Digital content convergence: Intellectual property rights and the problems of preservation: A US perspective

John N. Gathegi

School of Library and Information Science, University of South Florida
4202 E. Fowler Ave, CIS1040, Tampa, FL 33620
jgathegi@cas.usf.edu

Abstract

One of the issues that this conference explores is the continuing phenomenon of convergence of communication, caused in part by the convergence of media and digital content. In this paper, we will review some of the intellectual property challenges that loom in this environment, with an emphasis on the situation in the Unites States. We shall discuss some of the peculiar features inherent in digital content that exacerbate the intellectual property problem, such as non-permanence, multiple, heterogeneous. We shall examine a couple of cases that illustrate some of the problems in this area. We shall then conclude with the problem of intellectual property and the multiple goals of digital content collections.

Keywords: digital rights, intellectual property, digital content

1. Introduction

The literature continues to indicate a continuing phenomenon of convergence of communication, caused in part by the convergence of media and digital content. We will review in this paper some of the intellectual property challenges that loom in this environment, with an emphasis on the situation in the Unites States. Some of the peculiar features inherent in digital content that exacerbate the intellectual property problem, such as non-permanence, multiple, and heterogeneous media, will be discussed. We shall examine a couple of cases that illustrate some of the problems in this area. We shall then conclude with the problem of intellectual property and the multiple goals of digital content collections.

2. The problem of non-permanency of dynamic content

Digital content has, unlike its print counterpart, some unique features that present challenges in both development and management, especially from a legal perspective. We shall examine three such features in this paper, the proposition that digital content (1) is dynamic, (2) may suffer issues of non-permanence, and (3) may have more than one media format.

Digital content is dynamic. As the need arises, items are constantly added and corrections and modifications made to specific files and databases. This means that a file may change from day to day. The problem then emerges on how to preserve digital content and vouch for its integrity.

Preservation efforts face many legal problems. The primary problem is how to ensure the non-infringement of copyright, by avoiding unauthorized exercise of the authors' exclusive rights, as well as determining what content is protected by copyright, to facilitate access to content as well as consent from copyright holders. A persistent question is whether the digital content manager still has the necessary rights in the e-content. Also, issues of privacy and confidentiality may be raised by the dynamism and non-permanence of digital content, as may ethical issues in health and personal data [1].

Despite perceptions to the contrary, "digital information is in fact fragile and at risk." Changes in technology can render some digital files corrupt and unreadable [2]. The longer the time frame required for future access, the more the uncertainty with information preservation. Challenges include changes in format, data definitions, and metadata content [1]. The format problem is exacerbated by the fact that many formats are proprietary and continue to evolve into more complex versions with newer features and functions, sometimes 'orphaning' earlier versions [3]. Legal access problems can occur when a proprietary owner contractually limits access or goes out of business [4].

One way of handling format changes in digital preservation is migration of data, both in terms of software and hardware. This will sometimes involve re-arranging structural and data elements sequence [3]. Two copyright problems arise. First, the act of migration usually will involve copying of the information, which may be an infringement of the author's exclusive reproduction right. Also, the re-arrangement of the structural and data elements may trigger the trampling of another right: the author's exclusive right to make derivative copies. Permission to migrate may have to be sought from the copyright holder. Other issues include, in the case of the

United States, whether a file conversion would be a violation of the Digital Millennium Copyright Act, and whether, for evidentiary purposes, a migrated file is the same as the original file.

3. Multi-media content and its complexity

As well as being dynamic and raising the problem of non-permanency, digital content may also contain a mixture of different media formats, including text, sound, graphics, video, and a variety of other file formats.

Good examples of multimedia digital content are electronic books, or e-books. An e-book could have, for example, an article about a country, a video about parts of the country, and a sound file of examples of music from the country.

E-books are usually in proprietary devices, but may also be accessible through a central server. An owner of an e-book collection has many of the similar features to a publisher of any other digital content in terms of the susceptibility of the content to be easily copied. Digital rights management technology is used to control access to e-book content that is copyright protected, to preserve the copyright owner's exclusive rights. There are, however, e-books available free of copyright protection that a digital content manager can link to from the digital collection [5]. E-book aggregators, such as netLibrary (a division of OCLC), provide access to a digital library's e-content on a 24/7 basis by negotiating intellectual property rights with publishers to provide access to content hosted on their servers. Aggregators usually provide their own digital rights management technology, thus easing legal issues for the digital content manager [6].

"Stocking" or publishing e-books in a digital collection requires that the digital collection manager understand the access limitations that come with the digital rights management, and the different pricing models. These models can range from outright purchases (much like print versions) to limited time and number of persons per access, and may also come with use restrictions that define practices such as printing, downloads, and amount of content that can be accessed.

Legal issues in this area are complicated by the fact that some media formats are covered by rules specific to the media (e.g. sound files). Also, conversion of media from one format to another may trigger copyright infringement (e.g. conversion of text into audio formats). More so than text works, dates on which a sound recording was first fixed determine the nature of the legal protection available in the United States. For example, no federal

copyright protection was available to sound recordings prior to February 15, 1972, but the Sound Recording Amendment Act of 1971, rectified the situation by providing federal copyright protection to works recorded or fixed after that date [7]. Pre-1972 works, however, may be protected by state criminal law statutes or common law, against unfair competition or misappropriation, until February 15, 2067 [8].

Page thumbnails and document icons

Other newer versions of familiar formats, such as document icons and page thumbnails, present new legal issues.

Document icons are small visual representations of documents [9]. Icons can include information about a document format or genre (e.g., pdf document, web page or folder). Page thumbnails, on the other hand, are small images of a page usually in reduced resolution, that can be enlarged by a reader for viewing.

In discussing thumbnails, two rights that are exclusive to the copyright holder are implicated. Because they make copies of the images they crawl, search engines may violate the author's exclusive right to make reproductions of a work [10]. Also, because the thumbnails are shown to the users, search engines may also violate the author's exclusive right to public display [10]. However, the use of thumbnails may rely, as we see in the Kelly v. Arriba [11] and Perfect 10 cases below, on one of the exemptions to the author's exclusive rights: fair use.

In Kelly v. Arriba Soft Corp., a photographer whose copyrighted images were displayed by a visual search engine operator on the operator's web site and those that it had licensed sued the operator. The operator had built its database by copying images from web sites and reducing these images into "thumbnails" that could be enlarged by clicking on the thumbnail. The lower court ruled the operator's use of the thumbnails fair use, as the character and purpose of its use was "significantly transformative and the use did not harm the market for or value of [the photographer's] works" [11]. The 9th Circuit affirmed the lower court's ruling that the display of thumbnails was fair use.

In Perfect 10 v. Google [12] it was a website operator's turn to sue Internet search engines. Perfect 10 published adult photographs in both a magazine and a web site, and had expended considerable resources to the development of the brand name for the magazine and web site. Google and Amazon, search engine operators, have an image search function that retrieves thumbnail images in response to a textual search string query. Some of the images so retrieved came from Perfect 10's website, and it sued both

Google and Amazon. The district judge, when considering plaintiff Perfect 10's motion for an injunction against Google, put the issue in a perfect context:

> The principal two-part issue in this case arises out of the increasingly recurring conflict between intellectual property rights on the one hand and the dazzling capacity of internet technology to assemble, organize, store, access, and display intellectual property "content" on the other hand. That issue, in a nutshell, is: does a search engine infringe copyrighted images when it displays them on an "image search" function in the form of "thumbnails…?" [12]

The district court was of the view that Perfect 10 was likely to succeed in its claim that the display of thumbnails was a direct infringement by Google of its copyrighted images, and issued a preliminary injunction from creating and displaying Perfect 10's images. The district court distinguished Arriba's use of thumbnails in Kelly, in that Perfect 10's market for downloading reduced-size adult thumbnails into cell-phones was superseded by Google's use of Perfect 10's thumbnails. However, the 9th Circuit later ruled that the thumbnails were fair use because they did not detract from the economic value of the images, and thus Google could continue displaying Perfect 10 thumbnails that came up following a search [13].

4. Multiple, heterogeneous content: the legal complexities

A related feature to the complexity of digital content discussed above is the multiple, heterogeneous nature of digital collections. This feature can usefully be analyzed in two parts: the different types of digital collections, and the different goals of digital collections. Digital collections often have content of different types designed to meet a variety of goals.

The Internet, for example, can be viewed as one giant digital collection; a sort of a "meta-collection." Individuals, libraries, and other organizations often take subsets of the Internet to form specific collections. This is generally done through book-marking and linking. While there have been no legal challenges yet to bookmarking, linking has generated some legal issues, less so in the United States than in some European countries.

Likewise, libraries and other organizations may have, as part of their digital collections, commercial databases. The major issues here involve copyright protection and licensing issues. Data sets, on the other hand, may

be viewed in the same category as commercial databases, with less emphasis on copyright and more emphasis on licensing issues.

One type of content, collective works and compilations requires a more extended discussion here. Section 101 of the U.S. copyright code defines a compilation as "a work formed by the collection and assembling of preexisting materials or of data that are selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship. The term "compilation" includes collective works." [14].

Just to make sure it is understood that compilations fall under the subject matter of copyright specified in section 102, section 103 explicitly declares that compilations are indeed included, but points out that protection extends only to what the author has contributed, and not to the underlying or pre-existing material. Nor does it extend to preexisting material that has been used unlawfully [15].

Collective works and compilations may or may not have common characteristics. In a collective work, individual components are generally independent copyrightable works, while compilations may include material that is not necessarily copyrightable [16]. Separate contributions to a collective work can have copyright protection that is distinct from copyright in the collective work as a whole. Owning a copyright in a collective work entitles the copyright owner to "only the privilege of reproducing and distributing the contribution as part of that particular collective work, any revision of that collective work, and any later collective work in the same series" [17]. The court in New York Times Co., Inc. v. Tasini [18] explored the question of whether a copyright owner in a collective work who republished all or a part of the compilation in an electronic database could prevail against an assertion of copyright infringement from the author of a contribution in the compilation. The case involved freelance writers who had sued a newspaper publisher for making their articles available in electronic databases. The newspaper asserted a privilege offered by section 201(c) of Title 17. Under section 201(c):

> Copyright in each separate contribution to a collective work is distinct from copyright in the collective work as a whole, and vests initially in the author of the contribution. In the absence of an express transfer of the copyright or of any rights under it, the owner of copyright in the collective work is presumed to have acquired only the privilege of reproducing and distributing the contribution as part of that particular collective work, any revision of that collective work, and any later collective work in the same series. [17]

The Supreme Court focused on the perception of a user of the articles as presented in the database, and rejected the newspaper's reliance on the section 201(c) privilege. The privilege, however, continues to be available given the right circumstances. A second circuit court, for example, affirmed the granting of summary judgment to the National Geographic publisher who had made a searchable digital collection of past issues of the magazine (dubbing it the Complete National Geographic), against freelance authors and photographers who had sued the magazine for the use of their work in this new medium [19]. The Court relied on its holding the fact that the original context of the magazine was present in the new Complete National Geographic, and that the digital work was a new version of the National Geographic Magazine. The database at issue in Tasini, on the other hand, did not allow users to view the underlying works in their original context.

Divergent goals

As well as being heterogeneous and having multiple formats, digital collections also have different goals. One of the goals of digitization, as mentioned above, is preservation. Some institutions have a legal privilege to preserve. Section 108 of the U.S. Copyright Code, for example, addresses the need for preservation and conservation [20]. Legal issues that are likely to arise here not only include copyright, but also evidence.

The issue of non-permanence that we discussed above acquires critical importance when it comes to maintaining documents for legal evidentiary purposes. In December 2006, changes were put into effect in the U.S. Federal Rules of Civil Procedure that institute a new category of evidence: the Electronically Stored Information (ESI), which is designed to work within the existing rules of production of "documents." Under the rules, a party must provide to other parties: "a copy--or a description by category and location--of all documents, electronically stored information, and tangible things that the disclosing party has in its possession, custody, or control and may use to support its claims or defenses…" [21]. While the rule does not specify the version of the electronically stored information that should be produced, Rule 26(f) does oblige the parties to conference and "…discuss any issues about preserving discoverable information" as well as "any issues about disclosure or discovery of electronically stored information, including the form or forms in which it should be produced."

A closely related issue to non-permanence for evidence is authenticity. Digital information can be vulnerable to tampering or corruption. Depending on the nature of the collection, authentication methods such as digital signatures, version control, and encryption may be necessary [22].

Finally, we cannot conclude without mentioning something about access. Access and preservation are much intertwined. There could be any number of reasons for seeking access, including for entertainment, research, or safeguarding culture. Copyright is always an issue, as, for example, not all copyrightable works have the same protection duration. Different publications are covered under different copyright protection terms, depending on when they were created or published. However, the issues most likely to rise are those of licensing for access. By access, we are also here referring to use. Unfortunately, this issue is for the moment outside the scope of this paper.

Notes and References

[1] ANDERSON, WL. Some Challenges and Issues in Managing, and Preserving Access to, Long-Lived Collections of Digital Scientific and Technical Data. Data Science Journal 3, 2004, p. 191-201. Available at http://www.jstage.jst.go.jp/article/dsj/3/0/191/_pdf (June 29, 2009).

[2] ADAMS, G. Partners Go Dutch to Preserve the Minutes of Science. Research Information, September/October, 2004. Available at http://www.researchinformation.info/risepoct04archiving.html (March 10, 2010).

[3] ARMS, CR. Keeping Memory Alive: Practices for Preserving Digital Content at the National Digital Library Program of the Library of Congress, RLG DigiNews, 4(3) 2000, Available at http://www.rlg.org/preserv/diginews/diginews4-3.html#feature1

[4] BARNES, I. *The Preservation of Word Processing Documents.* Canberra: Australian Partnership for Sustainable Repositories, 2006. Available at http://www.apsr.edu.au/publications/word_processing_preservation.pdf (June 29, 2009).

[5] ESPOSITO, J. The Processed Book, First Monday, 8 (3) 2003 (Updated 2005). Available at http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1038/959 (March 25, 2010)

[6] GARROD, P. Ebooks in UK Libraries: Where Are We Now? Ariadne, no. 37 (2003). Available at http://www.ariadne.ac.uk/issue37/garrod/

[7] BESEK, J.M. *Copyright Issues Relevant to Digital Preservation and Dissemination of Pre-1972 Commercial Sound Recordings*. Washington,

DC: Council on Library and Information Resources and Library of Congress, 2005. p. 16-17.

[8] 17 USC 301(c) provides: "With respect to sound recordings fixed before February 15, 1972, any rights or remedies under the common law or statutes of any State shall not be annulled or limited by [Title 17] until February 15, 2067."

[9] JANSSEN, WC. *Document Icons and Page Thumbnails: Issues in Construction of Document Thumbnails for Page-Image Digital Libraries*. Palo Alto, CA: Palo Alto Research Center, 2004

[10] 17 USC 107

[11] 280 F.3d 934 (9th Cir. 2002)

[12] 416 F. Supp. 2d 828 (C.D. Cal. 2006)

[13] PERFECT 10, Inc. v. Amazon.com, Inc., 508 F.3d 1146 (9th Cir. 2007).

[14] 17 USC 101

[15] 17 USC 102, 103

[16] ROY Export Co. Establishment of Vaduz, Liechtenstein, Black Inc., A. G. v. Columbia Broadcasting System, Inc., 503 F. Supp. 1137 (S.D. N.Y. 1980), judgment aff'd, 672 F.2d 1095 (2d Cir. 1982)

[17] 17 USC 201(c)

[18] NEW York Times Co., Inc. v. Tasini, 533 U.S. 483

[19] FAULKNER v. National Geographic Enterprises Inc., 409 F.3d 26 (2d Cir. 2005)

[20] 17 USC 108

[21] FEDERAL Rules of Civil Procedure 26(a)(1)(A)(ii)

[22] GERMAIN, CM. Legal Information Management in a Global and Digital Age: Revolution and Tradition, International Journal of Legal Information, 35 (1) 2007. p.134-163, 156-157