Mass digitisation and the moral right of integrity

Maurizio Borghi[1]

Author's moral rights seem to have hard-living in the digital age. In spite of the often repeated statement that moral rights should become one of the central issues in the international debate on digital copyright, the legislative initiative in the field has been so far lazy, to say the least. To be sure, international bodies and governments have been very reactive to adapt economic rights to the challenges of new information technologies. By contrast, no significant initiative has been taken regarding moral rights. Both the TRIPs Agreement of 1994 and the WIPO Copyright Treaty of 1966 are silent on moral rights. At European level, moral rights have not been addressed in the seven directives constituting the *aquis communautaire* and are expressly excluded from the scope of the Information Society Directive 29/2001.[2] Thus, with the sole exception of the WIPO Performances and Phonograms Treaty of 1996, which mandates for the extension of the moral rights of attribution and integrity to performers,[3] the international legislator has been completely silent on moral rights from far before the beginnings of the digital age. To date, the Berne Convention of 1886 – lastly amended exactly 40 years ago[4] – is still the most important and almost unique international legal instrument in the field. Art. 6 *bis*, which was first introduced in the Convention at the Rome Conference of 1928 after prolonged controversies,[5] recognises two relevant moral rights, namely paternity (or attribution) and integrity. However, no special provision has ever since been enacted to address the respect of moral rights in the digital environment.

One of the reasons of this void is certainly the lack of a general feeling of urgency in this matter. Digital technologies and the networked environment have dramatically facilitated the way by which copyright works are *reproduced* and *distributed*. This has obviously had an impact on the two major economic rights of copyright, namely reproduction and distribution, and the legislator's promptness to expand the scope of these rights came with no surprise. As far as the European harmonisation process is concerned, moral rights were mentioned in the Green Paper of 1995 as one of the issues

[1] Brunel University Law School, London <maurizio.borghi@brunel.ac.uk>

[2] See Rec. 19.

[3] WPPT, art. 5.

[4] Berne Convention for the Protection of Literary and Artistic Works of 9 September 1886; lastly revised in Paris, 24 July 1971.

[5] See Sam Ricketson & Jane Ginsburg International *Copyright and Neighbouring Rights: The Berne Convention and Beyond*, 2nd ed., OUP 2005, pp. 223-230.

that, with the advent of the information society, were deemed to become more urgent than before.[6] In a Follow up of 1996, the Commission informed that an overwhelming number of interested parties had stressed the importance of moral rights in the digital environment, and particularly the right of integrity, and recommended further studies on the issue.[7] Yet in 2000 a comprehensive study carried out by Salokannel and Strowel on behalf of the Commission concluded that, although the level of protection of moral rights differs significantly amongst Member States, there is no evidence that the existing differences affect the functioning of internal market. The study discouraged legislative initiative in the field and recommended the adoption of "soft law" mechanisms, such as good practices, as a more appropriate instrument to address the respect of moral rights in the digital environment.[8] As a result, the Information Society Directive passed on 2001 by leaving moral rights outside of its scope.

Certainly, this lack of initiative in the part of the world where moral rights are supposed to be championed has deeply affected the international legislation process. However, the overall negligence on this matter has still another explanation. Digital networked technologies have caused new forms of creating and using works to become prominent. In this respect, the "cyber-space" has been described as dominated by large scale peer-production instead of centralised information production,[9] re-mix or "read-and-write" (RW) culture instead of "read only" (RO) culture,[10] briefly a space where the notional concepts of the public sphere are fated to blur. Along this line, commentators have extensively emphasised that the cyber-space challenges the traditional concepts of authorship, work and use, upon which copyright norms are framed. For instance, it is commonly observed that the world wide web tends to promote forms of collective authorship (e.g. Wikipedia) that do not fit well the copyright notion of "the author". Similarly, works are less used in a purely consumptive way, and are increasingly transformed and re-used in a more or less creative way, so that a more "cavalier" approach towards work's integrity and attribution is already adopted in practice. In order to take full advantage of peer-production and re-mix culture, certain barriers of copyright should be better removed. It is on this background that "traditional" author's moral rights seem to have become, as Guy Pessac has put it, "*persona non grata* within the cyberian

[6] Green Paper on Copyright and Related Rights in the Information Society, European Commission Report, COM (95) 382 final, 19 July 1995, p. 29.

[7] Follow up to Green Paper on copyright and related rights in the information society, European Commission Report no. COM (96) 568 final, 20 November 2006.

[8] Marjut Salokannel & Alain Strowel (with the collaboration of Estelle Derclaye), Study contract concerning moral rights in the context of the exploitation of works through digital technology, Study contract n° ETD/99/B5-3000/E°28, Final Report, April 2000.

[9] See Yochai Benkler, "Sharing Nicely: On Shareable Goods and the Emergence of Sharing as a Modality of Economic Production", 114 *Yale L.J.* 273 (2004); Erez Reuveni, "Authorship in the Age of the Conducer", *Journal of the Copyright Society of the U.S.A.*1801(2007).

[10] Lawrence Lessig *Remix. Making Art and Commerce Thrive in the Hybrid Economy*, Penguin Press 2008.

discourse"[11]. As a matter of fact, scholars who have addressed the issue of moral rights in the digital environment have mainly advocated for a narrower construction of their scope.[12]

Mass digitisation projects as a new challenge to moral rights

Yet with the advent of mass digitisation projects the scenario has changed. Mass digitisation is the conversion of works in digital format on an industrial scale. Projects such as Google Books, Internet Archive or Europeana, just to mention the most popular ones, come with the unprecedented and well-trumpeted promise to resuscitate the dream of the great library of Alexandria. As a matter of fact, mass digitisation projects are engaged in the mission of transplanting the whole cultural heritage of humankind, as deposited in books and in other physical carriers, in the digital networked environment. Millions of items that carry works born in the pre-digital era are turned into sequences of bites that form part of the cyber-space. In this way, mass digitisation of cultural heritage does not only improve the quantity of information that is available in the cyber-space, but it adds to the cyber-space something that the latter would have never produced on its own. The to-be-digitised works are not "products" of the cultures of the cyber-space; they are, rather, transplanted into the cyber-space by virtue of a mechanical operation. Works are transplanted from a context to another. We may term "authorial context" the environment from where they originate, and "cyber-context" the environment to which they are transplanted by effect of mass digitisation.

In the authorial context, works are, so to speak, *dense* units. Books are not just accidentally physical containers of information that would otherwise flow unrestrained in the public sphere; they are rather "discrete documents that operate with internal cohesion more than external linkages".[13] Books, as physical objects, are gatekeepers of this cohesion and comprehensiveness that belongs essentially to literary works. Similarly, dramatic or musical works are normally internally coherent units whose intimate sense and meaning depends on the maintenance of this cohesion. This holds true for most of the works that are generated in an authorial context. To be sure, although works are dense units, they can be "fragmented" by use. For instance, parts of the work may be extracted and used in other works as quotation or as part of a compilation, while the content of a work may be abridged, summarised or even partially reused in different context. All these ways of fragmenting the original whole in which the work consists presuppose a creative effort on the side of the subsequent author. The fact that the latter can legitimately fragment the work only on condition that

[11] Guy Pessach, "The author's moral right of integrity in cyberspace - a preliminary normative framework", 41(2), *IIC* 2010, 187.

[12] See the literature quoted in Pessach *supra*.

[13] Siva Vaidhyanathan, *The Googlization of Everything (and Why We Should Worry)*, California University Press, Berkeley 2011, p. 171.

he spends his own creative effort is an empirical proof of what we have called the "density" of the work itself.

In the cyber-context, by contrast, works tend to be *fluid* units. Not only is the internal cohesion no longer externally sustained by physical supports, but, most importantly, it is disarticulated in the first place. This happens essentially at three levels. First, at a pure technical level, works in digital format are reduced to a sequence of ones and zeros which flows in packages that can be disaggregated and re-aggregated as will. The case of Torrent files is probably at the moment the best exemplification of this technical disaggregation of the work's unity. Second, at the level of human perception, works are increasingly exploited on a piece-by-piece basis instead of in whole; for instance, fragments of films or of broadcasts are displayed on YouTube, single tracks or parts of larger musical works are downloaded from iTunes or from peer-to-peer networks and snippets of books and of articles appears in search engines or on digital libraries as response to search queries. These forms of exploitation change the way by which works are commonly used, and one may wonder whether all these fragments of works do not acquire a new economic significance per se.[14] Third, each work or part thereof is virtually linked to any external entity and can be equally internalised by any other work *via* framing, cut-and paste or remix. As it has been pointed out, books, once digitised, are no longer bound by "the paradigm of the physical paper tome" and, through the search engine technology, they become part of a "single liquid fabric of inter-connected words and ideas".[15] This "fabric" contains virtually the entire works of humankind in all languages, and it is not limited to books but extends to images, sounds and films. Once these items are transplanted from the authorial context to the cyber context via mass digitisation, then the "real magic" can start: "each word in each book is cross-linked, clustered, cited, extracted, indexed, analyzed, annotated, remixed, reassembled and woven deeper into the culture than ever before. In the new world of books, every bit informs another; every page reads all the other pages. [...] Once a book has been integrated into the new expanded library by means of this linking, its text will no longer be separate from the text in other books."[16] As a matter of fact, in the "single liquid fabric" containing virtually all works produced by the whole humanity in all forms and languages, everything can be linked to anything without opposing resistance.

As a result, the transplant from authorial context to cyber context, that is from a dense to a fluid context through mass digitisation may bring about an unprecedented, and so far almost unnoticed, threat to author's moral rights. As Hector MacQueen has pointed out, in the context of the latter

[14] As suggested by the Belgian case *Google Inc v Copiepresse SCRL*, Court of First Instance of Brussels, [2007] E.C.D.R. 5, discussed below.

[15] Kevin Kelly, "Scan this Book!", *The New York Times Magazine*, 14 April 2006, <http://www.nytimes.com/2006/05/14/magazine/14publishing.html?_r=1>

[16] Id.

developments of the digital world as determined by the advent of mass digitisation, "moral rights of attribution and integrity can be the bulwark of other authorial interests, offset by appropriately framed exceptions for education, research, news reporting, public libraries and parodies".[17] In the following, we discuss how the right of integrity can help defining the legal boundaries of mass digitisation projects.

Defining integrity

The right of integrity is commonly defined, in general, as the right to object to distortion, mutilation, or any other derogatory action in relation to the work which could be detrimental to the author's honour and reputation.[18] Although referred to the honour and reputation *of the author*, the right applies to actions performed in relation to *the work*. To understand the proper meaning of the right of integrity it is essential to consider it from this particular angle.

Copyright subsists in what is generally called a "work". Yet most of the rights that the law grants to the author of a work apply in fact to actions that third parties perform in relation to *copies* of the work. The author has in fact a right to make copies (reproduction), to issue copies to the public (publication) and to control the circulation of copies (distribution). The rights of reproduction, publication and distribution represent undisputedly the pith and marrow of the author's economic rights and they are recognised as exclusive rights in all copyright systems. In certain jurisdictions the author has also the inalienable rights to have the copies effectively and efficiently disseminated (divulgation) and to remove the copies from the public sphere (withdrawal).[19] All these rights enable the author to sue third parties for actions that they might perform in relation to *copies* of his work. They are copy-rights in a literal sense.

However, copyright does not enable control over copies only. So, for instance, the reproduction right is infringed not only when copies – namely "specimen of the same work"[20] – are made without authorisation. In fact, also non-literal and non-exact reproductions may be infringing. Here the infringing action is addressed to the work and not only to the copy as such. More precisely, the action is infringing to the extent that the *form* of the work is substantially taken, and not only the content or the matter expressed. The right of the author extends in fact to the form in which the work is expressed. The concept of form is broad enough to include "external" and "internal" form, the first being the sequence of words or of sounds or the arrangement of colours and figures by which the work is composed, and the latter being the construction of the expression in a broad

[17] Hector MacQueen "The Google Book Settlement", 40(3) *IIC*, 247 (2009), p. 249.
[18] See Berne Convention, art. 6 *bis*.
[19] The right of divulgation is recognised, in different forms, in Germany and in Italy; the right to withdraw or retract is recognised in France, Germany, Greece, Italy, Portugal and Spain.
[20] Oxford English Dictionary, entry "copy".

sense, including the style and manner applied to that particular matter.[21] This means that any modification, alteration or adaptation of the form of the work is also restricted to the author. Translation of a literary work in a different language, for instance, is at the same time the creation of a new work and the alteration of the external form of the original work.

The right of integrity is closely related to the right to modify the work, that is to make translations, adaptations, arrangements or, as in the American wording, "derivative" works – however, it does not merge with it. Modifying the work is only a condition, and not even a necessary condition, for breaching the work's integrity. So, while an unauthorised translation impinges upon the right to translate the work, a bad translation may infringe the right of integrity regardless of whether it is authorised or not. The two claims of infringement are clearly distinct. Moreover, integrity can be infringed by placing the entire work in an inappropriate context, without even modifying or mutilating the work as such. The scope of application of the integrity right is broader than that of the right to modify the work. Similarly, the integrity right is broader than the right of reproduction as applied to non-literal reproductions. Both modification and reproduction rights are in fact limited to the *form* of the work. The integrity right does not apply only to actions carried out in relation to the work's form. In a way, it can be infringed by any action performed in relation to the work itself.

Work as such	Work's form (both 'external' and 'internal')	Work's content	Copy of the work
Integrity	• Reproduction (non-literal) • Adaptations, translations, etc.	No copyright	• Reproduction (literal) • Publication • Distribution • Divulgation • Withdrawal

While copyright as such subsists in what is generally called a "work", the right of integrity puts the work itself at the forefront. Yet what is a copyright "work"? In spite of the centrality of the notion of work in all copyright systems, one can hardly find a definition of the work in either statutes or case law[22]. It is generally known that in the civil law tradition the work is seen more as

[21] The distinction between internal and external form has been introduced by Joseph Kohler (1849-1919) on the basis of the classical argument developed by Johann Gottlieb Fichte in its 1793 essay "Proof of the Unlawfulness of Reprinting" (now available from *Primary Sources on Copyright (1450-1900)*, eds Lionel Bently & Martin Kretschmer, www.copyrighthistory.org.) On the concept of form in copyright law see Maurizio Borghi, "Owning Form, Sharing Content: Natural-Right Copyright and Digital Environment", in Fiona MacMillan (ed.) *New Directions in Copyright Law*, vol. 5, Elgar 2007.

[22] See Brad Sherman, "What Is a Copyright Work?" 12 *Theoretical Inquires in Law* (2011).

an act of one's own personality rather than an intangible object of property. This explains the existence of certain rights in relation to the work that are not at full disposal of the author, or that are simply inalienable. The notion of the work as an act instead as an object of property is rooted back in Kant's philosophy of right and his understanding of the "book" as a *public speech*.[23] As an object vested with certain rights, the work is essentially an act of communication between human beings, namely a speech addressed by an author to the public by means of a publisher. The latter is thus the mandatory of an act of speech over which the author maintains all rights and responsibilities. The right of integrity flows naturally from this understanding of the work as an inalienable act of speech, since distorting, mutilating or derogatorily treating a work is in a way the same as abusing one's own words.

From the stand point of the work as an object of property, the justification of the integrity right is not so straightforward. In particular, it is not clear why such right should be inalienable. Through the lenses of the utilitarian doctrine, in general, the fact that certain rights are inalienable is just a bizarre nonsense, since it would be in the interests of the author himself to have full disposal over the whole range of his rights. For instance, a "ghost writer" may monetise his voluntary status of "ghost" by transferring his paternity right to a third party; similarly, an author can allow any mutilations of his work upon payment. Yet the notion of the work as an act of communication between the author and the public provides a better understanding of the integrity right in its full meaning. This right does not simply reflect the author's self-interest in maximising the exploitation of his works. As a matter of fact, integrity is not even an author's interest only. It is rather, at the same time, in the interest of the public at large as well to have the words of the author addressed in a proper and comprehensive way. In Kant's terms, there is "a right of the public to a transaction with the author" which mirrors the right of the author to communicate his own thoughts to the public by means of publication. On this basis, it is reasonable that the right on work's integrity does not exhaust with the death of the author, not it expires once the work falls in the public domain.

Integrity and mass digitisation

From a legal point of view, the conversion of a work to a different format and its inclusion into a different context does not justify an alteration of the rights that apply to the work itself. This means that rights are neither expanded nor shrunk by modification of the form and context. For example, the Public Domain Charter of the Europeana makes clear that digitisation of content in

[23] See Immanuel Kant, *Die Metaphysik der Sitten. Rechtslehre* [*The Metaphysics of Morals. Doctrine of Right*], *in* 8 Werkausgabe Immanuel Kant 404 (Wilhelm Weischedel ed., 1968) (discussion under heading "What Is a Book?"). Kant addressed the issue of unauthorized reprint also in the 1785 essay *On the Unlawfulness of Book Reprinting*. *See* Immanuel Kant, *On the Unlawfulness of Reprinting* (1875), now available from Primary Sources on Copyright (1450-1900) (Lionel Bently & Martin Kretschmer eds., 2008), http://www.copyrighthistory.org.

public domain does not create new rights in it.[24] At the same time, digitisation does not shrink existing economic or moral rights. So, for instance, the same standard of integrity should apply when works are transplanted from "authorial" to "cyber" context, i.e. from physical to digital format. This point has been made in *Pink Floyd v EMI*, a UK case on the use of musical works originally recorded on physical support and subsequently exploited piece-by-piece as digital files and as ringtones. The High Court faced the problem of whether a contractual clause whereby the record label agreed not to sell physical albums of the rock band Pink Floyd "in any form other than as the current albums" and to exploit them "in exactly the same form" applied also to online digital distribution of the albums. The Court, observing that the clause had the undisputed purpose "to preserve the artistic integrity of the album", stated that "that purpose is as relevant to online distribution as to the sale of physical product". As a consequence, the record label had a duty "to treat online distribution so far as possible the same way as the exploitation of the physical product", and this duty was contravened by the selling of single tracks of the albums on streaming modality and of parts of tracks as ringtones.

The fragmentation of the work into separate pieces is one of the activities that potentially abridge the right of integrity in the cyber context. As far as mass digitisation projects are concerned, there are four main activities that may affect the work's integrity. These are, in order of complexity: digitisation, indexing, search and automatic association.

Digitisation

The conversion of a work into digital format leads inevitably to a more or less visible loss of quality, depending on the type of the work and on the technology used to digitise. A first straightforward issue is whether, and to what extent, such a loss amounts to an abridgment of the work's integrity. It must be observed that reproducing a work in lower quality does not per se amount to breach of integrity. For instance, reproduction of visual works in smaller size has not been found infringing in UK.[25] Similarly, the use of reduced size photographs as "thumbnails" in search engines has been considered to produce a "minimal loss of integrity" which does not pre-empt the finding of fair use in the US.[26] In a case on mobile ringtones, the German Supreme Court ruled that modifications due to the needs of technology do not violate the integrity of a musical works when the author has given his consent to that particular use.[27] One can safely assume that the

[24] http://www.europeana-libraries.eu/web/europeana-project/publications/
[25] *Tidy v Natural History Museum Trustees* [1995] 37 IPR 501.
[26] *Kelly v. Arriba Soft Corporation*, 336 F.3d 811(CA9 2003).
[27] BGH 18 Dec 2008, I ZR 23/06. Mira, pp. 357-361.

loss of quality which can be reasonably expected by the use of a particular reproduction technology does not amount to breach of integrity.

However, digitisation is not just a technology to "reproduce" works. It is also, and more importantly, a technology to make works usable in a certain way. Take the example of digitisation of books in mass digital projects. Here digitisation implies two distinct operations. First, the book is "scanned", namely it is reproduced and converted into a series of digital images. Second, the digital images are processed by optical character recognition (OCR) software, which converts the image into a text file.[28] The first operation creates a *copy* of the work, same as with other reproduction technologies, such as photography and reprography – the only difference being that the image can be displayed on screens and copied in hard drives. Here the integrity of the work may be violated only when the loss of quality is not imputable to the technology as such, or when the book page is not entirely or truthfully reproduced, or when watermarks such as the brand of the digital project are superimposed to the image.[29] The second operation, *i.e.* OCRing, is qualitatively different than mere reproduction. Its purpose is not to reproduce the work, but to make the copy readable by computers. Through ORing, an object which can be read (only) by humans is turned into a machine-readable entity, and the work can be processed by software. This processing enables operations that are qualitatively different than those of "perceiving" the work itself, such as reading the book, viewing the image or listening to the music track or ringtone. Examples of automated processing that normally follow scanning and OCRing of texts are the extraction of information in order to index the work and the making of the text at disposal of users for search on a word-by-word basis and for search of terms in context.

While scanning per se may only amount to loss of integrity in exceptional circumstances – namely when the loss of quality is not justified by technological necessity – OCRing creates the conditions for systematic breach of work's integrity by means automated processing. We consider these operations in the following.

Indexing

To index a digitised work means to provide the work with metadata to facilitate its identification and retrieval by search engines. Although this operation can be done manually, it is now increasingly performed with the aid of software that extracts information automatically not only from texts, but also from images and sounds. Indexing content is the key operation of search engines and it is preliminary of any use of digitised content in digital libraries and repositories.

[28] See *Infopaq International A/S v Danske Dagblades Forening*, Case C-5/08 [2009] E.C.D.R. 16, § 18-19.

[29] This example has been discussed in Adolf Dietz "Authenticity of Authorship and Work" 165 (Copyright in Cyberspace, ALAI Study Days 1997).

It is a well established principle that indexing and creating indexes of copyright works do not generally amount to copyright infringement nor to infringement of author's moral rights. In a French case of the pre-digital era, the Court de Cassation ruled that the indexing of journal articles by selection of keywords and the making of short *résumés* does not infringe copyright in the articles and does violate their integrity, insofar as these activities are functional to a "purely informative" purpose and represent "the expression of free choice of the author of the secondary work".[30] It is important to note that the "free choice" of the subsequent author – that is, the author of the index – must be clearly recognised in the expressive form of the index itself, namely in the selection of keywords and in the composition of *résumés*. Failing this element, the index may not be an independent work of authorship but may unduly overlap with the indexed work. For instance, the selection of keywords could be perceived as an "objective fact", and the *résumé* may be understood as expressing the view of the author of the indexed work. By contrast, both the keywords and the *résumés* are the outcome of a choice which may – but does not have to – coincide with the intended purpose of the original work.

Moreover, according to the French court, the index must be "purely informative", that is it must have the sole purpose of informing the public about the existence of a certain work. It has not to replace the original work in any way.[31]

The combination of these two elements – the "informative" and the "free choice" element – provides a good set of criteria to test whether indexing in mass digitisation projects violates the work's integrity. The problem that might arise is when metadata on books and on other content are too poor or too inaccurate to properly indentify the work, or are such that the work is systematically identified in an improper way.[32] This problem is further enhanced when metadata are not created by some expenditure of independent skill and judgement, but are automatically generated via data mining on books or on other content. Is there still an element of "free choice" in the way by which the indexed works are presented? And if not, how does one has to classify, from a copyright perspective, the relationship between the index and the indexed work? This question is particularly relevant when referred to the activity that is commonly associated with indexing in mass digitisation projects, namely searching.

Search

[30] *Le Monde c. Microfor*, Cour de Cassation, 30 octobre 1987, 86-11918.

[31] See also *Ty Inc. v Publications International Ltd*, 292 F.3d 512 (7th Cir. 2002) (distinction between "complementary" and "substitutional" copying in the fair use analysis).

[32] For a detailed analysis of Google Book's metadata see Geoffrey Nunberg "Google Books: a Metadata Train Wreck" < http://languagelog.ldc.upenn.edu/nll/?p=1701>.

Once digitised and indexed, works are available for searching within search engines. As discussed earlier, in the example of books searching is made possible by OCRing, a process by virtue of which printed pages become searchable "inside", that is on a word-by-word basis. This processing of the book's pages does not represent *per se* a violation of the integrity of the work which is fixed in the printed medium. This is because, as a technical operation, OCRing modifies only the format in which the work is fixed but does not touch upon the work's form or even the work as such. Modifications that are technically necessary in this respect do not amount to infringement of moral rights. Nevertheless, OCRing creates the *conditions* for possible violations of integrity on a large scale. The case of Google Books is illustrative of this situation. Here books and other literary works are digitised and made searchable by users, which can search inside books on a word-by-word basis and view short excerpts or "snippets" in response of their queries. By entering a search query, subject to display are excerpts from a book that contain that particular word or string of words. Moreover, excerpts are displayed alongside excerpts from other books that contain the same word or string. And this is not the only case. The large corpus of digitised books is at the same time integral part of the large Google's database which include web pages, news, geographical locations and virtually all "world's information".[33] This means that excerpts from a book are made available not only to users that search inside that particular book, and not even only to users that search into the corpus of Google Books, but to any user who enter a search query in the Google search engine, and those excerpts are displayed alongside excerpts from any other web resource.[34] Accordingly, every search query generates automatically a fragment of the work which in turn is presented in a context which is determined by the algorithm of the search engine. Is there in this practice violation of integrity?

In Europe, two courts responded in the affirmative.[35] In *Google v Copiepresse*, the Belgian Court of first instance of Brussels faced the problem of whether the service Google News, which provides internet users with an automatic selection of news items from the web servers of the written press, infringes copyright in newspapers articles. *Inter alia* the court examined the closely related issues of the applicability of the exception for quotations and of the infringement of the moral right of integrity. As to the first point, the court correctly argued that under Belgian law – but

[33] As it is well known, the Google's corporate mission is to "organize the world's information and make it universally accessible and useful" (<http://www.google.com/corporate>.

[34] It can be observed, incidentally, that books forming the corpus of Google Books are only available to Google search engine, and do not feature in the results of other search engines like Yahoo! or Bing.

[35] *Google Inc v Copiepresse SCRL*, Court of First Instance of Brussels, [2007] E.C.D.R. 5; *Editions du Seuil et autres c. Google Inc et France*, Tribunal de grande instance de Paris 3ème chambre, 2ème section Jugement du 18 décembre 2009.

similar finding should be made in most European jurisdictions[36] – the "right to quotation" does not apply when excerpts of works are extracted and grouped automatically, without human intervention.[37] To the court, a "quotation" is meant to be used in a subsequent work to illustrate a proposition or defend an opinion; by contrast, Google "does no more than incorporate the 'quotations' and owes its substance solely to the extracts of works reproduced, which is contrary to the spirit of the right to quotation".[38] Given that Google's excerpts are not "quotations", their automatic extraction and grouping may violate the integrity of the work. In this respect, Google contended that there is no damage to "the integrity of the work where the quotation of a text is to be found with another quoted text or photos" in that "the internet user knows very well that it is dealing with a quotation, and sees the original text each time in its original context by clicking on a hyperlink".[39] The court disagreed and observed that "the circumstance that the internet user is aware that it is only dealing with a fragment of the work does not seem relevant in relation to respect for the integrity of the work." This is because "Google groups the different extracts of articles, which can originate from any source, by theme, so that *the editorial or philosophical approach espoused by the author may be altered*".[40]

This conclusion can easily apply to mass digitisation projects, and in particular to projects like Google Books where the digitised works form part of larger search engines databases. This is confirmed by the French decision in *Editions du Seul c. Google France*, where the court found that "the display of excerpts from works that Google Inc. recognises to be truncated randomly and in form of ripped banner of paper undermine the integrity of the works".[41] In the context of mass digitisation projects like Google Books, works are increasingly exposed to the possibility of being associated with, or even linked to, content originating from different sources. Such automatic association creates further conditions for abridgement of integrity.

---

[36] The exception for quotations is harmonised by art. 5(3)(d) of the Directive 2001/29/EC, which reads: "Member States may provide for exceptions or limitations to the rights provided for in Articles 2 [Reproduction] and 3 [Communication to the public] in the following cases:

[...]

(d) quotations for purposes such as criticism or review, provided that they relate to a work or other subject-matter which has already been lawfully made available to the public, that, unless this turns out to be impossible, the source, including the author's name, is indicated, and that their use is in accordance with fair practice, and to the extent required by the specific purpose."

The exception for quotations is not enacted in the Czech Republic and in Slovenia. In Ireland and the UK there is no express mention to quotes and the quotation exception falls under fair dealing for criticism and research.

[37] *Google v Copiepresse*, § 130.

[38] *Google v Copiepresse*, § 131. In *Editions du Seuil c. Google* the French court affirmed that the exception for "short quotations" (Code de la Propriété Intellectuelle, art. L. 122-5) cannot apply to the display of snippets in Google Books.

[39] Id. § 159.

[40] Id. §§ 160-161 (emphasis added).

[41] *Editions du Seuil c. Google*.

Automatic association

As seen in *Copiepresse*, integrity is violated when the work is displayed in contexts where the "editorial or philosophical approach espoused by the author" is altered. It is a well established principle that the author has a right to object publication or re-publication of his work in contexts that do not fit the intended purpose of the work. In the USA as well, where the moral right of integrity is not codified, the contributor of a collective work has a right not to have his work republished separately or in other collective works that do not belong to "the same series".[42] Although the rationale of this norm is to protect the economic interest of the author, the result is that the author can object any alteration of the context in which the work is published.

In case of mass digitisation projects, works are *de facto* republished in contexts where the original editorial approach cannot be easily reproduced. Here again a distinction must be made between what is technologically necessary and what exceed this necessity. A digital library is by definition an environment which is different than any context in which works as such or physical copies thereof are normally accessed, such as galleries or libraries.[43] This, however, does not exempt digital repositories from a duty to present works in an environment which respects the editorial and philosophical approach of the work itself.

The typical case in which integrity of the work may be violated is that of "bad association", namely when a work is associated with other information in a derogatory way. For instance, a book containing explicit language or an art photograph of a nude model could be automatically presented in connection with pornographic content or linked to information which has nothing to do with the book or of the photograph. The fact that these associations are generated automatically, that is without human intervention, does not mean that there cannot be violation of the work's integrity. In recent cases on search engines, French and Italian courts have found that associations generated by Google's "autocomplete"[44] are capable of determining defamation.[45]

More generally, mass digitisation projects present works in "reading environments" that are increasingly shaped by automatically generated associations. While the fact of associating works with other works is a natural one, and it is the principle upon which libraries are made, its implementation by digital technology may lead to drawbacks. As put forward in an opposition to the Google Books Settlement Agreement, the systematic lack of human control may cause works to

[42] 17 U.S.C. § 201(c). See *New York Times Co. v Tasini*, 533 U. S. (2001).

[43] For a discussion of the difference between libaries and digital libraries in the context of user's privacy see Elisabeth A. Jones & Joseph W. Janes, "Anonymity in a World of Digital Books: Google Books, Privacy, and the Freedom to Read", 2(4) *Policy & Internet*, 43 (2010).

[44] On the functioning of Autocomplete see <http://www.google.com/support/websearch/bin/answer.py?answer=106230>

[45] M. X... /Google Inc., Eric S. et Google France, TGI Paris 17ème chamber, 8 septembre 2010; Mme C. / Google France et Inc., TGI Montpellier, 28 octobre 2010; Tribunale di Milano, ordinanza 31 marzo 2011.

be "associated with authors' work in ways that are objectionable, offensive, or harmful to the author".[46] For example, automatic associations may be generated as effect of "groups attempting to create a false appearance of endorsement".[47] More subtly, behaviours of groups of readers which use an author's work for own purposes, for instance to support own political views, may influence algorithms so that the author will be automatically associated with those views. Moreover, computational analysis on corpuses of millions of books and of other digitised works, coupled with aggregated data mining, are able to associate an author with other authors, information, products, or even "ideas" independently from their expression. For instance, a book may be displayed alongside a list of "key ideas" as extracted via automatic text processing, and the reader may be re-addressed to authors that allegedly share "similar ideas" based on associative and taxonomic data analysis.[48] A range of "proxies" can be associated with works, spanning from snippets, quotes, cross-referencing or even "key ideas". Are all these associations, which potentially violate the work's integrity, justified by the "technological necessity" of transplanting a work from physical format to the digital environment?

A critical issue is whether association with advertisements is generally compatible with the "editorial and philosophical" approach of works that are hosted by a digital library. In commercial digitisation projects, "discrete advertisement" may be displayed in connection with books or articles. This may be in contrast with the editorial approach that underlies most of the works that are hosted by the service. It is no accident if advertisement is normally not included in book's pages, or if scientific journals who host advertisement pay particular attention not to place it alongside articles. These are forms of respect for the work that are deeply rooted in the publishing practice, and any alteration of these forms which exceed what is technologically necessary to digitise the work is a potential threat to the integrity of the work.

Conclusion

Mass digitisation has become paramount in the online world as the activity by which the cultural heritage of humankind is digitised in bulk to form part of the collection of online archives, repositories and digital libraries. While most of the rights pertaining to the digitised works, and particularly the economic rights, may be expired, certain moral rights are perpetually in force in most European jurisdictions. As we have seen, the moral right of integrity does not merely protect

[46] Objections of Arlo Guthrie, Julia Wright, Catherine Ryan Hyde, and Eugene Linden to Proposed Class Action Settlement Agreement, *Author's Guild, Inc. v Google, Inc.*, No. 1:05-CV-08136 (2 September 2009), p. 12

[47] Id.

[48] See Bill N. Schilit & Okan Kolak "Exploring a Digital Library through Key Ideas", Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries (Pittsburgh, Pennsylvania, USA, June 16-20, 2008), available at http://sites.google.com/site/schilit2/fp035-schilit.pdf

an author's interest, but is representative of a more general interest of the public at large. As a matter of fact, it is in the interest of the whole society to preserve the integrity of works that form part of our cultural heritage, and in this respect the integrity right can be seen as a bulwark of other more fundamental societal interests. If this is the case, then the question arises as to whether, and by which means, integrity is threatened by mass digitisation.

It has been rightly observed that rights are technology neutral, and violations of moral rights must be determined in similar way irrespective of whether they are performed with digital or analogue technologies.[49] Similarly, the rights pertaining to works are not altered when they are transplanted from the analogue to digital environment. However, mass digitisation projects brings about a technological shift in the way in which works of authorship are used: works are no longer dealt with with human intervention only, but are increasingly subject to automated processing. Such processing inevitably modifies both the work itself and the context in which the work is displayed to the public. Not all modifications introduced by automated processing of works are as apparent as those that occur by means of human intervention. Still, they are all but irrelevant. In particular, they are not all justifiable on technological grounds. Here again, the legal challenge is to draw a distinction between modifications that are "technologically necessary" and modifications that exceed technological necessity.

[49] Salokannel & Strowel, p. 206.