

Fordham Intellectual Property, Media and Entertainment Law Journal

Volume 18 *Volume XVIII*
Number 3 *Volume XVIII Book 3*

Article 5

2008

The Mereology of Digital Copyright

Dan L. Burk
University of Minnesota Law School

Follow this and additional works at: <https://ir.lawnet.fordham.edu/iplj>



Part of the [Entertainment, Arts, and Sports Law Commons](#), and the [Intellectual Property Law Commons](#)

Recommended Citation

Dan L. Burk, *The Mereology of Digital Copyright*, 18 Fordham Intell. Prop. Media & Ent. L.J. 711 (2008).
Available at: <https://ir.lawnet.fordham.edu/iplj/vol18/iss3/5>

This Article is brought to you for free and open access by FLASH: The Fordham Law Archive of Scholarship and History. It has been accepted for inclusion in Fordham Intellectual Property, Media and Entertainment Law Journal by an authorized editor of FLASH: The Fordham Law Archive of Scholarship and History. For more information, please contact tmelnick@law.fordham.edu.

The Mereology of Digital Copyright*

Dan L. Burk[†]

INTRODUCTION

The development of the Internet as a means of communication has facilitated widespread access to a vast array of digitized works in a variety of electronic formats.¹ Increased access to such digitized works has heightened the need for robust systems that can identify and index online resources in order to allow users to locate and access the new wealth of digitized materials in what amounts to a global virtual library.² Cataloging and indexing has always been critical to library functions, but never more so than in the decentralized, emergent library that constitutes the Internet.³ Consequently, search engines such as the Google⁴ database have developed into key tools for facilitating access to online resources;

A PDF version of this article is available online at <http://law.fordham.edu/publications/article.ihtml?pubID=200&id=2735>. Visit <http://www.iplj.net> for access to the complete Journal archive.

* Copyright 2007–08 by Dan L. Burk

[†] Dan L. Burk holds the Oppenheimer, Wolf and Donnelly Professorship at the University of Minnesota Law School, where he teaches courses in Patent, Copyright, and related topics. Professor Burk holds a B.S. (1985) in Microbiology from Brigham Young University, an M.S. (1987) in Molecular Biology and Biochemistry from Northwestern University, a J.D. (1990) from Arizona State University, and a J.S.M. (1994) from Stanford University.

¹ See Paula Samuelson, *Fair Use for Computer Programs and Other Copyrightable Works in Digital Form: The Implications of Sony, Galoob and Sega*, 1 J. INTEL. PROP. L. 49, 108 n.246 (1993).

² See Eric Goldman, *Search Engine Bias and the Demise of Search Engine Utopianism*, 8 YALE J. L. & TECH. 188, 189 (2006).

³ See Lucas D. Introna & Helen Nissenbaum, *Shaping the Web: Why the Politics of Search Engines Matters*, 16 INFO. SOC'Y 169, 170 (2000).

⁴ Google Home Page, <http://www.google.com> (last visited Jan. 2, 2008).

if the resource is not indexed via such a database, it effectively ceases to exist.⁵

Such control over access to digital resources implicates control over the use and disposition of those resources.⁶ Information cannot be used until it is found. In the physical world, control of information has been incident to physical or legal ownership that determines access or exclusion. Works of creative authorship have long been subject to ownership under the rubric of copyright law, which offers certain defined exclusive rights as an incentive for creation and publication of expressive works.⁷ However, the copyright system that developed in a world of hardcopy print is challenged both by the technology of digitization and by the construction of metadata indexes for digitized works.⁸ Rules of ownership developed to control access to atoms apply only uncertainly when used to control access to bits.⁹

The troubled interplay between copyright and digital expression is nowhere more apparent than in the controversial Google Book Search¹⁰ project, where search engine technology intersects with indexed databases scanned into electronic format from hardcopy materials.¹¹ In this Article, I analyze the copyright status of this project, using it as a vehicle to develop certain themes that are emerging as fundamental issues in the copyright of digitized texts. I begin by describing the Google Book Search project, touching briefly upon the legal rationale relied upon by Google for scanning copyrighted works into its database without permission of the copyright holders. I then move to the issue that has received less attention: the copyright status of the metadata relational database that is core of the project. This database, I argue, is emblematic of the broader issues facing copyright in an age of digitization, and I discuss several cases that bear upon the

⁵ See Introna, *supra* note 3, at 171.

⁶ See *id.* at 169–75.

⁷ 17 U.S.C. §§ 101–1332 (2000).

⁸ See Emily Anne Proskine, *Google's Technicolor Dreamcoat: A Copyright Analysis of the Google Book Search Library Project*, 21 BERKELEY TECH. L.J. 213, 213 (2006).

⁹ See *id.*

¹⁰ Google Book Search, <http://books.google.com> (last visited Jan. 2, 2008).

¹¹ About Google Books, <http://books.google.com/intl/en/googlebooks/about.html> (last visited Jan. 2, 2008).

legal status of such a meta-database. I conclude by sketching the challenges that copyright law will need to encompass as works of authorship move from fixation as atoms to recordation as bits.

THE GOOGLE BOOK PROJECT

The Google Book Search project is an ambitious—even audacious—attempt to make available via the Internet a searchable database of texts previously available only as bound, printed matter.¹² The project uses Google search engine technology as the technical vehicle to provide storage, indexing, and retrieval of texts that have been scanned from printed format to digital format.¹³ But the scanning process requires access to the physical printed texts. This has been accomplished via two complementary initiatives.¹⁴ The first of these initiatives is the relatively straightforward “publisher program.”¹⁵ In this component of their overall strategy, Google seeks the cooperation and permission of publishers for inclusion of books in the database, and many publishers see value in participating. Publishers in this part of the program provide copies of books to be scanned or, if available, they provide electronic files of book text, under terms agreed upon with the publisher.¹⁶ This portion of the overall Google agenda is relatively uncontroversial; scanning and inclusion of books within this portion of the database is done with permission and approval of the publishers, who generally hold the copyright, and who can set whatever terms of use they choose.¹⁷

What has been more controversial is the library scanning portion of the Google effort. In this portion of the project, Google

¹² See Jonathan Band, *The Google Print Library Project: Fair or Foul?*, 9 J. INTERNET L. 1, 1 (2005).

¹³ See *id.*

¹⁴ See *id.*

¹⁵ Partner Program, https://books.google.com/partner/?hl=en_US (last visited Jan. 2, 2008); see also Book Search Intro, http://books.google.com/intl/en-US/googlebooks/book_search_tour/index.html (last visited Jan. 2, 2008).

¹⁶ Help Users Find Your Books, http://books.google.com/googlebooks/book_search_tour/books2.html (last visited Feb. 11, 2008).

¹⁷ See Band, *supra* note 12.

has entered into agreements with several libraries,¹⁸ including those of Harvard University, the University of Michigan, the New York Public Library, and Oxford University, to gain access to and to scan all or part of the library's holdings.¹⁹ The exact terms of the agreements are undisclosed and appear to differ from library to library, but it is clear that at least some of these agreements will result in scanning of books for which the library owns the physical volume, but for which the library does not have the right to make or authorize making of copies, including scanned digital copies.²⁰

In both portions of the project, book pages are scanned to produce high quality, but not archival quality, images.²¹ The scanned images are then parsed by Google search technology.²² At present, Google is keeping the original images stored, and has agreed to provide copies of the original images back to the library from which the book was scanned.²³ But the stored images

¹⁸ Library Partners, <http://books.google.com/googlebooks/partners.html> (last visited Jan. 2, 2008).

¹⁹ *Id.*

²⁰ See Can I see a copy of a contract you have with a library partner?, Google Book Search Help Center, <http://books.google.com/support/bin/answer.py?answer=43752&query=contracts> (last visited Jan. 2, 2008); Suzanne Bresson Bis, *Book Search Is Beautiful?: An Analysis of Whether Google Book Search Violates International Copyright Law*, 17 ALB. L.J. SCI. & TECH. 271, 281 (2007).

²¹ "If you look at what Google does, it's really a bitonal representation. It's as if the book were brand new, which is just to say that the page is white [and] the ink for the font is black. Whereas if you look at the Microsoft [Windows Live Book Search] presentation, it's a color image, so you get the sense of it as an artifact." Dian Schaffhauser, *Google Book Search: The Good, the Bad, & the Ugly*, Campus Technology, Jan. 1, 2008, at 7, available at <http://www.campustechnology.com/articles/57064> (quoting Linda Becker, Vice President, Marketing for Kirtas, a digitization technology solutions company). Google has not released any specific information about the exact scanning system it uses for the Book Search project. Its engineering director, Dan Clancy, has stated that the company developed its own scanning process rather than use a commercial scanning solution. *Id.* at 4–6.

²² The digitization process involves the parsing of the scanned image and converting that information into "other file formats for online viewing." *Id.* at 7. Therefore a user can "just enter the keyword or phrase . . . into the Google Book Search box" and Google will return "snippets—of [the] search term in context." How does Google Book Search work?, Google Book Search Help Center, <http://books.google.com/support/bin/answer.py?answer=43724&topic=9259> (last visited Jan. 15, 2008).

²³ "[T]he library will get a digital copy of the book as a part of their collection." Information for publishers and authors about the Library Project, http://books.google.com/googlebooks/publisher_library.html (last visited Jan. 15, 2008).

themselves are not the database from which users search or from which search results are returned.²⁴ Rather, a relational database built up from the scanned images is the core of the project. An index is built of each word in the scanned text and its relationship to nearby words. This relational database is made available via a user interface on the Google website.²⁵ When a user searches the database using keywords, a snippet of the text comprising the keyword sought and a certain number of surrounding words is returned.²⁶ If a book is deemed to be in the public domain, the full text may be made available; if the book is still in copyright, the availability of text is restricted.²⁷ Rather than providing full text access for most books in the library program, Google provides, along with the search results, links to bookstores or libraries where the physical book may be purchased or borrowed.

Although access to the reconstructed text is parsimonious, the mere scanning of the books to produce the relational database has created a firestorm of controversy.²⁸ The holdings of any of these libraries will certainly include both works currently protected by copyright and works for which the copyright has lapsed or expired.²⁹ The latter type of work may of course be copied freely, but it will often be difficult to determine which category a particular work falls into. Google has argued that obtaining permission to scan the books would be prohibitive—even determining whether permission is needed would be prohibitive.³⁰ Prior to 1978, copyright in the United States was granted for a term

²⁴ As stated above, Google has disclosed little information about its search technology. See Schaffhauser, *supra* note 21. I rely here in part on conversations with personnel at Google.

²⁵ See How does Google Book Search work?, Google Book Search Help Center, <http://books.google.com/support/bin/answer.py?answer=43724&topic=9259> (last visited Jan. 15, 2008)

²⁶ See *id.*

²⁷ See Why can't I read the entire book?, Google Book Search Help Center, <http://books.google.com/support/bin/answer.py?answer=43729&topic=9259> (last visited Jan. 15, 2008).

²⁸ See Band, *supra* note 12, at 12.

²⁹ See *id.* at 1.

³⁰ See Proskine, *supra* note 8, at 219.

of years, with renewable extension.³¹ Some authors may have failed to renew the copyright for the extended term, in which case the work may have fallen into the public domain and may be copied freely—but it will often be difficult to determine whether a work was renewed.³² Additionally, since the early 1990s, U.S. copyright law has not required that a copyright notice be placed on published works.³³ Prior to that time, publication of a work without notice automatically placed it into the public domain, so that lack of a notice signaled that the work could be freely copied.³⁴ This is no longer a requirement for published works, so more recently published works that are protected by copyright may have no indication of who held the copyright when the work was published.³⁵

And, the complexity of determining a book's copyright status does not end there. Copyright may have been transferred to an entity other than the author or publisher; publishers may have gone out of business; authors may be deceased; the heirs of authors, who may or may not have inherited the copyright to the work, may be difficult or impossible to locate. Consequently, the cost of simply locating the copyright holders of many books, in order to obtain permission for their works to be scanned into the database, is potentially enormous.³⁶ Naturally, if copyright holders for the books can be located, some may decline permission to scan the book, diminishing the usefulness of the resultant database with each permission denied. But such refusals are a relatively simple and straightforward problem; it is equally likely that copyright owners, once located, could demand idiosyncratic fees or place restrictions on the use of the scanned work—the cost of such fees, as well as the cost of negotiating such permissions, would further add to the expense of creating the database.

³¹ See Copyright Act of 1909, § 23, available at <http://www.copyright.gov/history/1909act.pdf>.

³² See Library of Congress Copyright Office, How to investigate the copyright status of a work, Circular 22 (2004), <http://www.copyright.gov/circs/circ22.pdf>.

³³ 17 U.S.C. § 401, amended by The Berne Convention Implementation Act of 1988, Pub. L. No. 100-568 (Oct. 31, 1988).

³⁴ See Copyright Act of 1909, § 1.

³⁵ See Notice of Copyright, *supra* note 33.

³⁶ See Proskine, *supra* note 8, at 219.

This scenario appears to threaten the prospect of an “anti-commons” that might stifle the development of a comprehensive book database, whether compiled by Google or by anyone else.³⁷ Typically, property rights are allocated to prevent the so-called “tragedy of the commons,” where resources are misallocated because no one is motivated by ownership interests to maintain the resource.³⁸ But an “anti-commons” can potentially occur in situations where permissions from multiple property holders are necessary to complete a project and where the transactions cost of obtaining such permissions becomes prohibitive—that is, where there are too many property rights and rights holders, rather than too few.³⁹

Google has attempted to solve this problem by moving ahead with the scanning of books, but giving publishers and authors the option of requesting that their work not be scanned into the database.⁴⁰ This approach effectively shifts the burden of determining and asserting exclusive rights to copyright holders, requiring them to come forward and “opt out” of the project.⁴¹ Copyright holders have complained that this approach impermissibly inverts the basic exclusivity premise of intellectual property: that copying is prohibited unless authorized.⁴² However, the Google “opt out” procedure operates from the premise that Google has the right to copy the works, but as a courtesy will refrain from doing so if asked.⁴³ Much of the legal controversy to date over the Google “opt out” assertion has focused upon

³⁷ See Michael A. Heller, *The Tragedy of the Anticommons: Property in the Transition from Marx to Markets*, 111 HARV. L. REV. 621, 661 (1998).

³⁸ See *id.* at 677.

³⁹ See *id.*

⁴⁰ See What about books that I don't want in Google Book Search at all?, Google Book Search Help Center, <http://books.google.com/support/bin/answer.py?answer=43755&topic=9011> (last visited Jan. 15, 2008). Partners can upload a list of books to be excluded while others must participate in Google's opt-out process. *Id.*; see also Library Exclusion List Upload Instructions, Google Book Search Help Center, <http://books.google.com/support/partner/bin/answer.py?answer=44050> (last visited Jan. 15, 2008); Library Project Exclusion Registration, https://books.google.com/partner/exclusion-signup?hl=en_US (last visited Jan. 15, 2008).

⁴¹ See Band, *supra* note 12, at 12.

⁴² See *id.*

⁴³ See *id.* at 14.

determining how Google could claim to be in a position to make copies without obtaining advance copyright permission.⁴⁴

The legal justification for Google's opt-out position rests largely upon the American doctrine of fair use, a highly flexible, fact-specific exception to the rights of copyright holders.⁴⁵ Most countries have within their copyright law a series of privileges and exceptions that allow the users of copyrighted works to engage in unauthorized activities that would otherwise violate the rights of the copyright owner: quoting a protected work for news reporting, reproducing a work for educational purpose, or certain other activities socially beneficial activities.⁴⁶ But the United States, in addition to a list of such exceptions, also permits unauthorized uses on a case by case basis, depending upon the amount of protected expression taken, the type of work from which it is taken, the purpose to which it is put, and the effect of the unauthorized activity upon the market for the original work.⁴⁷

This U.S. exception to the exclusive rights of copyright holders has been explained by some commentators as a solution to the problem of high transaction costs.⁴⁸ This theory of "fair use as market failure" argues that fair use is necessary when the transaction costs of reaching agreement on authorized use is too high—when the copyright owner cannot be easily found, or demands a fee in excess of the value of the use, or the negotiations are protracted and cumbersome, and so on.⁴⁹ In such cases, the

⁴⁴ See, e.g., Edward Wyatt, *Ruling May Undercut Google in Fight Over Its Book Scans*, N.Y. TIMES, Feb. 25, 2006, <http://www.nytimes.com/2006/02/25/technology/25google.html> (speculating as to whether a relevant ruling against the creation of thumbnails from copyrighted images undermines the legality of Google's "opt-out" policy); see also Publishers Sue Google Over Plans To Digitize Books, Association of American Publishers (Oct. 19, 2005), http://publishers.org/main/PressCenter/Archives/2005_Oct/Oct_03.htm; Authors Guild Sues Google, Citing "Massive Copyright Infringement", The Authors Guild (Sept. 20, 2005), http://www.authorsguild.org/news/sues_google_citing.htm.

⁴⁵ See 17 U.S.C. § 107 (2007).

⁴⁶ See Band, *supra* note 12, at 14.

⁴⁷ See 17 U.S.C. § 107.

⁴⁸ Wendy J. Gordon, *Fair Use as Market Failure: A Structural and Economic Analysis of the Betamax Case and Its Predecessors*, 82 COLUM. L. REV. 1600, 1601 (1982); Lydia Pallas Loren, *Redefining the Market Failure Approach to Fair Use in an Era of Copyright Permission Systems*, 5 J. INTELL. PROP. L. 1 (1997).

⁴⁹ Gordon, *supra* note 48, at 1618; Loren, *supra* note 48, at 33 n.140.

law permits the user to circumvent the negotiation process and move ahead with the use, effectively taking a compulsory license at a zero royalty.⁵⁰ Often this will occur in the case of minor, de minimis uses, where the value of the use is relatively low relative to the costs of search, negotiation, and so on. But it could also occur when the value of the use is high and aggregate search costs are prohibitive, as in the anti-commons scenario.⁵¹

Several U.S. cases have held that producing a temporary or intermediate copy, which is produced in the process of developing a product different than the copyrighted work, and which is then discarded, or at least which is not part of the product eventually marketed to the public, is a fair use. For example, courts have repeatedly held that a copy of software made in the process of decompiling the software for reverse engineering, in order to produce an interoperable complimentary product, or even a competing product, is fair.⁵² The copy made is temporary—it is made in order to extract unprotected information about functionality, the result is a different product altogether, and the intermediate copy is not part of the product marketed to the public.⁵³ A similar analysis has been applied to extracting public domain artworks from a copyrighted publication.⁵⁴

This same rationale might be applied to the Google Book Search database. The scanned images are unnecessary to the final product—although images are being provided to the partner libraries, they can be discarded. They are not the end product that is to be offered to the public; rather, they are a mechanism or vehicle for creating the end product, which is the searchable database. The database is transformative, that is, it is not at all the

⁵⁰ See Dan L. Burk, *Muddy Rules for Cyberspace*, 21 CARDOZO L. REV. 121, 133 (1999).

⁵¹ See Michael A. Heller, *The Tragedy of the Anticommons: Property in the Transition from Marx to Markets*, 111 HARV. L. REV. 621, 624 (1998); see also Burk, *supra* note 50, at 158 & n.185.

⁵² Samuelson, *supra* note 1, at 87–88 (discussing *Sega Enters., Ltd. v. Accolade, Inc.*, 977 F.2d 1510 (9th Cir. 1992), *Atari Games Corp. v. Nintendo of Am.*, 975 F.2d 832 (Fed. Cir. 1992), and *MAI Sys. Corp. v. Peak Computers, Inc.*, 991 F.2d 511 (9th Cir. 1993)).

⁵³ Samuelson, *supra* note 1, at 66.

⁵⁴ *Id.* at 117–18.

same as, or a substitute for, the copyrighted works employed in its creation.⁵⁵ What is being offered to the public is more than the sum of the scanned images—Google has added value by disaggregating and indexing the texts to make the relations between individual words searchable. The database has value all its own, but its production would be impossible without the step of scanning the images.⁵⁶

At the same time, it might be argued that the market that is undermined by the use of the copyrighted works is the market for use of the works in databases. This is in fact a source of complaint about paying for permission to use individual works in the creation of the database.⁵⁷ Some previous copyright decisions suggest that in cases where a mechanism for licensing and payment of licensing royalties exists, the market potentially harmed by a fair use is the market for that particular use.⁵⁸ This argument may be unavailing in the context of the Google project; previous cases relying on this rationale involve photocopying periodicals when payment for such photocopies could be made through an existing mechanism. For example, cases holding that photocopying of materials for university course packets, or for archival reference, were based upon the existence of a mechanism for paying copy-licensing fees via the Copyright Clearance Center, which would distribute copying royalties to the owners of the works.⁵⁹ But we have already seen that in the case of the Google project, many of the books to be scanned are “orphan” works for which the copyright status is uncertain, and the possible copyright owner unknown. Consequently, it is unclear whether development of an effective licensing and payment mechanism is possible.

These considerations make Google’s fair use position at least tenable, and perhaps even decisive in the United States—but reliance on this U.S. doctrine has prompted Google to restrict Book

⁵⁵ See Loren, *supra* note 48, at 30.

⁵⁶ See Band, *supra* note 12, at 11.

⁵⁷ *Id.* at 13.

⁵⁸ See generally Loren, *supra* note 48, at 27–49 (analyzing recent case law).

⁵⁹ Princeton Univ. Press v. Mich. Document Servs., Inc., 99 F.3d 1381, 1381 (6th Cir. 1996).

Search access to Internet users in the United States.⁶⁰ Even with regard to U.S. law, such aspects of the Google project are fascinating and critically important. But they have already received some treatment elsewhere, and are likely to be the subject of further analysis and critique.⁶¹ For the discussion here, they are necessary primarily as a backdrop to a different set of issues that are foregrounded in the Google project, but which are endemic to digital copyright. Here I wish to focus upon the more fundamental and potentially far-reaching problem of ownership over metadata in digitized works.

This analysis centers on the database that Google is building from the scanned texts, rather than the act of scanning images. The database appears to consist of words and of metadata defining the relationship between those words.⁶² A given word will typically be too short to satisfy the requirements of originality and creativity for copyright;⁶³ only creative expression that originates with an author is eligible for copyright protection.⁶⁴ And even in the rare case where a word might qualify as original, it will likely fail copyright's "merger" doctrine: fundamental to the law of copyright is the principle that it protects only original expression, and never the idea that is expressed.⁶⁵ If there is only a single way, or very limited number of ways to express an idea, those expressive choices are excluded from copyright, since including them would be tantamount to protecting the idea itself.⁶⁶

Thus, as a matter of black-letter copyright law, it would appear that words are not protected by copyright, so no book publisher or author can claim copyright infringement of individual words.⁶⁷ But more troubling is the status of the meta-database that records the relationship *between* those words. The question I consider there, then, is whether Google has created an infringing copy of copyrighted books by building a database that allows

⁶⁰ Band, *supra* note 12, at 14.

⁶¹ *Id.* at 14.

⁶² Proskine, *supra* note 8, at 217.

⁶³ *Morrissey v. Procter & Gamble Co.*, 379 F.2d 675 (1st Cir. 1967).

⁶⁴ *Feist Publications, Inc., v. Rural Telephone Serv. Co.*, 499 U.S. 340 (1991).

⁶⁵ *See* 17 U.S.C. § 102(b) (2006).

⁶⁶ 17 U.S.C. § 102(b) (2006).

⁶⁷ *See id.*

disaggregated words to be re-assembled into the text of those books.

DEFINING DIGITAL COPIES

The first question in determining whether Google's database infringes the copyright in the scanned book texts is whether Google has made a *copy* for purposes of the copyright statute. Somewhat surprisingly, the first answer to this question of disaggregated copyrightable works was perhaps first apparent back to the beginning of the 20th Century, with the advent of automated player pianos, which played popular songs from paper rolls in which the music was coded as punched holes.⁶⁸ Music composers objected that such piano rolls—precursors to the punch cards on which computer data was later stored—when made without their permission infringed the copyright in their musical compositions.⁶⁹ But the United States Supreme Court held that such piano rolls did not fall under the copyright statute, but were rather a piece of a machine that produced music, akin to a cog or toothed wheel in a music box.⁷⁰ Machines and other functional devices are not covered by copyright; if covered by intellectual property law at all, they belong to the patent system.⁷¹ Consequently, the encoding of copyrighted music as punched holes was held not to constitute a violation of the copyright.⁷²

Congress responded to this decision, and to the ensuing distress of music composers, by creating a new category of copyrightable work, the phonorecord, intended to bring sound recordings within the Copyright Act.⁷³ More important than player piano rolls, the category of phonorecords also encompassed the phonograph record in which sounds were recorded as grooves in vinyl discs—a critically important format for the distribution of popular music.⁷⁴

⁶⁸ See *White-Smith Music Publ'g Co. v. Apollo Co.*, 209 U.S. 1, 9 (1907).

⁶⁹ *Id.* at 9–11.

⁷⁰ *Id.* at 18.

⁷¹ See 17 U.S.C. § 102(a) (2006).

⁷² *White-Smith*, 209 U.S. at 18.

⁷³ Copyright Act of 1909, *supra* note 31, at § 5.

⁷⁴ See Dan L. Burk, *Patenting Speech*, 79 TEX. L. REV. 99, 127 (2000).

In fact, this category of copyrightable fixation would eventually extend to encoding of music in a series of successive technological media: the magnetic flux of reel-to-reel, eight-track and cassette tapes, as well as the optical pits of compact discs.⁷⁵ Since each of these formats, as well as other new media such as videotape, could be considered parts of a machine in the same sense as a player piano roll, Congress also amended the statutory definition of copies to include both those media “now known or later developed”⁷⁶ from which the work could be perceived by a human being unaided, and from which perception of a work required the aid of a machine.⁷⁷

The media “later developed” to instantiate copyrightable works of course came to include digital media, perceived with the aid of a computer—the logical successors to the player piano roll.⁷⁸ Coin-operated video arcade games were among the earliest digitized works considered under this provision. The popularity of these games during the 1970s and 1980s led to unauthorized “knock-offs” of the most popular games, and to copyright suits against the copyists.⁷⁹ Game developers had registered videotapes of the game displays with the federal Copyright Office as proof of ownership. But copyists argued that registration of the game displays did not satisfy the requirements for copyright because the game output was not fixed in the circuits that generated the display—what was fixed was a computer program that produced the output, but the two were separable.⁸⁰ As evidence that the display was unfixed, the copyists pointed out that the game display changed each time the game was played, in response to player input. In a related argument, the copyists pointed out that no copy of the work was contained in the chips; rather, the work was generated or developed from instructions programmed into the chip.⁸¹ A variety of instructions or programs might produce the

⁷⁵ *Id.* at 127.

⁷⁶ Copyright Act of 1909, *supra* note 31, at § 26.

⁷⁷ *Id.*

⁷⁸ Burk, *supra* note 74, at 127.

⁷⁹ *See, e.g.*, Stern Elec. Co. v. Kaufman, 669 F.2d 852, 855 (2d Cir. 1982); Williams Elec., Inc. v. Artic Int'l, Inc., 685 F.2d 870, 873 (3d Cir. 1982).

⁸⁰ *Stern*, 669 F.2d at 855.

⁸¹ *Williams*, 685 F.2d at 877.

same output, and registration of the display should not entitle the copyright holder to every set of instructions producing such a display.⁸²

In cases such as *Stern Electronics Co. v. Kaufman*⁸³ and *Williams Electronics v. Artic International, Inc.*,⁸⁴ the courts rejected such arguments on the basis of the statutory definition, holding that if the work was fixed in the game's semiconductor chip and could be perceived with the aid of a machine then it must be a statutory copy. But under this holding, "perceive" must implicitly include *generation* or re-assembly of a work. The pattern of voltages in a chip or the pattern of magnetic flux on a disc are both profoundly and subtly different from the grooves of a vinyl phonograph record. Such analog recordings use one physical quantity to represent another. Consequently, analog media maintain some relationship within the record of a work corresponding to the relationships within the work itself. Digital records need not maintain such analogous relationships, but are instead series of sequences of bits that can be read to re-construct the work.⁸⁵ While the digital version of the work is in some sense a record of the work, it is not a *recording* of the work as found in previous media.

Thus, unlike a microform reader which simply amplifies human perception, computer code constitutes a set of instructions for generating the work perceived—but the courts held that perception of output implied existence of a digital copy.⁸⁶ As a corollary, these cases also skirted the contested definitional lines between "data" and "software"—a sequence of bits containing the *instructions* to generate music or text, and a sequence of bits constituting the *record* of digitized music or text appear to be treated identically under these opinions.⁸⁷ The *Stern* and *Williams* courts were likely less concerned with the technicalities of digital processing than they were with the end product of the process—an

⁸² *Stern*, 669 F.2d at 855.

⁸³ *Id.* at 856.

⁸⁴ *Williams*, 685 F.2d at 877.

⁸⁵ *Stern*, 669 F.2d at 855–56; *Williams*, 685 F.2d at 874.

⁸⁶ *Stern*, 669 F.2d at 857; *Williams*, 685 F.2d at 873–74.

⁸⁷ *Stern*, 669 F.2d at 855; *Williams*, 685 F.2d at 871.

audiovisual work that, to all appearances, fell within copyright's statutory subject matter—and with developing a plausible social policy for such end products.⁸⁸ Consequently, these early gaming cases proved pivotal to establishing the copyrightability of computer programs, but their implications for storage of digital copies may not be consistent with more recent decisions, such as those discussed below.

NEW YORK TIMES CO. V. TASINI

The relationship between the arrangement of text in hardcopy materials and arrangement of digitized texts within a database was central to the decision of the United States Supreme Court in *New York Times Co. v. Tasini*.⁸⁹ The *Tasini* case involved infringement claims by independently contracted or freelance writers who had licensed stories to a wide variety of periodical publications: newspapers and magazines including the New York Times, Time magazine, and Newsday.⁹⁰ Many of these hardcopy periodicals had, since licensing or purchasing the freelance stories, begun making their contents available in searchable full-text electronic format.⁹¹ Some periodicals developed online databases, while others provided their contents on CD-ROM or similar electronic formats.⁹² Because the licenses or copyright transfers from the authors were executed before electronic versions of periodicals became common, the transfers did not address publication of the stories in databases or other electronic formats.⁹³ Consequently, the authors claimed that inclusion of their stories in electronic formats was an unauthorized re-publication of their work, for which no rights had been granted to the periodical publishers.⁹⁴

In defense, the publishers relied upon § 201 of the copyright statute,⁹⁵ which both establishes copyright in collective works and

⁸⁸ *Stern*, 669 F.2d at 855; *Williams*, 685 F.2d at 877.

⁸⁹ 533 U.S. 483, 506 (2001).

⁹⁰ *Id.* at 490.

⁹¹ *See id.*

⁹² *See id.*

⁹³ *See id.*

⁹⁴ *See id.*

⁹⁵ *Id.*

distinguishes such copyright from that in the individual contributions to a collective work.⁹⁶ Under this section of the statute, copyright may subsist in the original selection and arrangement of copyrighted works in a collective work, such as an anthology.⁹⁷ This copyright in the collective work is separate and distinct from the copyright in any given work included in the collective work—copyright in the constituent works need not be transferred to the author or compiler of a collective work.⁹⁸ The holder of the collective work copyright is limited in the subsequent uses to which she may put the individual works comprising the collective work.⁹⁹ Absent transfer of the copyright in the constituent works, the holder of the collective work copyright may use the constituent works only in a revision or re-issue of the initial collective work.¹⁰⁰

The publishers argued that the databases constituted a “revision” of the hardcopy periodical issues in which articles originally appeared, and that therefore the publishers had the right to include the articles in the database collections.¹⁰¹ But, this defense was rejected by the Supreme Court, which held that the periodical databases or electronic versions of the periodicals were not equivalent to the print versions.¹⁰² The Court reasoned that the articles in the databases had been disaggregated from their sequence and relationship in the print edition.¹⁰³ Because the articles were no longer in the context of their original publication, but rather divorced of their print media relationships, the § 201 revisioning privilege did not apply.¹⁰⁴

The Court rejected the contention that the electronic and CD databases contained the original periodicals despite the fact that articles retrieved in searches would display the pagination and publication markings of the periodical in which they were

⁹⁶ 17 U.S.C. § 201 (2000).

⁹⁷ *Id.*

⁹⁸ *Id.*

⁹⁹ *Id.*

¹⁰⁰ *Id.*

¹⁰¹ *See* N.Y. Times Co. v. Tasini, 533 U.S. 483, 485 (2001).

¹⁰² *See id.* at 484.

¹⁰³ *See id.* at 486.

¹⁰⁴ *See id.*

originally published.¹⁰⁵ The Court held that such indicia were indicative of the article having previously appeared in the periodical, but were not indicative of the article as retrieved being part of the periodical.¹⁰⁶ Rather, the articles were better thought of as components of a super-compendium or library consisting of all the disaggregated articles, and such a super-compendium was not equivalent to the original periodical.¹⁰⁷ The Court specifically rejected an analogy to microform records, as microforms, unlike electronic databases, recorded the original sequence of the periodicals.¹⁰⁸ The court also rejected the argument that because the users of the databases could re-assemble the articles into the original sequence, that the database was a revision of the original periodicals, anymore than a hardcopy library from which a patron could retrieve and re-assemble a periodical sequence would be a “revision” of those periodicals.¹⁰⁹

MATTHEW BENDER & CO. V. WEST PUBLISHING CO.

A similar set of issues is found in copyright cases considering the protectability of relationships within the compiled text of volumes of judicial opinions. The development of legal research databases from these opinions occasioned the disputes in such cases, when West Publishing Company attempted to prevent rival database publishers from adopting a standardized case citation format.¹¹⁰ West Publishing was and remains the major publisher of bound, hardcopy volumes of judicial reports, collecting the judicial opinions from essentially every jurisdiction in the United States.¹¹¹ While some jurisdictions published their own reporters, in many instances the West reporter volumes were the only judicial reporters for certain jurisdictions.

¹⁰⁵ See *id.* at 485.

¹⁰⁶ *Id.*

¹⁰⁷ See *id.*

¹⁰⁸ See *id.* at 485–86.

¹⁰⁹ See *id.* at 486.

¹¹⁰ See generally *Matthew Bender & Co. v. West Publ'g Co. (West)*, 158 F.3d 693 (2d Cir. 1998).

¹¹¹ Company Overview, Thompson West, <http://west.thomson.com/overview> (last visited Nov. 16, 2007) [hereinafter Thompson West].

Consequently, citation to the West reporter volumes became effectively an industry standard: anyone practicing in the legal profession used citations to the West bound hardcopy reports, not only in office documents and memoranda, but in official documents filed with the court system.¹¹² Law students across the United States were trained to use West citations as an essential component of their professional preparation.¹¹³ Judicial opinions routinely used West citations.¹¹⁴ Indeed, many courts required attorneys practicing before them to use West citations in motions and briefs, and allowed the use of no other system, because the West reporter volumes were most commonly available to judges searching for, and citing to, judicial precedent.¹¹⁵

With the advent of electronic storage and retrieval systems containing searchable, full text versions of judicial opinions, West transferred its reporter volume text to electronic format.¹¹⁶ West's subscription database, Westlaw, used the "star pagination" system, which inserted into the electronic text of opinions numerical markers corresponding to the location of pages in the bound hardcopy reporter volumes.¹¹⁷ A rival legal publisher provided a competing product made up of judicial opinions on CD-ROM discs that could be accessed by a purchaser's own machine, rather than via on-line database access.¹¹⁸ The text of these opinions on CD-ROM included "star pagination" markers relating judicial opinion text to the published West reporter volumes—without such citations to the West volumes the electronic databases would be essentially useless to lawyers and other legal professionals.¹¹⁹

However, West sued to prevent its competitors from using such pagination markers, arguing that appropriation of the star pagination citations was essentially appropriation of their published volumes—arguing, in other words, that such markers, indicating the position of text in the published volumes, mapped

¹¹² *Id.*

¹¹³ *Id.*

¹¹⁴ *See West*, 158 F.3d at 696.

¹¹⁵ *See id.*

¹¹⁶ *See Thompson West*, *supra* note 111.

¹¹⁷ *See West*, 158 F.3d at 695.

¹¹⁸ *See id.*

¹¹⁹ *See id.*

onto the published volume in such a way as to effectively constitute a copy of the hardcopy book.¹²⁰ According to West, the CD-ROM discs comprised “copies” of the West reporter volumes because the selection and arrangement of cases in the West reporters could be perceived with the aid of a machine by use of the star pagination markers.¹²¹ In other words, by following the star pagination markers, a user could employ the automated functions of the CD-ROM to view or print cases in the same order found in the bound West reporter volumes, and this meant that the CD-ROMs contained copies of the West volumes.¹²² West also argued that the layout, selection, and arrangement of opinions in the published volumes constituted original, copyrighted expression.¹²³ Hence, under this theory, use of “star pagination” markers constituted copying of the protected West volumes.¹²⁴

This citation infringement argument was striking because West advanced no argument that copying the content of the volumes was an infringement—it hardly could, because in the United States, by statute, federal government documents, including federal judicial opinions, cannot be copyrighted and lie in the public domain.¹²⁵ Many state government documents receive similar treatment, meaning that most of the text at issue in West’s claims was unquestionably unprotected by copyright. Neither did they argue that the layout of the judicial opinion text was directly appropriated—formatting and layout was not captured in early electronic text databases, which was indeed what made the star pagination system of markers necessary.¹²⁶ Rather, West’s argument was that citation that indicated the position of information—essentially what would come to be called metadata—was protected by their copyright.¹²⁷

¹²⁰ See *id.* at 695–96.

¹²¹ See *id.*

¹²² See *id.*

¹²³ See *id.* at 700.

¹²⁴ See *id.*

¹²⁵ See *id.* at 698.

¹²⁶ West claimed that Bender’s product allowed users to “to view (and print) judicial opinions in the same order in which they are printed in a West volume,” not that it copied West’s publications. *Id.* at 697.

¹²⁷ “West’s alternative argument is that even though the page numbering is not (by itself) a protectable element of West’s compilation, (i) plaintiffs’ star pagination to

As an initial matter, the court declined to find any significant expression in the editing, selection, and arrangement of opinions in the West reporters—the court found that West chronologically published all the opinions made available, with only routine and minimal editing.¹²⁸ The court rejected the copying argument, holding that offering the capability to re-generate the West volumes was not equivalent to offering copies of the West volumes.¹²⁹ The sequence of page breaks signified by star pagination was not itself original, as it was created by a mechanical typesetting process, and not by any creative selection and arrangement of West's.¹³⁰ Moreover, even though star pagination markers might reveal to a reader how the West arrangement could be recreated, that arrangement was not fixed in the CD-ROM disc.¹³¹ Only manipulation of the data by a user would produce the West volume case sequence, and the products of user manipulation were copies fixed in the discs.¹³² Distinguishing the video game cases such as *Stern*,¹³³ the court reasoned that adopting West's argument regarding fixation would effectively extend West's copyright to all arrangements or re-arrangements that could be generated by a user, and West was not entitled to control user-generated arrangements.¹³⁴

MICRO STAR V. FORMGEN, INC.

The copyright status of metadata descriptions has also been considered by the Ninth Circuit in *Micro Star v. FormGen Inc.*¹³⁵

West's case reporters embeds West's arrangement of cases in plaintiffs' CD-ROM discs, thereby allowing a user to perceive West's protected arrangement through the plaintiffs' file-retrieval programs, and (ii) that under the Copyright Act's definition of 'copies,' 17 U.S.C. § 101, a work that allows the perception of a protectable element of a compilation through the aid of a machine amounts to a copy of the compilation." *Id.* at 700.

¹²⁸ *See id.* at 705.

¹²⁹ *See id.* at 704.

¹³⁰ *See id.* at 705.

¹³¹ *See id.* at 703.

¹³² *See id.*

¹³³ *See id.* at 704; *see generally* *Stern Elec. Co. v. Kaufman*, 669 F.2d 852 (2d Cir. 1982) (discussion *supra*).

¹³⁴ *See West*, 158 F.3d at 703.

¹³⁵ 154 F.3d 1107 (9th Cir. 1998).

The dispute in that case centered on advanced playing levels for the computer game “Duke Nukem 3-D.”¹³⁶ The game developer, FormGen, made available to its users the tools to develop alternate game levels and encouraged the sharing of such user-created game files on its web site.¹³⁷ The advanced or alternate game versions existed as “MAP” files, or sets of game instructions, that could draw upon a graphic library of character and object images provided with the game itself, but which would sequence, arrange, and display the library images in such a way as to provide a more challenging game experience.¹³⁸ In the opinion, the court compared such MAP files to “paint by number” instructions that tell where to place graphics from the library.¹³⁹ The library of graphics itself was not traded or distributed with the MAP files; rather, it was part of the game purchased by users from FormGen.¹⁴⁰ Thus, the MAP files operated together with other components with the Duke Nukem game, but could not themselves independently generate game output, lacking the necessary content to do so.¹⁴¹

The defendant in the case, Micro Star, assembled on compact disc collections of the user-created files and marketed the disc without the authorization of either FormGen or of the users who developed the files.¹⁴² The user-created files, having been tacitly, if not explicitly, authorized by FormGen were presumably original works of authorship, for which the users would hold copyright and for which they might have a claim against Micro Star for infringing reproduction and distribution of the files.¹⁴³ But suit was brought by FormGen, which had not created the files.¹⁴⁴ This raised the question as to whether Micro Star had taken any of FormGen’s original expression.¹⁴⁵ If none of FormGen’s

¹³⁶ *See id.* at 1109.

¹³⁷ *Id.*

¹³⁸ *See id.* at 1110.

¹³⁹ *Id.*

¹⁴⁰ *See id.*

¹⁴¹ *See id.*

¹⁴² *See id.* at 1109.

¹⁴³ *See id.*

¹⁴⁴ *Id.*

¹⁴⁵ *See id.*

expression were found in the files, FormGen had no infringement claim.¹⁴⁶

FormGen claimed that the taken MAP files constituted derivative works that contained protected expression, giving them standing to sue.¹⁴⁷ In response, Micro Star argued that the copied and distributed MAP files failed the requirements of a derivative work, that they incorporated no protected expression, and did not constitute any concrete or permanent alteration to the FormGen's original expression in the game.¹⁴⁸ Micro Star's arguments hinged upon the MAP files comprising only instructions for assembling an audiovisual display, without incorporating any of Formgen's content. The characters, objects, and other graphics of the game were contained in the separate game library that was not copied or distributed by Micro Star, and which was unaltered by the MAP files.¹⁴⁹

The appellate opinion authored by Judge Kozynski rejected this line of argument by reasoning somewhat circularly first, that the MAP files were permanent and concrete instantiation of a derivative work because they fully described an infringing audiovisual output.¹⁵⁰ Analogizing the MAP files to sheet music, which Kozynski claimed similarly "describes" a musical composition, the court reasoned that a description of a derivative work is equivalent to a derivative work.¹⁵¹ The court further held that the MAP files incorporated protected expression as a sort of narrative; because the MAP files dictated the placement and sequence of the Duke Nukem characters, the MAP files therefore constituted a "story" about the FormGen game characters.¹⁵² Relying on the early video game cases, Kozynski's opinion reasons that the alternate game levels were derivative works of FormGen's Duke Nukem story, and that by "describing" the placement and arrangement of the graphics in those derivative works, the MAP

¹⁴⁶ *See id.*

¹⁴⁷ *Id.* at 1110.

¹⁴⁸ *See id.* at 1112.

¹⁴⁹ *See id.*

¹⁵⁰ *Id.* at 1111–12.

¹⁵¹ *Id.* at 1112.

¹⁵² *See id.*

files constituted in effect the plot, the sequence, and the narrative of those derivative works.

The result in *Micro Star* might be read to imply that a relational database constitutes a copy of the work that it maps, since the court characterized the question presented as “whether an exact, down to the last detail, description of an audiovisual display” constituted a permanent fixation of a derivative work.¹⁵³ Although this question is not exactly the question presented by the Google Book Search database, the questions are clearly closely related. In Judge Kozynski’s words, the Google database comprises “an exact, down to the last detail description” of texts scanned into the database.¹⁵⁴ If instructions for re-arranging copyrighted elements into a derivative work constitute a permanent and concrete instantiation of the initial work, then instructions for re-assembling elements into the exact arrangement of a copyrighted work might be seen to constitute a copy of the initial work.¹⁵⁵

But it is unclear how far the logic of the opinion can be extended, as even on its own terms the logic of the opinion is more than a little dubious. To begin with, the premise that the opinion relies upon is by definition flawed—the MAP files were not, as the court claimed, an exact description of an audiovisual display down to the last detail description; had they been, there would have been no need for the game graphics library.¹⁵⁶ The MAP files were instructions on the placement and arrangement of data objects drawn from the graphics library, but those instructions would be agnostic as to exactly which images were so arranged. One could imagine substituting a different graphics library, perhaps of Disney cartoon characters, or even caricatures of federal judges, for the FormGen graphics library.¹⁵⁷ So long as the files in such graphic libraries were labeled with the same designators as those in the FormGen library, the images could be “called” in place of the Duke Nukem files, and a narrative devoid of any FormGen would

¹⁵³ *Id.* at 1111–12.

¹⁵⁴ *Id.* at 1111.

¹⁵⁵ *See id.* at 1112.

¹⁵⁶ *See id.* at 1111–12.

¹⁵⁷ *See generally id.* at 1112.

be generated.¹⁵⁸ The point is, of course, that the MAP files contained no distinctive “story” regarding Duke Nukem unless paired with the content of the FormGen graphics library.

Kozynski’s comparison of the MAP files to sheet music, as instructions to produce a copyrighted work, fails for similar reasons. Sheet music seldom offers a staff and time signature leaving the performer to fill in the actual notes to be played, nor for that matter does it merely offer instructions on how to assemble the score from a musical library such as “insert third note of the Mozart Requiem,” “insert twelfth note of the J.S. Bach Minuet in G,” and so on.¹⁵⁹ Similarly, if the MAP files constituted detailed instructions about a derivative work, then it seems dubious that they could constitute a derivative work. On Kozynski’s logic, a book of instructions on how to compose a book about Duke Nukem would itself qualify as a derivative work.¹⁶⁰ The problem with this conclusion seems clear from the court’s comparison of MAP files to “paint by number” instructions—the result of painting by numbers might be infringing a copyrighted work, but it seems inconceivable that instructions such as “paint the spaces with the number 2 red” could be said to have copied any original expression from such work.¹⁶¹ Such instructions might perhaps constitute some sort of inducement to infringement, but that is another claim entirely.

COPYRIGHT AND METADATA

Although the Google project is novel in scope and vision, digitized works have been the subject of copyright controversy for well over a quarter century, so that previous cases offer some suggestion as to how Google’s meta-database should be regarded.¹⁶² The precedent that seems most relevant to the status of the Google database may appear to point in different

¹⁵⁸ *Id.*

¹⁵⁹ *See id.*

¹⁶⁰ *See id.*

¹⁶¹ *See id.* at 1110.

¹⁶² *See, e.g., id.* at 1107.

directions.¹⁶³ Taken on their own, the early videogame cases seem to suggest that the form of digital fixation is itself irrelevant: so long as the copyrighted work can be re-generated as output from the circuits of a machine, it doesn't matter what kind of coding or instructions one would find at the machine level.¹⁶⁴ This in turn seems to indicate that the set of instructions, or metadata, necessary to re-construct a copyrighted work itself constitutes an infringing copy of the work—that the metadata describing a work reproduces the originality in the work. The *Micro Star* holding appears somewhat consistent with the results in the early videogame cases, although it also appears to be inconsistent with the results in *Tasini* and *West*. An examination of the MAP files in *Micro Star* would not only fail to reveal the physical relational structure sought by the courts in *Tasini* and *West*, but would fail to find any representation whatsoever of the graphics displayed by the game output, as such graphics resided in the game library.¹⁶⁵

But careful consideration of the inapposite analogies in the *Micro Star* opinion also makes clear that this result is not altogether sensible and that instructions leading to a particular selection and arrangement are necessarily equivalent to selection and arrangement itself.¹⁶⁶ Certainly this would not be the result if considering written instructions to a human, rather than coded instructions to a machine—if, for example, an art expert meticulously examined a famous painting and then wrote out detailed instructions re-creating the painting, brush stroke by brush stroke, it seems fairly clear that such instructions would not be considered to constitute a copy of the initial painting.¹⁶⁷ The same would surely be true if the instructions were implemented by a machine; if a copyist developed a mechanical painting arm and programmed it to reproduce in fine detail the brush strokes of a painting, even if the product of the mechanical arm constituted a copy of the initial painting, the program instructing the movements of the mechanical arm would not be a copy. Such instructions

¹⁶³ See *id.*

¹⁶⁴ See *id.*

¹⁶⁵ See *id.*; N.Y. Times Co. v. Tasini, 533 U.S. 483 (2001).

¹⁶⁶ See *Micro Star*, 154 F.3d at 1111–12.

¹⁶⁷ See *id.* at 1112.

might themselves warrant copyright protection, although that protection would likely be minimal, since the underlying process is excluded from copyright protection, and only whatever creativity might be found in the expression of the instructions could be covered by copyright. But the copyright in the instructions would clearly not cover the *result* of following the instructions.¹⁶⁸

Other recent cases dealing with digitized hardcopy works tend to hold that when individual uncopyrightable elements of a copyrighted work are stored electronically so as to disrupt the relationship found between them in the original work, no copy has been made.¹⁶⁹ Although decided in the context of a very specific statutory provision—§ 201—the reasoning in *Tasini*¹⁷⁰ suggests that digital versions of hardcopy materials do not infringe copyright in the hardcopy text due to the disaggregation that occurs in digital storage and retrieval.¹⁷¹ Neither did the Supreme Court view user-initiated reassembly of the hardcopy sequence as infringement, due to the disaggregated nature of the database from which the user was working. The presence of metadata sufficient to reassemble the original texts did not change this view. This tends to suggest that neither the Google database of disaggregated book text nor the meta-database of book text relationships should be viewed as infringing copies of the books.¹⁷²

Much as in the *Tasini* decision, the analysis in *West* suggests that neither disaggregated digitized text nor relational metadata regarding that text constitute a copy of the original text for purposes of the copyright statute.¹⁷³ But the reasoning in the *West* decision bears even more directly upon the Google situation, as this analysis is not through the lens of § 201 republication. Unlike the databases in *Tasini* and *West*, the Google Book Search database does not maintain the works scanned into it as discrete retrievable

¹⁶⁸ See *id.* at 1111–12.

¹⁶⁹ See generally *Tasini*, 533 U.S. 483.

¹⁷⁰ *Id.*

¹⁷¹ See *id.* at 503–04.

¹⁷² See *id.* at 504.

¹⁷³ See *Matthew Bender & Co., Inc. v. West Publ'g Co. (West)*, 158 F.3d 693, 703–04 (1998).

works, but atomizes them to the level of individual words.¹⁷⁴ And far from assembling individually copyrightable images into what Judge Kozynski might consider “narrative,” the largest chunk available to the Google end user is a snippet of a few dozen words.¹⁷⁵ Consequently, the Google database resembles its original texts even less than did the databases in *Tasini* and *West*, and seems even less likely to constitute and infringing copy.

The analysis in these cases demarcates a general set of concerns for not only the Google project, but for digital copyright generally. Although the law of copyright was developed in an analog world, creative works of all kinds are now captured as series of sequences of bits rather than as analog records.¹⁷⁶ This change in the fixation of works has several consequences that are problematic for the basic doctrines of copyright.¹⁷⁷ Most of these consequences flow from the fungible nature of bits. Previous analog media typically encoded different types of works in different formats—motion pictures were not recorded as grooves in vinyl; musical compositions were not recorded as grains of silver nitrate on celluloid. But digitized music or software or text are all just sequences of bits, not anything that can as encoded be differentiated as pictures or music or text.¹⁷⁸ Data processors make no distinction between bits that represent a photograph or painting or a piece of music or a piece of text, or for that matter between bits representing a copyrightable work and bits representing something uncopyrightable, such as a Fourier transform series.¹⁷⁹

Because digital records use this common building block of the bit, it is possible to arrange that common building block into all kinds of copyrightable works. In this sense, digital media turns everything—all kinds of copyrightable works—into databases, into

¹⁷⁴ See How does Google Book Search work?, Google Book Search Help Center, <http://books.google.com/support/bin/answer.py?answer=43724&topic=9259> (last visited Jan. 15, 2008).

¹⁷⁵ See *id.*

¹⁷⁶ See Dan L. Burk, *Expression, Selection, Abstraction: Copyright's Golden Braid*, 55 SYRACUSE L. REV. 593, 594 (2005).

¹⁷⁷ See *id.*

¹⁷⁸ See *id.* at 611.

¹⁷⁹ See *id.*

compilations of fungible elements.¹⁸⁰ Everything is reducible to discrete elements, none of which are individually original.¹⁸¹ This in turn means that there must necessarily be some type of metadata, some type of organizational instruction, as to the manner in which the bits are to be reassembled.¹⁸² And it is at this level that the original expression necessary for copyright protection must reside, in the manner in which the bits have been arranged to encode music or text or graphics.¹⁸³ There is nothing original in any of the individual bits that can be read to constitute a photograph or text or musical composition. The originality lies not in the components, the bits of information, but in the way that the bits of information are arranged

Such atomistic reduction of copyrighted works to fungible units may not necessarily be the product of digital technologies; previous media show the same characteristics to some degree.¹⁸⁴ For example, it is possible to view print media as an arrangement of individual letters from the alphabet into words, and at the next level of organization, as an arrangement of individual words into a novel, play, or poem. This reductionist view presents the same doctrinal problem: surely the letter “A” is not of itself protectable in copyright, nor is the letter “B,” nor are individual words.¹⁸⁵ The only original aspect of the work must be the author’s selection and arrangement of the words and letters. The same is true in the case of other copyrightable works, such as a musical composition or an Impressionist painting. No individual dab of paint or musical note will entail the requisite originality for copyright. Rather, the arrangement of the dabs of paint, the arrangement of individual notes, structured to communicate a particular idea, is original.

Thus, the essence of copyright seems to lie in original selection and arrangement of fungible elements.¹⁸⁶ Digitization makes this result more apparent because it facilitates the disaggregation of

¹⁸⁰ *See id.*

¹⁸¹ *See id.*

¹⁸² *See id.*

¹⁸³ *See id.*

¹⁸⁴ *See id.* at 612.

¹⁸⁵ *See id.*

¹⁸⁶ *See id.*

individual elements.¹⁸⁷ Although this principle could be applied to previous analog or physical types of media, it was largely a matter of academic or philosophical speculation—an exercise in determining how many works of original authorship could dance on the head of a pin. But the issue is now unavoidable, and therein lies the paradox for copyright doctrine: copyrightable selection and arrangement cannot exist in a vacuum; there must be selection and arrangement of *something*.¹⁸⁸ This seems to be the message of the *Tasini* and *West* decisions. Relational metadata, as generated in *Tasini* or *West* or in the Google Book Search database, is a description of the selection and arrangement of atomistic elements in the work described, but cannot itself be a copy of the work, because the metadata does not incorporate the atomized elements of the work.¹⁸⁹

This is presumably good news for the Google project, as it argues powerfully against its disaggregated textual database constituting an infringing copy, quite apart from an analysis of fair use. But this reductionist conclusion also in some sense places copyright doctrine on a collision course with itself. If no individual bit of data warrants copyright protection, and the metadata used to arrange such bits also fails the criteria for copyright, then it is unclear which digital content might warrant copyright protection. Indeed, this outcome implies that the Google metadatabase itself, for all the investment that has gone into it, is no more eligible for copyright protection than any other digital work in the database of databases that comprises the emerging global information structure.

¹⁸⁷ *See id.*

¹⁸⁸ *See id.* at 612–13.

¹⁸⁹ *See id.* at 613.