12-4547-cv

United States Court of Appeals

for the

Second Circuit

AUTHORS GUILD, INC., AUSTRALIAN SOCIETY OF AUTHORS LIMITED, UNION DES ECRIVAINES ET DES ECRIVAINS QUEBECOIS, ANGELO LOUKAKIS, ROXANA ROBINSON, ANDRE ROY, JAMES SHAPIRO, DANIELE SIMPSON, T.J. STILES, FAY WELDON,

(For Continuation of Caption See Inside Cover)

ON APPEAL FROM THE UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF NEW YORK

BRIEF OF DIGITAL HUMANITIES AND LAW SCHOLARS AS *AMICI CURIAE* IN SUPPORT OF DEFENDANTS-APPELLEES AND AFFIRMANCE

JASON SCHULTZ*
ASSISTANT CLINICAL PROFESSOR OF LAW
UC BERKELEY SCHOOL OF LAW
396 Simon Hall
Berkeley, California 94720
(510) 642-1957
jschultz@law.berkeley.edu

Attorneys for Amici Curiae

On the Brief:
MATTHEW SAG*
ASSOCIATE PROFESSOR
LOYOLA UNIVERSITY OF
CHICAGO SCHOOL OF LAW

* Filed in their individual capacity and not on behalf of their institutions.

AUTHORS LEAGUE FUND, INC., AUTHORS' LICENSING AND COLLECTING SOCIETY, SVERIGES FORFATTARFORBUND, NORSK FAGLITTERAER FORFATTERO OG OVERSETTERFORENING, WRITERS' UNION OF CANADA, PAT CUMMINGS, ERIK GRUNDSTROM, HELGE RONNING, JACK R. SALAMANCA,

Plaintiffs-Appellants,

v.

HATHITRUST, CORNELL UNIVERSITY, MARY SUE COLEMAN, President, University of Michigan, MARK G. YUDOF, President, University of California, KEVIN REILLY, President, University of Wisconsin System, MICHAEL MCROBBIE, President, Indiana University,

Defendants-Appellees,

NATIONAL FEDERATION OF THE BLIND, GEORGINA KLEEGE, BLAIR SEIDLITZ, COURTNEY WHEELER,

Intervenor Defendants-Appellees.

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Cases

Statutes

Constitutional Provisions

Other Authorities

STATEMENT OF INTEREST OF *AMICI*[1]

Amici are over 100 professors and scholars who teach, write, and research in computer science, the digital humanities, linguistics or law, and two associations that represent Digital Humanities scholars generally.[2] *Amici* have an interest in this case because of its potential impact on their ability to discover and understand, through automated means, the data in and relationships among textual works. Legal Scholar *Amici* also have an interest in the sound development of intellectual property law. Resolution of the legal issue of copying for non-expressive uses has far-reaching implications for the scope of copyright protection, a subject germane to *Amici*'s professional interests and one about which they have great expertise. *Amici* speak only to the issue of copying for non-expressive uses. A complete list of individual *amici* is attached as Appendix A.

---

[1] Pursuant to Fed. R. App. P. 29(a), (c)(4), (c)(5) and Rule 29.1 of the Local Rules of the United States Court of Appeals for the Second Circuit, *Amici* hereby state that none of the parties to this case nor their counsel authored this brief in whole or in part; no party or any party's counsel contributed money intended to fund preparing or submitting the brief; and no one else other than *Amici* and their counsel contributed money that was intended to fund preparing or submitting this brief. *Amici* also hereby state that all parties have consented to the filing of this brief, and we rely on that consent as our source of authority to file.

[2] *See* Association for Computers and the Humanities, http://www.ach.org/; Canadian Society for Digital Humanities, http://csdh-schn.org.

.

SUMMARY OF ARGUMENT

Mass digitization, especially by libraries, is a key enabler of socially valuable computational and statistical research (often called "data mining" or "text mining"). While the practice of data mining has been used for several decades in traditional scientific disciplines such as astrophysics and in social sciences like economics, it has only recently become technologically and economically feasible within the humanities. This has led to a revolution, dubbed "Digital Humanities," ranging across subjects like literature and linguistics to history and philosophy. New scholarly endeavors enabled by Digital Humanities advancements are still in their infancy but have enormous potential to contribute to our collective understanding of the cultural, political, and economic relationships among various collections (or *corpora*) of works—including copyrighted works—and with society. The Court's ruling in this case on the legality of mass digitization could dramatically affect the future of work in the Digital Humanities.

This Court should affirm the decision of the district court below that library digitization for the purpose of text mining and similar non-expressive uses present

no legally cognizable conflict with the statutory rights or interests of the copyright holders. Where, as here, the output of a database—*i.e.*, the data it produces and displays—is noninfringing, this Court should find that the creation and operation of the database itself is likewise noninfringing. The copying required to convert paper library books into a searchable digital database is properly considered a "non-expressive use" because the works are copied for reasons unrelated to their protectable expressive qualities — the copies are intermediate and, as far as is relevant here, unread.

The type of non-expressive use at issue here – statistical analysis of text – is common among copy-reliant technologies: for example, Internet search engines and plagiarism detection software do not read, understand, or enjoy copyrighted works, nor do they deliver these works directly to the public. Such platforms copy the works only incidentally, in order to process them as "grist for the mill"—raw materials that feed various algorithms and indices. *See* Matthew Sag, *Copyright and Copy-Reliant Technology*, 103 NW. U.L. REV. 1607 (2009); Matthew Sag, *Orphan Works as Grist for the Data Mill*, 27 BERKELEY TECH. L. J. 1503 (2012).

Further, generating data about a copyrighted work (often called "metadata") does not infringe the original work because, as has been recognized for over a century, copyright law protects only an author's original expression, not facts. That a "fact" might pertain to or describe an expressive work does not change its factual

character—or render it an author's exclusive intellectual property under the law. Indeed, making such factual information freely available to all is crucial to the harmony between copyright law and the First Amendment—hence the existence of rules like the "idea/expression" distinction (*see* 17 U.S.C. § 102(b)), the doctrine of *scenes à faire*, and the "merger" principle.

The act of copying works into a database in order to enable the generation of metadata about those works should thus be deemed noninfringing. As numerous courts have found, making intermediate copies that enable socially beneficial noninfringing uses and/or outputs constitutes a protected "fair use" under Section 107 of the Copyright Act. *See, e.g., A.V. ex rel. Vanderhye v. iParadigms, LLC*, 562 F.3d 630, 645 (4th Cir. 2009); *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1168 (9th Cir. 2007); *Sony Computer Entm't, Inc. v. Connectix Corp.*, 203 F.3d 596, 609 (9th Cir. 2000); *Sega Enters. Ltd. v. Accolade, Inc.*, 977 F.2d 1510, 1527-28 (9th Cir. 1992). Similarly, the mass digitization of books for text-mining purposes is a form of incidental or "intermediate" copying that enables ultimately non-expressive, non-infringing, and socially beneficial uses without unduly treading on any expressive—*i.e.*, legally cognizable—uses of the works. The Court should find such copying to be fair use.

ARGUMENT

I. The Freedom to Make Non-expressive Use of Copyrighted Works is Vital to the "Progress of Science" in the Digital Humanities

Where large-scale electronic text collections are available, advances in computational power and a proliferation of new text-mining and visualization tools offer scholars of the humanities the chance to do what biologists, physicists, and economists have been doing for decades—analyze massive amounts of data.

"Digital Humanities" scholars fervently believe that text mining and the computational analysis of text are vital to the progress of human knowledge in the current Information Age. The potential of these non-expressive uses of text has already been revealed in the life sciences, where researchers routinely use a variety of text-mining tools to facilitate the search for relevant research across disparate fields and to uncover previously unnoticed "correlations or associations such as protein-protein interactions and gene-disease associations." *See* Sophia Ananiadou et al., *Text Mining and its Potential Applications in Systems Biology,* 24 TRENDS IN BIOTECHNOLOGY 571, 571 (2006) (citing Toshihide Ono et al., *Automated Extraction of Information on Protein–Protein Interactions from the Biological Literature*, 17 BIOINFORMATICS 155 (2001) and Christian Blaschke et al. *Information Extraction in Molecular Biology*, 3 BRIEFINGS IN BIOINFORMATICS 154 (2002)).

Similar breakthroughs are on the horizon in the humanities. Traditionally, literary scholars have relied upon the close and often anecdotal study of select works. Modern computing power, advances in computational linguistics and

natural language processing, and the mass digitization of texts now permit investigation of the larger literary record.

Digitization enhances our ability to process, mine, and ultimately better understand individual texts, the connections between texts, and the evolution of literature and language. As University of Nebraska Professor Matthew Jockers explains, by exploring the literary record writ large, researchers can better understand the *context* in which individual texts exist, and thereby better understand the texts themselves. *See* Matthew Jockers, MACROANALYSIS: DIGITAL METHODS FOR LITERARY HISTORY (2013). Along similar lines, Stanford University Professor Franco Moretti has noted that "a field this large cannot be understood by stitching together separate bits of knowledge about individual cases, because it *isn't* a sum of individual cases: it's a collective system, that should be grasped as such, as a whole . . . ." Franco Moretti, GRAPHS, MAPS, TREES: ABSTRACT MODELS FOR LITERARY HISTORY 4 (2005) (emphasis in original).

Researchers working in the field of information retrieval frequently use text mining and computer-aided classification to identify and retrieve relevant documents. Using similar techniques, researchers in the Digital Humanities are able to identify and retrieve relevant texts, often from unlikely places. Humanities researchers can thereby expand their traditional study of a few canonical works to a study of several million in the larger archive of literary history—an archive that

has hitherto remained hidden because of the limitations of humans' reading capacity. As part of this process, such non-expressive uses often leads to additional expressive uses, expanding the audience (and the potential market) for enjoyment of individual works.[3]

Mass digitization also results in the creation of data that enables scholars to reimagine relationships between texts—for example, by linking texts with maps. Thus, Google's "Ancient Places Project" links the text of public domain books like *Gibbon's Decline and Fall of the Roman Empire* to a map of the ancient world.[4] The interface allows the user to browse the books, including the full text, at the same time as she browses a map. The places mentioned are marked on the map and hyperlinked.[5] Similar maps could be made with reference to works still under

---

[3] For example, Matthew Jockers used text mining and computer aided classification to identify an overlooked tradition of whaling fiction predating (and arguably informing) Melville's writing of *Moby Dick*. *See* Jockers, *supra*.

[4] *See* Leif Isaksen, Elton Barker, Eric C. Kansa, Kate Byrne, *GAP: A NeoGeo Approach to Classical Resources*, 45 LEONARDO 82-83 (2012).

[5] In a similar vein, researchers at Stanford University have mapped thousands of letters exchanged during the Enlightenment and thereby devised a theory of how these individual networks fit into a coherent whole, which the scholars refer to as the "Republic of Letters." *See Tracking 18th-century "social network" through letters*, STANFORD UNIVERSITY (Dec. 14, 2009) (video), http://www.youtube.com/watch?v=nw0oS-AOIPE. Such aggregation yields surprising insights: for example, "the common narrative is that the Enlightenment started in England and spread to the rest of Europe," but the relatively low volume of correspondence between London and Paris suggests otherwise. *See* Patricia

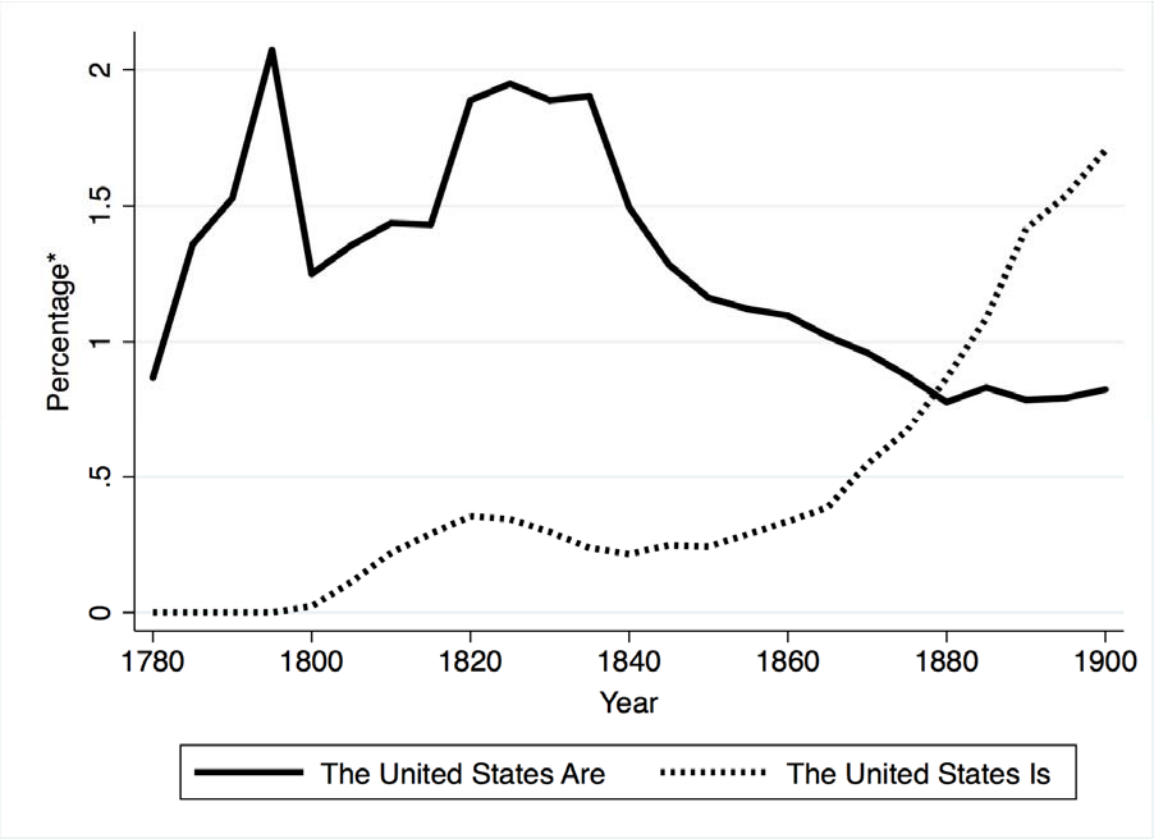
copyright—importantly, *without* ever making the text of the book available for free viewing. Extracting such data from texts to create these maps is a quintessential *non-expressive* use of the underlying texts that does not implicate any copyright-protected use—let alone infringe the copyrights of—the works in question.

Google's "Ngram" tool provides another example of a non-expressive use enabled by mass digitization—this time easily visualized. Figure 1, below, is an Ngram-generated chart that compares the frequency with which authors of texts in the Google Book Search database refer to the United States as a single entity ("is") as opposed to a collection of individual states ("are"). As the chart illustrates, it was only in the latter half of the Nineteenth Century that the conception of the United States as a single, indivisible entity was reflected in the way a majority of writers referred to the nation. This is a trend with obvious political and historical significance, of interest to a wide range of scholars and even to the public at large. But this type of comparison is meaningful only to the extent that it uses as raw data a digitized archive of significant size and scope.[6]

Cohen, *Digital Keys for Unlocking the Humanities' Riches*, N.Y. TIMES, Nov. 17, 2010, at C1.

[6] Google Ngram is available at http://books.google.com/ngrams.

Figure 1: Google Ngram Visualization Comparing Frequency of "The United States is" to "The United States are"[7]



To be absolutely clear, 1) the data used to produce this visualization can *only* be collected by digitizing the entire contents of the relevant books, and 2) not a *single sentence* of the underlying books has been reproduced in the finished product. In other words, this type of non-expressive use only adds to our collective

[7] Figure 1 is a reconstruction of data generated using Google Ngram, sampled at five-year intervals. The y-axis is scaled to 1/100,000 of a percent, such that 1 = 0.00001%.

knowledge and understanding, without in any way replacing, damaging the value of, or interfering with the market for, the original works.[8]

Google Ngram is just the tip of the iceberg.[9] In *Macroanalysis: Digital Methods and Literary History*, Professor Jockers draws on a corpus of Nineteenth Century novels to demonstrate how literary style changes over time. *See generally* Jockers, *supra*. Examining word frequencies, syntactic patterns, and thematic markers in the metadata-enriched context of author nationality, author gender, and time period, opens up literary study to an entirely new perspective.[10] Trendsetters

---

[8] For additional examples of Ngram's uses, *see*, *e.g.*, Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden; *Quantitative Analysis of Culture Using Millions of Digitized Books*. 331 SCIENCE 176 (2011) (a study of linguistic and cultural changes in over five million digitized books).

[9] The toolkit available to Digital Humanities researchers is becoming increasingly sophisticated. *See, e.g.*, Text Analysis Portal for Research ("TAPoR"), http://portal.tapor.ca/portal/portal (last visited May 21, 2013) (tools to map word usage over time, including peaks, density, collocations, and types); MALLET: MAchine Learning for LanguagE Toolkit, http://mallet.cs.umass.edu/ (last visited May 31, 2013) (a Java-based package for statistical natural language processing, document classification, clustering, topic modeling, information extraction, and other machine learning applications to text); MONK: Metadata Offer New Knowledge, http://www.monkproject.org/ (last visited May 31, 2013) (a digital environment designed to help humanities scholars discover and analyze patterns in the texts); Software Environment for the Advancement of Scholarly Research ("SEASR"), http://seasr.org (last visited May 31, 2013).

[10] A recently published study, led by mathematicians at Dartmouth, makes a similar point. *See* James M. Hughes et al., *Quantitative Patterns of Stylistic*

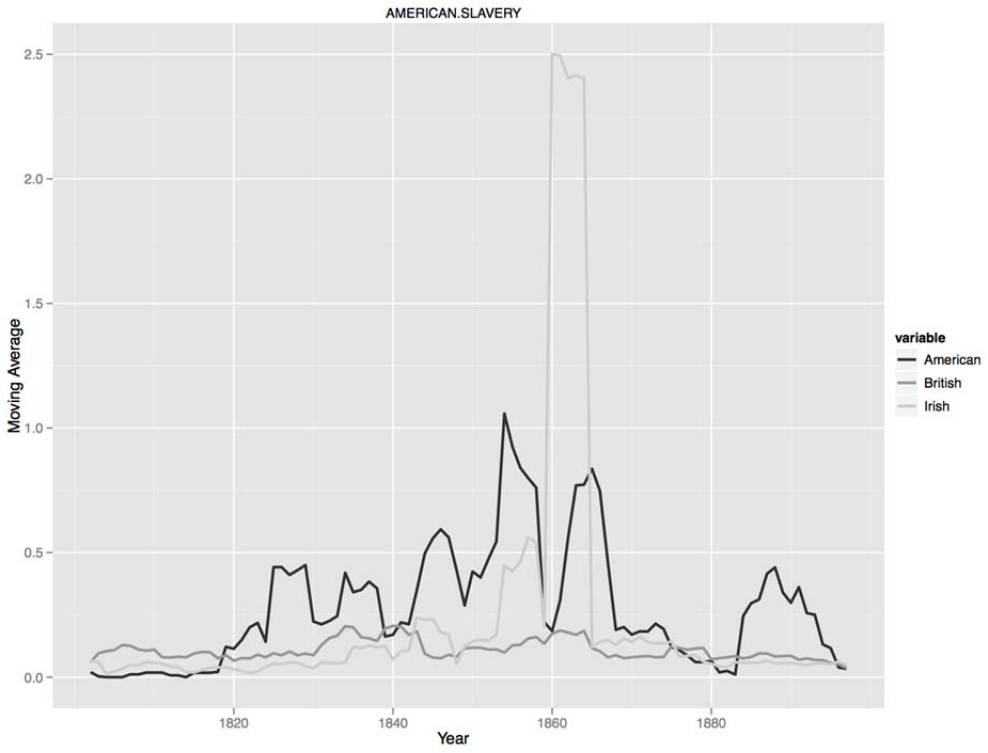
and outliers are revealed, as when Jockers' text mining and computational analysis demonstrated that Harriet Beecher Stowe's fiction is far more similar to the work of male authors of her generation than to the female-authored works of "sentimental fiction" among which her work has traditionally been categorized.[11]

Figure 2 provides another fascinating example of Professor Jockers' research. The chart shows the extent to which British, American, and Irish authors focused on the theme of American slavery during the Nineteenth Century, based on a corpus of 3,450 novels from that time period. Although it comes as no surprise that slavery was most often addressed by American authors, the strong Irish reaction to the American Civil War (note the spike in the light gray line beginning in 1860) compared with the decidedly muted response by British authors invites—indeed, demands—further investigation.

Influence in the Evolution of Literature, 109 PROC. OF THE NAT'L ACAD. OF SCI. OF THE U.S. 7682 (2012).

[11] Steve Lohr, *Dickens, Austen and Twain, Through a Digital Lens*, N.Y. TIMES, Jan. 26, 2013, at BU3, *available at http*://www.nytimes.com/2013/01/27/technology/literary-history-seen-through-big-datas-lens.html?pagewanted=all&_r=2&.

Figure 2: American Slavery in American, English, and Irish Literature, 1800-1899.



As Jockers' work reveals, "macroanalysis" of text archives has the potential to provide insight into historical literary questions, such as the place of individual texts, authors, and genres in relation to a larger literary context; literary patterns and lexicons employed over time, across periods, within regions, or within demographic groups; the cultural and societal forces that impact literary style and the evolution of style; the waxing and waning of literary themes; and the tastes and preferences of the literary establishment—and whether those preferences correspond to general tastes and preferences. However, *realizing this potential requires access to digitized texts*.

If libraries, research universities, non-profit organizations, and commercial entities are prohibited from making non-expressive use of copyrighted material, literary scholars, historians, and other humanists are restricted to becoming 19th-centuryists; slaves not to history, but to the public domain. History does not end in 1923.[12] But if copyright law prevents Digital Humanities scholars from using more recent materials, 1923 will be the effective end date of the work these scholars can do.

In short, the possibility of mining huge digital archives and manipulating the data collected in the process has inspired many scholars to re-conceptualize the very nature of humanities research. For others, it has played the more modest—but still valuable—role of providing new tools for testing old theories, or suggesting new areas of inquiry. *None of this*, however, can be done in the modern context if scholars cannot make non-expressive uses of underlying copyrighted texts, which (as shown above) will frequently number in the thousands, if not millions. Given copyright law's objective of promoting "the Progress of Science,"[13] it would be perversely counterintuitive if the promise of Digital Humanities were extinguished in the name of copyright protection.

---

[12] Due to repeated extensions of the copyright term, U.S. copyrights after 1923 do not automatically expire on an annual basis; thus, most modern works are still copyrighted. *See Eldred v. Ashcroft*, 537 U.S. 186 (2003).

[13] U.S. Const. Art I., Sec. 8. "Science," as used in the Constitution, referred to knowledge and learning.

II. COPYRIGHT LAW DOES NOT PROTECT NON-EXPRESSIVE ASPECTS OF WORKS

Fortunately, this Court need not contemplate such a scenario, as non-expressive aspects of copyrighted works—*e.g.*, the facts and ideas contained within the work and concerning it—are not protected by copyright. Such fundamental legal principles as the "idea/expression" distinction (reflected in Section 102(b) of the Copyright Act), the "merger" doctrine, the rule of "*scènes à faire*," and the "fact/expression" distinction all reflect this basic tenet. Metadata—information *about* copyrighted works collected through data mining and used by Digital Humanities scholars in the research described above—either does not implicate copyright protection at all, or is inoculated by the aforementioned doctrines that limit authors' rights to their works' expressive content.

A. The Idea/Expression Distinction

Copyright gives authors the right to set the terms upon which their original expression is made available to the public. But this right is not unlimited. As one of the fundamental—and Constitutional—limitations on those rights, the idea-expression distinction strikes a balance between "the interests of authors . . . in the control and exploitation of their writings . . . on the one hand, and society's competing interest in the free flow of ideas, information, and commerce on the other hand." *Harper & Row Publishers, Inc. v. Nation Enters.*, 471 U.S. 539 (1985) (quoting *Sony Corp. of Am. v. Universal City Studios, Inc.*, 464 U.S. 417, 429

(1984)); *see also Golan v. Holder*, 132 S. Ct. 873, 890 (2012) (describing the idea-expression distinction as one of copyright's "built-in First Amendment accommodations"). Copyright law protects only *expressive* use: "It is an axiom of copyright law that the protection granted to a copyrightable work extends only to the particular expression of an idea and never to the idea itself." *Reyher v. Children's Television Workshop*, 533 F.2d 87, 90 (2d Cir. 1976).

B. Section 102(b)

Recognizing the importance of access to ideas within expressive works, Congress has placed statutory limits on the rights of copyright holders through Section 102(b) of the Copyright Act, which provides: "In no case does copyright protection for an original work of authorship extend to any idea . . . concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work." 17 U.S.C. § 102(b) (2006). This provision has played a key role in modern copyright cases, ensuring that access to non-expressive aspects of works is not inhibited. *See, e.g., Peter F. Gaito Architecture, LLC v. Simone Dev. Corp.*, 602 F.3d 57, 67 (2d Cir. 2010) (holding that the principle behind § 102(b) required the court "to determine whether . . . 'similarities are due to protected aesthetic expressions original to the allegedly infringed work, or whether the similarity is to something in the original that is free for the taking' " (quoting *Tufenkian Imp./Exp. Ventures, Inc. v. Einstein Moomjy, Inc.*, 338 F.3d

127, 134-35 (2d Cir. 2003))). As noted above, text mining extracts and compiles ideas, concepts, and principles in copyrighted works into metadata. This process generates the very types of "discovery" that § 102(b) envisions.

C. Merger and *Scènes à Faire*

The policy of excluding non-expressive elements from copyright protection is so strong that—even in situations where expressive and non-expressive elements intertwine—doctrines like that of "merger" and "*scènes à faire*" preclude copyright protection *for expression* "in those instances where there is only one or so few ways of expressing an idea that protection of the expression would effectively accord protection to the idea itself." *Kregos v. Associated Press,* 937 F.2d 700, 705 (2d Cir. 1991); *see also New York Mercantile Exch., Inc. v. IntercontinentalExchange, Inc*., 497 F.3d 109, 118 (2d Cir. 2007). The "merger" doctrine is built upon the same principle as the idea/expression distinction: the protection of expressive elements of a work cannot, for Constitutional and practical reasons, interfere with the public's "free access to ideas." *New York Mercantile Exch., Inc.*, 497 F.3d. at 116. Relatedly, elements of a work that are *scènes à faire*—that is, "incidents, characters or settings which are as a practical matter indispensable, or at least standard, in the treatment of a given topic"—are not protectable. *Hoehling v. Universal City Studios, Inc*., 618 F.2d 972, 979 (2d Cir.

1980); *see also MyWebGrocer, LLC v. Hometown Info, Inc.*, 375 F.3d 190, 194 (2d Cir. 2004).

D. Fact/Expression Distinction

Finally, the monopoly rights of authors cannot extend to factual elements that "do not owe their origin to an act of authorship." *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co., Inc.*, 499 U.S. 340, 347 (1991). "The distinction is one between creation and discovery: The first person to find and report a particular fact has not created the fact; he or she has merely discovered its existence." *Id.* The Supreme Court in *Feist* made clear that if an "author clothes facts with an original collocation of words, he or she may be able to claim a copyright in this written expression;" *nevertheless*, "[o]thers may copy the underlying facts from the publication . . . ." *Id*. at 348.

In *National Basketball Association v. Motorola, Inc.*, 105 F.3d 841 (2d Cir. 1997), for example, a sports reporting service distributing real-time game statistics based on a data feed from reporters was held non-infringing. This Court reasoned that "[b]ecause [the service reproduced] only factual information culled from the broadcasts and none of the copyrightable expression of the games, appellants did not infringe the copyright of the broadcasts." *Id*. at 847. This Court has similarly held that one has "the right to avail himself of the facts contained in [another's] book and to use such information, whether correct or incorrect, in his own literary

work." *Hoehling*, 618 F.2d at 979. In other words, copyright law clearly distinguishes between expressive and non-expressive content, and deems only *expressive* content protectable.

E. Non-expressive Metadata Does Not Implicate the Statutory Rights of the Copyright Holder

Metadata about a copyrighted work does not implicate any legally cognizable interest of the copyright holder. Metadata may contain facts about the works themselves, might capture (in different terminology) the ideas contained within the text, or may convey information such as the number of times a given word appears in a particular text, how often a particular author uses a specific literary device, or the essence of what the work is about. Though it is true that metadata would not exist but for the underlying work, *it does not contain the expression of the work.*

Consider, for example, two facts about *Moby Dick*: first, that the word "whale" appears 1119 times; second, that the word "dinosaur" appears 0 times. While *a whale* is certainly central to the expression contained in *Moby Dick*, this data is not. Rather, metadata of this sort—a simplified version of the metadata surveyed in Section I—is factual and non-expressive, and incapable of infringing the rights of copyright holders.

The same principle can be illustrated using a recent decision of the court below, *Warner Brothers Entertainment Inc. v. RDR Books*, 575 F. Supp. 2d 513 (S.D.N.Y. 2008). Consider the following four statements:

> [1] "Goblin-made armour does not require cleaning, simple girl. Goblins' silver repels mundane dirt, imbibing only that which strengthens it."
>
> [2] "goblin-made armor does not require cleaning, because goblins' silver repels mundane dirt, imbibing only that which strengthens it, such as basilisk venom."
>
> [3] "Statement [1] contains twenty words, and other than 'Goblin', no word in expression [1] is repeated."
>
> [4] "Statement [2] is strikingly similar to Statement [1]."

Statement [1] originates with J.K. Rowling, the author of the *Harry Potter* novels. *See Warner Bros.*, 575 F. Supp. 2d at 527 (quoting J.K. Rowling, *Harry Potter and the Deathly Hallows* 303 (2007)). Statement [2] was held out as originating with a contributor to the *Harry Potter Lexicon* (a reference work for the "*Harry Potter* universe"), which was found to infringe because too much of its contents consisted of direct quotations or close paraphrases of vivid passages in the Harry Potter books, as the comparison between [1] and [2] illustrates. *Id.* at 527. Statements [3] and [4], by contrast, are classic metadata; they would not exist but

for the underlying work, and yet neither passage is substantially similar—or indeed, bears any resemblance at all—to the expressive elements of the underlying work.

Even more importantly, this metadata *does not originate with the author* of the underlying work. As the Supreme Court held in *Feist Publications*, "copying of constituent elements of the work that are *original*" is an essential element of a copyright infringement claim. 499 U.S. at 361 (emphasis added); *see also* 17 U.S.C. § 102(a) (2006).

Amici wish to emphasize that metadata is *not* the same thing as so-called "invented facts." J.K. Rowling's conception and description of goblin armor and thousands of other details in the Harry Potter series could be regarded as "invented facts" because, quite simply, she made them up. As laid out in the case law, if such facts and their associated expressive descriptions are reproduced in sufficient quantity, they may "constitute creative expression protected by copyright because characters and events spring from the imagination of the original authors." *Warner Bros.*, 575 F. Supp. 2d at 536 (quoting *Castle Rock Entm't Inc. v. Carol Publ'g Grp., Inc.*, 150 F.3d 132, 139 (2d Cir. 1998)). Metadata, however, cannot be accurately characterized as "invented facts," but only as facts *about* "invented facts." The distinction is significant: once again, facts are not eligible for copyright protection.

Nor does metadata infringe the author's right "to prepare derivative works based upon the copyrighted work[.]" 17 U.S.C. § 106(2) (2006). As the court below held in *Warner Brothers*, an analytical work that provides insight into a copyrighted work but does not "recast, transform, or adapt" that work does not violate the derivative work right. 575 F. Supp. 2d at 539; *see also Ty, Inc. v. Publ'ns Int'l Ltd.*, 292 F.3d 512, 520 (7th Cir. 2002) (holding that collectors' guide to certain copyrighted works did not violate 17 U.S.C. § 106(2) because the guides did not "recast, transform, or adapt the things to which they are guides").

Amici urge the Court to carefully distinguish the facts of the instant case from those in *Castle Rock Entertainment v. Carol Publishing Group*, 150 F.3d 132 (2d Cir. 1998). In *Castle Rock*, this Court held that a quiz book based on the popular television series "Seinfeld" was, quantitatively and qualitatively, substantially similar to that series, considered as a whole. *Id.* at 138–39. The quiz book in that case, however, was not an analytical work; rather, it essentially recast "Seinfeld's" copyrightable characters into a new format, as if the defendant had made miniature dolls of those same characters. *See Hasbro Bradley, Inc. v. Sparkle Toys, Inc.*, 780 F.2d 189, 192-93 (2d Cir. 1985) (upholding copyrightability of "Transformer" robotic action figures as sculptural works). The supposed "facts" conveyed in the "Seinfeld" quiz book were not truly *facts* about the television

program; they were "in reality fictitious expression created by *Seinfeld*'s authors." *Castle Rock Entm't*, 150 F.3d at 139.

By contrast, the many forms of metadata produced by the library digitization at the heart of this litigation *do not* merely recast copyrightable expression from underlying works; rather, the metadata encompasses numerous uncopyrightable facts *about* the works, such as author, title, frequency of particular words or phrases, and the like.

F. Non-expressive Metadata Is Also Noninfringing Because It Does Not Allow the Public to Perceive the Expressive Content of a Work

The significance of public perception runs deep in copyright law. Indeed, controlling authority suggests that the copyright holder's exclusive rights are limited to the right to communicate the expressive aspects of her work to the public. For example, in *New York Times Co. v. Tasini*, 533 U.S. 483 (2001), a case about the scope of the 17 U.S.C. § 201(c) "privilege" of the copyright owner to reproduce and distribute individual contributions "as part of [a] collective work," the Supreme Court held that "[i]n determining whether the Articles [at issue] have been reproduced and distributed as part of a revision of the collective works in issue, we focus on the Articles *as presented to, and perceptible by, the user[s]* of the Databases [containing the Articles]." 533 U.S. at 499 (emphasis added; internal quotation marks and citations omitted). The Court elaborated: "the question is not whether a user can generate a revision of a collective work from a database, but

whether the database itself *perceptibly presents the author's contribution* as part of a revision of the collective work." *Id.* at 504 (emphasis added).

This point is especially evident in cases where plaintiffs have argued that, although a defendant's final product does not support an allegation of infringement, the defendant has violated the Copyright Act by making a reproduction of the plaintiff's work that is merely intermediate and *imperceptible to the reading public*. In *Davis v. United Artists, Inc.*, for example, the court below rejected out of hand the allegation that the defendant's unpublished screenplays were substantially similar to plaintiff's novel, refusing to "consider the preliminary scripts" because "the ultimate test of infringement must be the film as produced and broadcast" to the public. 547 F. Supp. 722, 724 n.9 (S.D.N.Y. 1982). *See also Fuld v. Nat'l Broad. Co., Inc.*, 390 F. Supp. 877, 882 n.4 (S.D.N.Y. 1975) ("[T]he ultimate test of infringement must be the television film as produced and broadcast — and not the preliminary scripts . . . ."); *Walker v. Time Life Films, Inc.*, 615 F. Supp. 430, 434 (S.D.N.Y. 1985) ("The Court considers the works as they were presented to the public.").[14]

[14] Courts in other circuits have adopted the same view. *See*, *e.g.*, *Stromback v. New Line Cinema*, 384 F.3d 283, 299 (6th Cir. 2004) ("In deciding infringement claims, courts have held that only the version of the alleged infringing work presented to the public should be considered"); *Madrid v. Chronicle Books*, 209 F. Supp. 2d 1227, 1234 (D. Wyo. 2002) ("Since a court considers the works as they were presented to the public, discovery in this case . . . would be pointless") (internal quotation marks omitted).

III. Text Mining Creates Value by Facilitating the Advancement of Our Collective Knowledge; To Protect That Value, Mass Digitization and Similar Intermediate Copying for Data Mining and Other Non-expressive Purposes Should Be Considered "Fair Use"

As demonstrated above, non-expressive metadata itself is noninfringing. However, *Amici* recognize that this Court must also consider the legality of the process of making copies to generate that metadata. Fortunately, numerous courts have held that copying to enable purely non-expressive uses, such as the automated extraction of data, does not infringe the statutory rights of the copyright holder. Like copying employed for other transformative purposes, such as parody, criticism, and reverse engineering, intermediate copying for the purpose of extracting non-expressive metadata is fair use.

A. Non-expressive Copying to Expand Our Knowledge in the Digital Humanities Is An Activity of the Sort that Copyright Law Should Favor, Through Fair Use

First among the statutory factors relevant to a fair use analysis is "the purpose and character of the use, including whether such use is of a commercial nature or is for nonprofit educational purposes." 17 U.S.C. § 107(1). Like more traditional expressive transformative uses, the more "non-expressive" the use of a copyrighted work, the less it substitutes for the author's original expression. As such, non-expressive uses are properly considered equivalent to (or a subset of) highly transformative uses: their "purpose and character" is such that they do not

merely supersede the objects of the original creation. *Campbell v. Acuff-Rose Music, Inc.*, 510 U.S. 569, 583 (1994). *See also Cariou v. Prince*, No. 11-1197-cv,___ F.3d __, slip op. at 13 (2d Cir., April 25, 2013) (finding that defendant's "composition, presentation, scale, color palette, and media are fundamentally different and new compared to [Plaintiff's] photographs, as is the expressive nature of [defendant]'s work."); *Perfect 10, Inc. v. Amazon.com, Inc.*, 508 F.3d 1146, 1165 (9th Cir. 2007) (holding that search engines are "highly transformative" because "[a]lthough an image may have been created originally to serve an entertainment, aesthetic, or informative function, a search engine transforms the image into a pointer directing a user to a source of information"); *Kelly v. Arriba Soft Corp.*, 336 F.3d 811, 818 (9th Cir. 2002) (holding that use of images in search engine was transformative because they served "as a tool to help index and improve access to images on the internet and their related web sites" and their use was "unrelated to any aesthetic purpose"); *Bill Graham Archives v. Dorling Kindersley Ltd.*, 448 F.3d 605, 609 (2d Cir. 2006) (finding critical to fair use analysis that publisher's use of copyrighted images of concert posters in book was "plainly different from the original purpose for which they were created"). As the process of digitization for text mining is intermediate and non-expressive, and its purpose is to produce non-expressive metadata, this factor favors fair use.

Moreover, "there is a strong presumption that factor one [in the fair use analysis] favors the defendant if the allegedly infringing work fits the description of uses described in [17 U.S.C.] § 107," which includes "scholarship" and "research." *NXIVM Corp. v. Ross Institute*, 364 F.3d 471, 477 (2d Cir. 2004). The crucial role that mass digitization plays in promoting the progress of research and scholarship in the Digital Humanities weighs heavily in favor of fair use here. *See also* Pierre N. Leval, *Toward A Fair Use Standard*, 103 HARV. L. REV. 1105, 1111 (1990) ("If [a] secondary use adds value to the original – if the quoted matter is used as raw material, transformed in the creation of new information, new aesthetics, new insights and understandings – this is the very type of activity that the fair use doctrine intends to protect for the enrichment of society.")

Similarly, courts have ruled in favor of fair use when copying allowed defendants or third parties to use facts from copyrighted works in news reporting or court proceedings. *See*, *e.g.*, *Bond v. Blum*, 317 F.3d 385, 395 (4th Cir. 2003) (holding that "the narrow purpose of defendants' use of the manuscript . . . for the evidentiary value of its content" weighed "heavily" against a finding of infringement); *Religious Tech. Ctr. v. Lerma*, 908 F. Supp. 1362, 1366 (E.D. Va. 1995) (finding fair use in part because documents were copied for "news gathering, news reporting and responding to litigation," not to "scoop" copyright owner). Significantly, both the *Bond* and *Religious Tech. Ctr.* courts' fair use holdings went

further than the text mining at issue here, because the users in those cases had to glean the necessary facts by reading the materials, rather than mining the text with computers. *Bond*, 317 F.3d at 393; *Religious Tech. Ctr.*, 908 F. Supp. at 1364-65. If a human's *reading* of copyrighted expression to extract non-expressive material is fair use, the result should be the same when a computer performs the extraction.

B. The Nature of the Works in Question Is Favorable to the Fair Use Analysis of Mass Digitization for the Advancement of Digital Humanities Research and Scholarship

When the purpose of a secondary use is socially beneficial, the second fair use factor, "the nature of the copyrighted work," is rarely dispositive. *See, e.g.*, *Bill Graham*, 448 F.3d at 612 ("The second factor may be of limited usefulness where the creative work of art is being used for a transformative purpose.") This is especially true in "intermediate copying" cases like this one, where the material ultimately reaching the user is not the expressive content of the copyrighted work at all, but rather ideas contained within it or facts about it.

Nevertheless, to the extent that the second fair use factor is relevant here, it weighs in favor of fair use. Looking to this factor, "[c]ourts generally hold that 'the scope of the second fair use is greater with respect to factual than non-factual works'. . . . [F]ictional works, on the other hand, . . . require more protection." *Basic Books, Inc. v. Kinko's Graphics Corp.*, 758 F. Supp. 1522, 1533 (S.D.N.Y. 1991) (quoting *New Era Publications Int'l, ApS v. Carol Pub. Group*, 904 F.2d 152,

157 (2d Cir. 1990)). A detailed study of the copyrighted works in the collections from which Google has created its digitized corpus have concluded that the "overwhelming majority – 92 Percent . . . – were non fiction." Brian Lavoie & Lorcan Dempsey, *Beyond 1923: Characteristics of Potentially In Copyright Print Books in Library Collections*, 15 D-Lib Mag., http://www.dlib.org/dlib/november09/lavoie/11lavoie.html.

Furthermore, as one court explained, the second fair use factor weighs in favor of fair use where humans "cannot gain access to the unprotected ideas and functional concepts contained in [the copyrighted work] without . . . making copies." *Sega*, 977 F.2d at 1525. This is effectively the case for Digital Humanities scholars, as there are no plausible ways to conduct analyses of the sort described in Section I other than mass digitization and algorithmic analysis, both of which require making intermediate copies.

C. To the Extent Relevant, Mass Digitization Uses a Reasonable "Amount and Substantiality" of the Works in Question, in Light of the Socially Beneficial Purpose of Facilitating Data Mining for the Advancement of the Digital Humanities

The third fair use factor asks whether the amount and substantiality used are "reasonable in relation to the purpose of the copying." *Campbell*, 510 U.S. at 586–87. Because the metadata created here does not contain any infringing material, the third factor "is of very little weight." *See*, *e.g.*, *Connectix*, 203 F.3d at 606. This is

true even where many intermediate copies are made. *Id.* at 601. Moreover, as Section I shows, it is not only reasonable to use mass digitization of an entire set of works to enable the creation of noninfringing metadata about those works, it is a practical necessity, as there is no equivalent human means of doing so. In order for Digital Humanities research and scholarship to be as accurate and complete as possible, every word or image in a copyrighted work must be mined.

Other courts have relied upon similar rationales to support full copying in intermediate and non-expressive fair use cases. *See*, *e.g.*, *Cariou v. Prince,* No. 11-1197-cv,___ F.3d __, slip op. (2d Cir., April 25, 2013); *Vanderhye*, 562 F.3d at 642 (finding mass digitization of entire student essays to be fair use when reasonable as a means to check for plagiarism); *Perfect 10*, 508 F.3d at 1167-68 (finding thumbnail reproduction of entire photographs reasonable in light of defendant's use of the images to improve access to information on the internet versus artistic expression); *Kelly*, 336 F.3d 820-21 (same); *Bond*, 317 F.3d at 396 (noting that "[t]he use of the copyrighted material [as evidence in a custody proceeding], even the entire manuscript, does not undermine the protections granted by the [Copyright] Act"). In light of practical necessity and ample precedent in support, the Court should find that the "amount and substantiality" factor favors the making of intermediate copies for non-expressive use.

D. Allowing Intermediate Copying in Order to Enable Non-expressive Uses Does Not Harm the Market for the Original

Works in a Legally Cognizable Manner, As The Practice Does Not Implicate the Works' Expressive Aspects in Any Way

The fourth statutory fair use factor is "the effect of the use upon the potential market for or value of the copyrighted work." In the case of expressive uses such as parody, and non-expressive uses such as reverse engineering, courts have consistently held that the protection that copyright affords is limited to certain cognizable markets. *Campbell*, 510 U.S. at 591-92 ("[W]hen a lethal parody, like a scathing theater review, kills demand for the original, it does not produce a harm cognizable under the Copyright Act."); *Sega*, 977 F.2d at 1523-24. Transformative expressive uses do not usually affect the market in any relevant sense because the second author's expression does not substitute for that of the original author. *Campbell*, 510 U.S. at 591; *Fisher v. Dees*, 794 F.2d 432, 438 (9th Cir. 1986) ("This is not a case in which commercial substitution is likely . . . . The two works do not fulfill the same demand."). As illustrated by the examples in Section I, above, non-expressive uses have no potential substitution effect on any legally cognizable market for copyrighted works, because copyright only protects markets for *expression*, *and not* markets for discoveries, ideas, facts, principles, or concepts. *See*, *e.g.*, *Vanderhye*, 562 F.3d at 644 ("[N]o market substitute was created by [defendants], whose archived student works do not supplant the plaintiffs' works . . . so much as merely suppress demand for them . . . In our view, then, any harm here is not of the kind protected against by copyright law.") Indeed, in many

instances, the use of metadata made by scholars could actually enhance the market for the underlying work, by causing researchers to revisit the original work and reexamine it in more detail.

In short, there is no reason to disallow the digitization of libraries, whether by libraries themselves, or commercial search engine companies, so long as that digitization is for non-expressive use. Non-expressive uses such as those practiced in the Digital Humanities hold great promise for *Amici*, other scholars, society at large—and copyright owners, too.

DATED: June 4, 2013
New York, New York

Respectfully Submitted,

/s/ Jason Schultz

Jason Schultz
UC Berkeley School of Law
Counsel for Amici (with Matthew Sag)

CERTIFICATE OF COMPLIANCE WITH FRAP 32(A)

1. This brief complies with the type-volume limitation of Fed. R. App. P. 32(a)(7)(B) and 29(d) because this brief contains **6984** words, excluding the parts of the brief exempted by Fed. R. App. P. 32(a)(7)(B)(iii).

2. This brief complies with the typeface requirements of Fed. R. App. P. 32(a)(5) and the type style requirements of Fed. R. App. P. 32(a)(6) because this brief has been prepared in a proportionally spaced typeface using Microsoft Word in Times New Roman, 14 point font.

/s/ Jason M. Schultz

APPENDIX A

APPENDIX A

The Association for Computers and the Humanities
http://www.ach.org

The Canadian Society for Digital Humanities
http://csdh-schn.org

Andrew Adams
Student
Stanford University

Kristin W. Andrews
Social Sciences & Humanities Librarian
University of North Carolina Wilmington

Jonathan Askin
Professor
Brooklyn Law School
Founder/Director Brooklyn Law Incubator & Policy Clinic

Elton Barker
Reader
Open University

Christina Bell
Humanities Librarian
Bates College

Michael Black
Graduate Student
University of Illinois at Urbana-Champaign

Chris Bourg
Assistant University Librarian for Public Services
Stanford University

Daniel Boyarin
Hermann P. and Sophia Taubman Professor of Talmudic Culture
Departments of Near Eastern Studies and Rhetoric

University of California at Berkeley

Collin Gifford Brooke
Associate Professor of Rhetoric and Writing
Syracuse University

Susan Brown
Professor
University of Guelph/University of Alberta

Patrick J. Burns
Senior Teaching Fellow
Fordham University

Kate Byrne
Research Fellow
School of Informatics, University of Edinburgh

David Carroll
Associate Professor
Parsons The New School for Design

Carol Chiodo
PhD candidate
Yale University

Margaret Chon
Donald and Lynda Horowitz Professor for the Pursuit of Justice
Seattle University School of Law

Eve V. Clark
Professor
Stanford University

Dr. Daniel Cohen
Executive Director of the Digital Public Library of America
Digital Innovation Fellow, American Council of Learned Societies

James Coltrain
Assistant Professor of History

University of Nebraska

Paul Conway
Associate Professor
University of Michigan School of Information

Ryan Cordell
Assistant Professor of English
Northeastern University

Brian Croxall
Digital Humanities Strategist and Lecturer of English
Emory University

Michael Scott Cuthbert
Homer A. Burnell Associate Professor of Music
MIT

Johanna Drucker
Bernard and Martin Breslauer Professor of Bibliography
Department of Information Studies at the Graduate School of Education and Information Studies
UCLA

G. Cory Duclos
Assistant Professor
Spring Hill College

Hoyt N. Duggan
Professor emeritus
University of Virginia

Morris Eaves
Professor of English and Turner Prof. of Humanities
University of Rochester
Co-Editor, William Blake Archive (www.blakearchive.org)

Penelope Eckert
Albert Ray Lang Professor of Humanities and Sciences
Professor by Courtesy of Anthropology

Stanford University

Jacob Eisenstein
Assistant Professor
Georgia Institute of Technology

James Evans
Graduate Student
CUNY Graduate Center

Dr. Marco Forlivesi
Director of the Digital Archive of Inaugural Lectures at Renaissance and Early Modern Universities
Università degli Studi di Chieti e Pescara, Italy

Rosemary Franklin
Research Librarian
Langsam Library
University of Cincinnati

Bernard Frischer
Professor
Departments of Art History and Classics
University of Virginia

Shubha Ghosh
Professor
University of Wisconsin

Alex Gil
Digital Scholarship Coordinator
Columbia University

Melissa Girard
Assistant Professor of English
Loyola University Maryland

Matthew K. Gold
Associate Professor of English and Digital Humanities
City Tech and Graduate Center, CUNY

Les Harrison
Associate Professor
Virginia Commonwealth University

Charles van den Heuvel
Professor
Royal Netherlands Academy of Arts and Sciences
Huygens Institute for the History of the Netherlands

Jeremy Hunsinger
Center for Digital Discourse and Culture
Virginia Polytechnic Institute and State University (Virginia Tech)

Dan Hunter
Professor
New York Law School
Director of the Institute for Information Law & Policy

Leif Isaksen
Deputy Director, Web Science Doctoral Training Centre
University of Southampton

Matthew Jockers
Assistant Professor of English
Fellow, Center for Digital Humanities Research
University of Nebraska, Lincoln

Dr. Eric Kansa
UC Berkeley School of Information
Alexandria Archive Institute
Lead Developer, Open Context (www.opencontext.org)

Dennis S. Karjala
Jack E. Brown Professor of Law
Sandra Day O'Connor College of Law, Arizona State University

Matthew Kirschenbaum
Associate Professor of English
University of Maryland

Hubertus Kohle
Professor
University of Munich, Germany

Kari Kraus
Assistant Professor
University of Maryland

Lore Kuehnert
Instructor, History
Hagerstown Community College

John Laudun
Associate Professor
University of Louisiana

Konrad M. Lawson
Max Weber Postdoctoral Fellow
European University Institute

William R. Leben
Professor of Linguistics Emeritus
Stanford University

Jarom McDonald
Associate Research Professor
Brigham Young University
Director, BYU Office of Digital Humanities

Erin McKean
Founder
Wordnik.com

Mark P. McKenna
Professor of Law, Notre Dame Presidential Fellow
Notre Dame Law School

Elijah Meeks
Digital Humanities Specialist

Stanford University Libraries

Richard Menke
Associate Professor of English
University of Georgia

David Mimno
Postdoctoral Researcher
Princeton University

Sally Moffitt
Reference Librarian and Bibliographer
University of Cincinnati

Franco Moretti
Professor of English and Comparative Literature
Stanford University

Paige Morgan
PhD Student & Instructor
University of Washington

Brian D. Moss
Reference Coordinator
University of Kansas Libraries

Dr. James Murphy

Ira Steven Nathenson
Associate Professor
St. Thomas University School of Law

Bethany Nowviskie
Director, Digital Research & Scholarship
University of Virginia
Director of the UVA Library Scholars' Lab
President of the Association for Computers & the Humanities

Dr. Julianne Nyhan
University College London

Amy V. Ogden
Associate Professor
University of Virginia

Piotr Organisciak
University of Illinois at Urbana-Champaign

Moacir P. de Sá Pereira
PhD Candidate
University of Chicago

Dorothy Porter
Curator, Digital Research Services
University of Pennsylvania

Todd Presner
Professor and Chair
Digital Humanities Program
UCLA

Kenneth M. Price
Hillegass University Professor
University of Nebraska-Lincoln
Co-director of the Center for Digital Research in the Humanities.

Adam Rabinowitz
Assistant Professor
University of Texas

Dean Rehberger
Michigan State University
Director of MATRIX (a digital humanities center)

Kevin Reilly, MSN, RN
Doctoral Student
Graduate School of Eduction & Psychology, Pepperdine University

Allen Riddell
PhD Student

Duke University

Augusta Rohrbach
Associate Professor
Washington State University
Editor ESQ: A Journal of the American Renaissance

Brian Rosenblum
Head, Center for Faculty Initiatives
University of Kansas Libraries

Ivan A. Sag
Sadie Dernham Patek Professor in Humanities and Professor of Linguistics
Stanford University

Mark Sample
Associate Professor
George Mason University

Jeffrey T. Schnapp
Professor and faculty director of metaLAB (at) Harvard
Harvard University

Dr. Christof Schöch
University of Würzburg, Germany

Susan Schreibman
Professor
Trinity College Dublin

Kris Shaffer
Assistant Professor of Music Theory
Charleston Southern University

Crandall Shifflett
Professor Emeritus of History
Virginia Tech

Stéfan Sinclair
Associate Professor

McGill University

Timothy Tangherlini
Professor
The Scandinavian Section and Department of Asian Languages and Cultures
UCLA
2010-2011 Digital Innovation Fellow, American Council of Learned Societies

Laurie Taylor
Digital Humanities Librarian
University of Florida

Dennis Tenen
Assistant Professor
Columbia University

Rebecca Tushnet
Professor
Georgetown University, School of Law

Ted Underwood
Associate Professor of English
University of Illinois, Urbana-Champaign

Christopher Warren
Assistant Professor of English
Carnegie Mellon University
Project Director and co-principle investigator for Six Degrees of Francis Bacon

Scott Weingart
Indiana University

Andrew Whalen
Researcher
University of California at Berkeley

Roger Whitson
Assistant Professor of English
Washington State University

Matthew Wilkens
Assistant Professor of English
University of Notre Dame

Glen Worthey
Digital Humanities Librarian
Stanford University

Vika Zafrin
Institutional Repository Librarian
Boston University