonable for the purpose. In order to create a full-text searchable database, all of the works need to be digitized. And finally, harm in the market-place was not accepted. The facts of the Google-case are slightly different, as Google also presents snippets of text in its search results. The Court has yet to decide, but acceptance of fair use of the copyrighted works is not ruled out here.

Also the initiative authorsalliance.org was put forward, where authors want people to find their books. Here again, the risk of losing cultural heritage and the public's perception that what is not available online does not exist are stressed. A conclusion is thus that fair use is now a tool for mass-digitization, and that many heritage institutions will also be able to engage in this practice.

Will the dream become reality?

That the answer to this question is not an easy one, became clear in the general discussion. The importance of finding ways to overcome copyright obstacles was stressed. Would tailor-made solutions for digitisation projects be an option? Do we need publishers anymore? Is it relevant for the EU with its strict definition of reproduction not to look at the copy but the purpose? What counts is not the copying, but what happens after: the communication to the public. It is questionable whether the text of the European library exception in the Copyright Directive can be interpreted as a way for mass-digitization. But what would be the harm to copyright in the case of making works searchable? In the US, for example, there is more attention for the purpose and for considering the copying as an intermediate step. The focus is on the 'end-product' and the question whether that is a non-infringing work. The HathiTrust activities, for example, are similar to intermediate copies and lead to a positive use of copyright and the copies. Therefore, they are non-infringing copies. Also, a library's mission to index and catalogue was emphasized.

Furthermore, the step after making copies was stressed: making material available. As limitations and exceptions to copyright pursue social goals, it was argued that a re-balancing act is essential as there may be a conflict between interests of right holders and interests of cultural heritage institutions in making works available. Not performing such a re-balancing act would only continue the current state of affairs.

A suggestion from the audience was to digitize and make available material that is not commercially available with a notion and takedown (NOTD) procedure or a liability rule and without an obligation to pay. The default setting would then be that everything can be digitized and made available, but with a NOTD-system in place for right holders who object. Would it be against the three-step test if it is empirically demonstrated that not many authors will in fact object? What also came up, is the necessity for non-use to have consequences, as is the case in other areas. It was deemed strange that this is not the case in copyright law. Added to this was a suggestion of abuse of rights and unnecessary infringement of free speech, and that how longer a rights lasts, the narrower it should become.

And finally, the role of cultural heritage institutions that have digitized works was assessed. Are they really the 'good guys'? It was stressed that what is in the public domain should stay there and not be commercialized by cultural heritage institutions. When works have finally been digitized and made available, it should not be these institutions that claim rights in public domain works afterwards.