

Title: Changing scholarly communication: the implications of granting copyright for born digital objects

Presenter: Lori M. Jahnke

Abstract:

Changing modes of scholarly communication that incorporate digital methodologies prompt the question of whether or not born digital objects should receive copyright under the Library of Congress system. This change in procedure could have a profound impact on the preservation of digital data, the development of standards, and the distribution of databases as born digital objects. Although copyright ostensibly protects the intellectual creator, it can facilitate information monopolies that stifle scholarly communication. This paper provides a critical review of the legal, social, and academic implications of copyright for born digital objects versus pursuing open access models of publication.

Author Information: Lori M. Jahnke, Ph.D.
S. Gordon Castigliano CLIR Fellow
The College of Physicians of Philadelphia
Philadelphia, PA 19103
lmjahnke@gmail.com
(504) 905-6605

CHANGING SCHOLARLY COMMUNICATION: THE IMPLICATIONS OF GRANTING COPYRIGHT FOR BORN DIGITAL OBJECTS

“Scholarly communication is redundancy. There is no scholarship without communication. When I write something, it’s not to make money, I write so somebody will read it. I just want to get it out there and easily found.”

– Sidney Verba, Harvard University Faculty of Arts and Sciences, quoted in SPARC Innovators, June 2008.

Now retired Harvard University Librarian, Sidney Verba, uttered these words following a landmark faculty vote to make open access the default policy for published journal articles produced by members of the Faculty of Arts and Sciences. The policy requires faculty authors to deposit an electronic copy of their articles in the university’s digital repository and, unless the policy is waived for a particular article, faculty authors automatically grant copyright to the university through the University License (Nguyen 2008). The University License attaches to faculty work before the transfer of copyright to the publisher allowing Harvard to standardize the terms of the Institution’s right to host the work of their faculty, make the work accessible, and it grants the right to make copies of the articles (e.g., for archiving). Through the work of several internal champions in building consensus and fostering collaboration among the faculty and librarians, Harvard University now has the right to maintain a comprehensive record of the work of its faculty.

Several universities have since followed Harvard’s example and have adopted a similar policy or are in the process of drafting an open access policy (e.g., Stanford University, UCLA, and University of Pennsylvania). Debates over open access and the use of a University License

to further institutional goals have been contentious in many circles with faculty members fearing that publishers will automatically reject submissions carrying license restrictions. This is a particular concern for younger faculty who may not have achieved tenure yet.

The efforts of the Harvard faculty and staff underscore many of the ways in which research and scholarly communication have changed since the early 1990's in response to technological advancements and emerging business models. Responding to concern over the enclosure of knowledge and copyright law that is out of step with current technology, open access has rapidly grown into a full-fledged movement that has profound implications for the day-to-day business of scholars (i.e., the communication of our work). For many, a commitment to open access is based on deeper philosophical motivations, such as the democratization of knowledge and stewardship of the commons.

There has been much discussion among legal scholars, librarians, and faculty regarding the challenges that born digital objects pose to a legal system and a publishing market that was based on the dissemination of analog (i.e., print) materials. Born digital objects are any materials that originate in digital form. Generally, this category does not include analog materials that have been digitized. Within the context of scholarly communication, the term born digital object usually refers to electronically published journal articles. Although continued discussion is needed on this front, in this paper I focus on another type of born digital object—research databases.

Traditionally research databases have not met the criteria for copyright and thus rarely enter the copyright/open access/scholarly communication debate. As we reinvent the ways in which we conduct research, collaborate across institutions, and disseminate our work, it may be worthwhile to take a step back and learn what lessons we can from our brief history of digital

scholarly communication. As we move toward building data conservation infrastructure (Lee, et al. 2009), my hope is that we take an active role as a discipline in defining our best practices for attribution, access, and data preservation.

THE DIGITAL DILEMMA FOR COPYRIGHT

Although I worked in a law library for years and handled countless law books, I rarely read beyond the title page to glean information other than that needed for item processing. It is now much to my surprise that I find the debate surrounding copyright, digital objects, and scholarly communication so fascinating. The rise of digital technologies and the rapid shift to the production and dissemination of knowledge electronically is culturally transformative. The dissonance between formalized structures of regulation (e.g., copyright and the publication system) and now routine modes of scholarly production and communication, demonstrate the profundity of this transformation. This is no minor transition in a profession where many of our traditions were established during the Middle Ages. Copyright is one of those traditions.

Copyright has its beginnings at the end of the 15th century in England with the invention of the printing press. In its earliest inception, copyright was perpetual and only members of the Company of Stationers could hold it (Ewing 2003:2). The Company of Stationers consisted of individuals in a number of professions related to the book arts (e.g., bookbinders, sellers, printers, text writers), but did not include authors. The purpose of copyright was to give a particular stationer the exclusive right to copy a work, which served as protection from rival publishers. Protecting authors by creating a system of attribution and financial incentives to produce was not the goal copyright in the early days of print.

The idea behind modern copyright is that it provides a limited period of monopoly as an incentive for authors to produce. As Maxwell (2004:12) summarizes, "...the U.S. Constitutional rationale for providing monopoly protection was based on the assumption that providing a limited monopoly for authors would lead to 'the progress of science and the useful arts' (US Constitution)." Since, in most cases copyright is not held by the author but transferred to the publisher, copyright gives *publishers* financial protection by guaranteeing a period of monopoly over the work (Ewing 2003:5). In the United States, the term of copyright is 70 years beyond the life of the author. Although the modern purpose of copyright is to balance the interests of authors, publishers, and the public, as numerous authors have pointed out copyright law is out of balance.

"Today's copyright laws and traditions are dissonant with modern culture and technology, and the dissonance has become more and more apparent in the past few years" (Ewing 2003). Ewing is far from alone in acknowledging the disconnection between modern copyright law and contemporary forms of communication. Legal scholar Lawrence Lessig, among others (Boyle 2007; Hess and Ostrom 2003; Suber 2007), has been an outspoken proponent of copyright law reform to accommodate electronic communication and prevent enclosure of knowledge (e.g., Lessig 2001, 2002, 2005). As Boyle (1996) pointed out, enclosure is caused by contradictions between intellectual property law and the capabilities of new technologies. While digital technologies have generated greater access to information of all kinds (e.g., science, art, current events, history), they also provide the means for profit-oriented corporations to extract value from resources that were previously held in common (Hess and Ostrom 2003:112).

The challenges of creating a copyright code suited to digital materials and the recognition that access to ideas and information is vital for promoting innovation (O'Sullivan 2008), has prompted some to question as to whether copyright is still an effective means for providing incentive for authors to produce in the digital age. To test whether US copyright code was the right policy choice to foster growth of the sciences, Maxwell (2004) developed a market simulation model to gauge the impact of strict vs. loose copy control on growth of the publishing market and its ability to support authors. The study showed that "...the same policy options have different effects depending on the condition of the publishing industry at any stage of development" (Maxwell 2004:12).

While strict copy control provided the greatest benefit in promoting the growth of an immature publishing industry and supported the largest number of authors, strict copy control had a very different effect on a *mature* publishing market. Maxwell found that looser copy control facilitated market growth in the digital age, which led to a greater variety of books available, more authors, and increased publisher profits. Maxwell (2004) concludes that in regards to copyright reform:

...the impacts of copyright and other policy options on mature publishing industries have important implications for current debates about copyright in the digital age...current efforts to strengthen copy control over information products in order to keep prices at a level similar to pre-digital periods is not necessarily the best policy direction (Maxwell 2004:13).

Maxwell's model assumes that authors receive royalties from their publications, which may not characterize most academic publishing arrangements. Academic authors are rarely concerned with royalties, but are typically interested in reputation and dissemination of their work since this is the means for professional credentialing. It would be interesting to see Maxwell revise this model to account for professional capital as an incentive to authors. I suspect that this revision

would amplify Maxwell's recommendation for weaker copy control over digital information products.

After wading through the technical legal debates over the impact of copyright on innovation and technological change, Verba's words, "there is no scholarship without communication," remind us of the deeper philosophical concern regarding the endangerment of the public domain and the circumscription of knowledge. "Restricting access to public domain material may eventually close off the raw material from which new works can be created as intellectual resources, whilst abundant, are not infinite." (O'Sullivan 2008:12)

THE PROMISE OF OPEN ACCESS AND THE CREATIVE COMMONS

Creative Commons was founded by professor in law and lawyer Lawrence Lessig to "hold back the top-down spread of copyright through the courts and the legislature" (O'Sullivan 2008:15). In other words, Creative Commons is a means to work within the system of copyright to provide less restrictive options for licensing and reassert some balance between the interests of publishers, authors, and the public by negotiating an intermediate course between privatization and complete abandonment of ownership. Lessig (2004) characterizes these license choices as voluntary steps to rebuild the public domain.

Creative Commons licensing has been popular with educators and researchers who want their work accessible, yet wish to ensure proper attribution. However, it does not fully account for the distinct nature of digital assets from print. Digital assets are, "...nonrival in consumption as a general rule. This means that one person's use of a digital asset does not impede or detract from another person's use of the same" (Blue Ribbon Task Force on Sustainable Digital Preservation and Access 2010:26). This is different from print items, which are *rival* in

consumption since they cannot be used by multiple people at the same time. The conditions of nonrivalry in consumption and non-excludability generally lead to market failure to supply even though the goods may benefit many consumers. Thus, public goods are usually supplied through non-market mechanisms such as government agencies (Blue Ribbon Task Force on Sustainable Digital Preservation and Access 2010:92-93).

Understanding the fundamental differences between print and digital assets has broad implications for the way we view intellectual property, the products of our scholarship, and for maintaining knowledge as a commons. While it is appealing to embark on a brave new course of open access in scholarly communication, we should remember that we conduct our scholarly activities in the shadow of a heavily capitalized publishing industry and it may be wise to consider market forces carefully. As Waters has astutely observed:

It is easy to imagine—especially in the absence of hard-nosed and aggressive strategic planning by, and collective action among, scholars, libraries, information technologists, and their universities—that the large, heavily capitalized publishing and other media firms will simply exploit open access repositories, cherry-pick the most valuable open access products, combine them with the most valuable new databases and resources, and sell them back to the academy at a significant profit, thereby chasing out sources of capital from within the academic community that are desperately needed to advance scientific, humanistic, and social science study (Waters 2008:3).

These sentiments regarding the imbalanced relationship between authors, academic institutions, and publishers are echoed in Kelty's response to the engagement of the American Anthropological Association in a five-year contract with Wiley, "...for a multinational corporation with shareholders and an enormous profit margin. I might now ask: why am I doing it [peer review] for free, for Wiley-Blackwell? Why isn't Wiley-Blackwell paying me to peer review?" (Kelty 2008:10). These observations may be taken as a fair warning to make an active investment in the future of our scholarly products—their preservation, ownership, and access.

WHAT DO WE DO ABOUT OUR DATA?

The thesis of this paper is simple. It is time to organize ourselves as a discipline and take a critical look at the products of our scholarship—all of them. The way in which data is collected, curated, and accessed varies considerably among archaeological projects. These methods are characterized as either centralized, distributed information gathering, or some combination of the two. The centralized end of the spectrum includes project databases that are designed, populated, and maintained by a core group of individuals who fulfill this role throughout the life of the project. At the other extreme, the growth of the database happens by a fluid process where researchers of varying ranks, experience, and institutional affiliations may come and go. During their tenure on the project, each individual adds information within established parameters to a predefined structure. Presumably, the structure of the database, along with data collection protocols, offers a measure of quality control.

Certainly, the size of the research project and its duration has a substantial impact on whether the information gathering is centralized or distributed. I would also venture that with the proliferation of digital data collection methods, even smaller projects begin to resemble the distributed type of information gathering. Large archaeological projects that run for several years may include contributions of specialists from a variety of disciplines. The result is a complex collaborative database that may include geographic, botanical, ceramic, lithic, biological, hydrological data, and more. With the associated metadata and data collection narrative, the databases become a type of scholarly communication themselves.

These differing modes of data gathering can affect the lifecycle of the data. Figure 1 is a diagram of a simplified data lifecycle that shows periods of relative activity and quiescence.

Databases may have very active periods of data deposition, correction, analysis and may even undergo revision to the design as its function is reevaluated. These aspects of the data lifecycle result in dynamic products that can pose challenges for preservation beyond that of versioning. As multiple researchers interact with the database at any given stage, the tracking of attribution becomes an additional challenge.

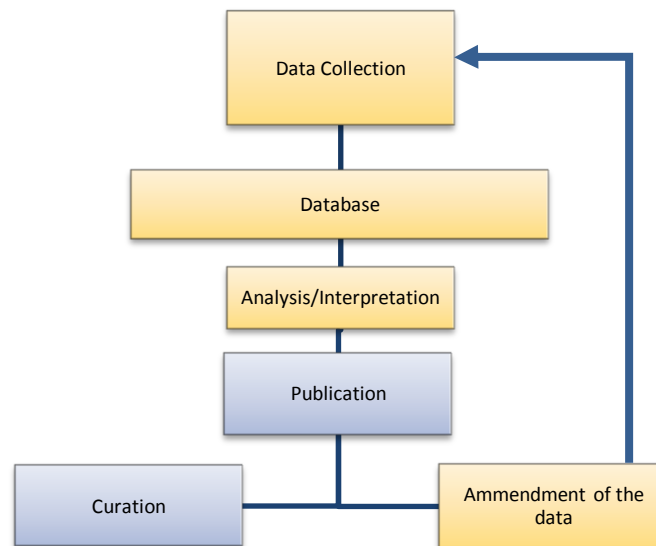


Figure 1. Diagram of a simplified data lifecycle. The orange boxes indicate periods of relative activity and greater interaction with researchers and the blue boxes represent periods of quiescence.

Although I am stating the obvious, I would like to make one other point regarding the nature of digital data collection. It is much easier to share data digitally than in print format. “...once created, digital materials can be shared at essentially no incremental cost, creating significant economies of scale” (Blue Ribbon Task Force on Sustainable Digital Preservation and Access 2010:27). I think it would be difficult to find someone who would disagree with the

statement, but I add that it is not necessarily easier to share digital data well. In the absence of metadata and other standards, data will quickly lose their context and even a well-planned and carefully constructed set of data may be worthless.

Provenance and authenticity of the information needs to be established; rights cleared, and databases and database schemas created; textual objects may need to be translated and marked up for grammatical and structural features as well as semantically according to certain knowledge structures... (*Waters 2008:3*).

Although Waters is referring to the study of texts, the production of archaeological databases would benefit from similar planning and standards development. The variety of data collection methods and the complexity of many archaeological databases pose significant challenges for the preservation of these data.

As a rule, the more complex the file format, the more effort must go into preserving it. Having well-documented file formats is essential for the future intelligibility of data, so metadata and format documentation are integral elements in preservation. (Blue Ribbon Task Force on Sustainable Digital Preservation and Access 2010:26)

Archaeology is a data intensive field in which individual scholars primarily bear the responsibility for organizing and preserving their data. Although there are some benefits to a distributed system (e.g., autonomy, flexibility), the costs of data loss, inconsistency of preservation methods, and metadata creation are much greater. By organizing ourselves as a discipline rather than waiting until our institutions have done it for us, we can promote consistency in archaeological data across institutions and ensure that developing standards address the practicalities of archaeological data.

WHOSE RESPONSIBILITY IS IT?

The collection of the scholarly record has long been the province of libraries. However, aside from archives of unpublished scholarly papers, libraries have traditionally collected the

formal products of scholarship (e.g., books, journal articles). As the practice of universities establishing and maintaining open access repositories through their libraries spreads, it may be the natural conclusion that libraries also establish data repositories. Governmental organizations, disciplinary societies, or other academic organizations could also fill this role (Waters 2007). The Data Conservancy, led by Sayeed Choudhury at the Johns Hopkins University Sheridan Libraries, is one example of such an initiative (Lee, et al. 2009). As part of NSF's DataNet program, the Data Conservancy will receive \$20 million over five years to build infrastructure for the management of digital research data.

Depositing research data in repositories for the purposes of preservation and dissemination requires the development of rigorous metadata standards to ensure adequate documentation of collection methodologies, data provenience, etc. In addition to improving our ability to share data and preserve it over time, the development of standards could have broader disciplinary impacts such as improving data quality through peer review or other vetting processes. There are already examples of such changes taking hold in other fields. In the United Kingdom, JISC is supporting the development of a data journal where the peer review process is built around the data. This type of journal is no replacement for the traditional interpretive research article, but it highlights new opportunities afforded by digital technologies and suggests how we might better take advantage of them. Cognizant of the developments that are already taking place in the IT and library world, by actively engaging with those who will build the cyberinfrastructure we can promote the development of standards that address our needs as archaeological researchers.

REFERENCES CITED

- Blue Ribbon Task Force on Sustainable Digital Preservation and Access
2010 *Sustainable economics for a digital planet: ensuring long-term access to digital information.*
- Boyle, J.
1996 *Shamans, software, and spleens: law and the construction of the information society.* Harvard University Press, Cambridge, MA.

2007 Mertonianism unbound? Imagining free, decentralized access to most cultural and scientific material. In *Understanding Knowledge as a Commons*, edited by C. Hess and E. Ostrom, pp. 123-143. The MIT Press, Cambridge, MA.
- Ewing, J.
2003 Copyright and authors. *First Monday* 8(10):1-8.
- Hess, C. and E. Ostrom
2003 Ideas, artifacts, and facilities: Information as a common-pool resource. *Law and Contemporary Problems* 66:111-145.
- Kelty, C.
2008 The state of open access anthropology. *Anthropology News* (February):9-10.
- Lee, J. W., J. Zhang, A. S. Zimmerman and A. Lucia
2009 DataNet: An emerging cyberinfrastructure for sharing, reusing and preserving digital data for scientific discovery and learning. *AIChE Journal* 55(11):2757-2764.
- Lessig, L.
2001 *The future of ideas.* Random House, New York.

2002 The architecture of innovation. *Duke Law Journal* 51(6):1783-1801.

2004 *Free culture.* Penguin, New York.

2005 The people own ideas! *Technology Review* 108(6):46-53.
- Maxwell, T.
2004 Is copyright necessary?
- Nguyen, T.
2008 *Open doors and open minds: what faculty authors can do to ensure open access to their work through their institution.* SPARC/Science Commons.

O'Sullivan, M.

2008 Creative Commons and contemporary copyright: A fitting shoe or "a load of old cobblers"?

Suber, P.

2007 Creating an Intellectual Commons through Open Access. In *Understanding Knowledge as a Commons*, edited by C. Hess and E. Ostrom, pp. 171-208. The MIT Press, Cambridge, MA.

Waters, D.

2007 Preserving the Knowledge Commons. In *Understanding Knowledge as a Commons*, edited by C. Hess and E. Ostrom, pp. 145-167. The MIT Press, Cambridge, MA.

2008 Open access publishing and the emerging infrastructure for 21st-century scholarship. *JEP: the journal of electronic publishing* 11(1).