

April 16, 2006

Mary Rasenberger
Policy Advisor for Special Programs
U.S. Copyright Office

Ms. Rasenberger,

We wish to thank the US Copyright Office, the Library of Congress, and the Section 108 Study Group for holding an enquiry into the provisions of Section 108 of the Copyright Act, and opening these to interested parties.

Please accept the enclosed written comments to the 108 Study Group respectfully submitted by the following individuals on behalf of the International Research on Permanent Authentic Records in Electronic Systems (InterPARES): Luke Meagher (University of British Columbia), Mahnaz Ghaznavi (Getty Research Institute, J. Paul Getty Trust), Howard Besser (New York University), Terry Maxwell (State University of New York, Albany), and Jim Suderman (City of Toronto), who chairs the InterPARES 2 Policy Cross-Domain team, under the guidance of Luciana Duranti, Univesity of British Columbia, Director of the InterPARES Project and Philip Eppard, Professor of Archives and Records Management, State University of New York, Albany, principal investigator of the U.S. InterPARES Team.

InterPARES is a large multi-year multi-national research project of the archival world examining the long-term preservation of authentic records created and/or maintained in digital form. Since 2002, InterPARES 2 has been examining issues of authenticity, reliability and accuracy in the preservation of digital records, focusing on archival records produced in complex digital environments in the course of artistic, scientific and e-government activities. Through our research, InterPARES has identified issues and impediments to long-term preservation of digital records. A number of these issues and impediments lay in the policy arena (including both copyright and privacy issues), and we will be releasing our full report on those before the end of 2006. Because Section 108 provisions are directly related to our research findings, we are providing the Section 108 Committee with a look at some of our relevant findings in advance of the release of our final report later this year. For more information on InterPARES, please see our website at http://www.interpares.org/ip2/ip2_index.cfm (where our final report will be posted as well).

Preservation, a function typically, but not exclusively, undertaken by libraries and/or archives, plays an important social role by maintaining and providing access to records over time. In the analog world, the costs associated with preservation have usually not been borne by records creators or rights holders, but rather by public or non-profit organizations, even when the primary beneficiary (in the monetary sense) has been the rights holder. As discussed in the enclosed comments, preservation of digital records is a resource intensive activity, and it is not clear who is to bear the full costs of preservation. At the same time, our research indicates that electronic records are at risk from the

moment they are created – a point examined at length below. We therefore urge the Study Group to bear in mind that any changes to Section 108 exemptions should take into account these very real social and financial costs involved with preservation in the digital world.

Respectfully yours,

Mahnaz Ghaznavi
Records Manager
Institutional Records and Archive, Getty Research Library
J. Paul Getty Trust

Topic 1: Eligibility for Section 108 Exceptions

The Study Group asks whether additional criteria for eligibility for this exemption should be added to the current text of the statute, and whether further definition of the terms “library” and “archives” is required. In addition, the Study Group wonders whether or not “virtual” libraries and/or archives and museum should be included in the law. Historically, neither “libraries” nor “archives” have been defined by Section 108; rather, questions of whether exceptions apply have had to do with the extent to which copying is undertaken for the sole purpose of commercial gain at the expense of rights holders.¹ And it is precisely activities and functions, rather than physical or organizational entities, that should be considered when exemptions and indemnification are analyzed under section 108, especially given the more recent profound changes in the ways in which organizations (both commercial and non-profit) conduct business (changes that have prompted the Study Group to enquire about, say, outsourcing and various other novel production models).

Focused on records’ creation, maintenance, and preservation in the domains of the arts, sciences, and e-government, rather than on “libraries” and “archives” as such, InterPARES case study research findings suggest that digital records are at risk from the moment that these are created (more on this found below in comments addressing points 2 and 3). What a repository chooses to call itself -- whether “archives,” “library,” “museum,” or something else entirely -- should have no bearing on its eligibility for exemption under section 108(a), because it is the function of preservation, and the organization’s policies and procedures in support of that function that should be considered when framing the issues.² With this in mind, InterPARES would suggest that the term “records preserver” could replace the phrase “library or archives” throughout the section.³

The preservation policy, strategy or standard of an entity calling itself a records preserver *must* explicitly state its role as what InterPARES calls a “trusted custodian.” To be considered as a trusted custodian, the preserver must:

- a) act as a neutral third party, i.e., demonstrate that it has no stake in the content of the records and no reasons to alter records under its custody, and that it will not

¹ Mary Rasenberger and Chris Weston, *Overview of the Libraries and Archives Exception in the Copyright Act: Background, History, and Meaning*, 24, available at <http://www.loc.gov/section108/papers.html> (accessed April 5, 2005).

² Any organization without such a policy could not be considered to be eligible for exemption under this section. Lack of a preservation policy reasonably calls into question an organization’s trustworthiness as a preserver, the authenticity of the records it may claim to preserve, and, from a copyright perspective, whether or not the organization respects the moral and economic rights of creators.

³ “The records preserver is the entity responsible for acquiring (i.e., taking physical and legal custody) and preserving (i.e., protecting and ensuring continuous access to) records.” InterPARES 2, *Principles for Records Preservers (draft)*, Principle #1 (forthcoming)

allow anybody to alter the records either accidentally or on purpose,

b) be equipped with knowledge and techniques necessary to fulfill its responsibilities, which should be acquired through formal education in the field of archival and/or records management with professional practices, and to

c) establish a trusted preservation system that is capable of ensuring that reliable, accurate and authentic copies of the creator's records are acquired and preserved.

The Study Group will note that attribute "a)" of the trusted custodian is similar to the Section 108(a)(1)⁴ criteria for exemption. It should also be noted, though, that InterPARES's attribute is drawn more broadly, in that it states that the preserver have "no stake" in the records' content and "no reasons" – commercial or otherwise – to alter records under its custody.

The underlying rationale of InterPARES's concept of the trusted custodian is the protection of electronic records' authenticity in their preservation. Likewise, the protection of copyright holders' rights is also concerned with maintaining authenticity: the moral rights bestowed upon the creator of a work are meant to provide a safeguard against the distortion of that work by unauthorized parties. The protection of an author's moral rights are provided for in attributes "a)" and "c)" of the trusted custodian.

The trusted custodian concept also provides safeguards against the violation of a creator's economic rights, as seen above in attribute "a)," but also through attribute "b)." This characteristic of the trusted custodian requires that "knowledge and techniques" of digital preservation "be acquired through formal education in the field of archival and/or records management." Such an education would necessarily entail an exposure to the legal and ethical issues at play in records and archives management, especially those associated with copyright law.

By placing certain activities outside the normal constraints of copyright law, Section 108 exemptions aim to balance the economic interests of rights holders and social benefits of preservation and access.⁵ The trusted custodian's procedures should be driven by explicit policies on the management of records that seek to balance the interests of various stakeholders. However, trusted custodians and their institutions should also be held harmless from penalties emanating for infringement of laws in cases where they are:

- a) following explicit procedures laid out by their institutions; and
- b) attempting to balance competing values imbricated in organizational and legal mandates.

⁴ "[I]t is not an infringement of copyright for a library or archives [...] to reproduce no more than one copy or phonorecord of a work [...] if [...] the reproduction or distribution is made without any purpose of direct or indirect commercial advantage[.]"

⁵ Rasenberger and Weston, 1.

The implementation of policies should include procedures for

- a) clearing records of legal constraints at the point of accession; and
- b) locating and negotiating in good faith with rights holders and affected parties where prior clearance has not been acquired.

In cases where the rights holder is neither identifiable nor able to be located after reasonable efforts have been undertaken, or where the records have no current market value (that is, they are no longer for sale in normal retail establishments for reasonable market prices), the trusted custodian should be allowed to use their best professional judgment in determining electronic access.

Topic 2: Amendments To Current Subsections 108(b) and (c)

InterPARES research findings suggest that the preservation of digital records requires maintenance of the records throughout their lifecycle. In such digital preservation techniques as migration and emulation, the preserver must continuously make copies to make electronic works viewable over time. This work is particularly difficult and particularly necessary with works that are interactive, dynamic, and experiential, such as those that have been studied in InterPARES's case studies.

Through its research, InterPARES has determined that the proper preservation of digital records requires that multiple copies of the records must be periodically generated. In order to support the presumption of authenticity, preservers need to be able to refer back to previous representations of a work, all the way back to the authentic original record or its authentic copy, in order to verify whether or not the various migrations over time have altered the work. This means that we cannot destroy previous transformational copies when we make a new one, as they are points of reference. In the field of digital preservation, any limitation on the number of copies a preserver is allowed to make is artificial and becomes a barrier to the faithful preservation of a record.

InterPARES research indicates that preservation activity in some cases entails incidental capture (such as copies of web sites or records from other dynamic systems); to the extent that this activity is engaged in for the purposes of preservation, rather than to compromise the economic interests of the rights holder, records preservers must be indemnified under exemptions recommended in topic 1, above, including for copies that include content with underlying rights that relied upon legal principles like "fair use."

Questions of access, such as the ones posed by the Study Group, are not solely contingent on rights management, but also related to privacy legislation and technological realities. Access must be balanced not only against the economic interests of the rights holders, but also state and federal privacy and, in some cases, freedom of information laws intended to protect the interests of the individual and/or the public. As well, questions of access must contend with the technical infrastructure that enables access to content; in the digital environment, access itself is an implementation of code that is intellectual property, and rights issues are not exclusively bound up with content and its creators but also with our

ability to experience that content (that is, the code or software that enables us to view material online too engages copyright issues).⁶

InterPARES research suggests that in the absence of appropriate preservation of digital records, the question of access is often moot. In order to provide access, the records preserver will need to have the ability (both technical and legal) to reproduce the record. For instance, if permission is granted for “off-site” premises to copies of files, and yet the user does not have the appropriate hardware and software to render the file eye-readable and functional, then the user will not be able to reproduce the record. Hence, having the permission to provide the content is not the sole issue. Access privileges to preservation copies should exclusively apply to the trusted custodian. Preservation copies would support the work of the preserver, who would create authentic copies for access from them.

Topic 3: New Preservation-Only Exception

Within the questions posed by the Study Group in Topic 3, there are two concepts on which InterPARES is compelled to testify. First, the issue of identifying “at-risk” records is one which InterPARES can comment upon citing its empirical study of the problem. Secondly, and in relation to this first issue, the Study Group poses the question of how to determine which “trusted repositories” may take advantage of the proposed exception. InterPARES has crafted criteria for institutions to determine whether or not a repository is trustworthy – an organization can be considered a “trusted custodian” if it fulfills the criteria noted above (in Topic 1).

It is a foundational finding of InterPARES that *all* electronic records are by their very nature “at-risk,” that the preservation of such records must begin before or at the records’ creation. In empirical research involving dozens of multi-year case-studies, InterPARES has concluded that preserving a digital work and guaranteeing that the preservation copy is accurate, reliable, and authentic is a hugely complex process (that requires involvement of a preservation archivist throughout the life-cycle of a work).

The techniques for preserving electronic records are certainly different than those applied to the preservation of traditional records. More importantly, though, the preservation of electronic records requires a shift in thinking about when the preservation of the records begins as InterPARES has found that the preservation of digital records must take place at a different time from that of traditional records.

In the case of the more durable records of the past – such as those created on paper and some other analog materials -- records are appraised for preservation at the disposition stage, when they are no longer needed for business purposes. This preservation paradigm was identified by InterPARES as the *Chain of Custody* – the authenticity of the records could be presumed simply by the directness of the creator’s transfer of records to the preserver. With digital records, however, decisions regarding preservation must be made

⁶ See for instance, the works of Lawrence Lessig and Karen Coyle.

at or even before the record creation stage due to the ease with which they can be modified, manipulated or deleted in the creator's recordmaking system and/or that they may become inaccessible caused by the technological obsolescence in the creator's recordkeeping system. This new paradigm, referred to within InterPARES as the *Chain of Preservation*, is confirmed by but not unique to InterPARES research. It is a common finding of many research initiatives to date. Only by appraisal at the time of creation, followed by careful stewardship and monitoring, can the custodian ensure that the record to be preserved can be presumed authentic. In fact, if this course is followed, it is not simply a presumption of authenticity, but presumably that authenticity of the preserved records can be demonstrated; these points are further discussed in the following sections

Section 108(c) reads, in part: "The right of reproduction under this section applies [...] if the existing format in which the work is stored has become obsolete." The preservation of digital records is a constant warding-off of obsolescence -- an exercise in acting in the digital present and looking to the horizon of the digital future. Waiting until a record's format is obsolete virtually guarantees that its authenticity is diminished and/or compromised, and may in fact result in the complete loss of that record. Such a strategy finds the preserver in the unfortunate and constantly flat-footed position of perpetually responding to the crisis of the digital present as it is overtaken by the digital future. This is not only inefficient -- it puts the preservation of digital materials in peril and promises to be a costly affair.

The interactive and dynamic computer systems employed in organizations complicate the preservation situation in that such systems normally do not generate records with fixed form and stable content that are necessary conditions enabling preservation. This is the origin of the new notion that records preservation starts at the creation stage. This notion requires that preservation considerations be incorporated and manifested in the design of recordmaking and keeping systems. In addition, records should be monitored throughout their lifecycle so that appraisal decisions and preservation considerations can be updated and/or modified to accommodate any possible changes occurring after they are first made.⁷

In regard to the question posed by the Study Group concerning which libraries or archives should be allowed to preserve digital records, whether through a certification process or by self-certifying, InterPARES has several comments.

The "Digital Library Certification" project of the Research Libraries Group/National Archives (RLG/NARA) is an important step in identifying what "good practice" should be for a preservation repository. But what is "good practice" for a digital repository will continue to be a moving target for a number of years, and we cannot etch requirements for certification in stone or, indeed, into law. The certification requirements that result from this process will be important for large repositories handling general, wide bodies of content. However, there will always be differing needs for repositories that seek to preserve more limited, niche areas of content. These repositories may accession elements of their collection according to subject (electronic literature or old weather reports or

⁷ InterPARES 2, Principles for Records Creators. From explanation under Principle #7.

stories about Bears) or format (blogs or pop-up ads), or other criteria. Some of these repositories should ideally face more stringent certification than the RLG/NARA project will formulate, while others should ideally face more relaxed certification.

But a “one-size-fits-all” certification process would be a mistake. It is the InterPARES position that a preservation repository be a “trusted custodian” rather than it being certified as such. InterPARES believes that self-qualification in this regard is certainly better than strict adherence to a certification process, but does concede that both options - self-certification and certification by an external body – have their respective and particular problems. True, but would it be fair to say that self-qualification is superior because it places an educational responsibility on the preserver, which is better able to keep pace with changing technologies?

Regarding the issue of dark archives, InterPARES holds that the creation of dark archives in order to simultaneously protect rightsholders and preserve the records is generally a bad idea, due to the complexity of the archives’ management. In a dark archives, when the preserver discovers a problem with a record in the archives, it may be too late to correct the problem. The viewing or rendering software for a particular record may be unavailable or may not work on a current operating system, links in a webpage may be corrupted, or the files themselves may have lost sub-pieces. A recent study by the Center for Research Libraries study showed that 40% of materials requested from a “dark archives” had become inaccessible while the archives remained dark, unused. For a dark archives to work, the system must have some kind of function to alert the digital archivist of failures within the system, and the archivist must be notified while they can still correct the failures. Until such a reliable function is developed and successfully implemented, there will need to be at least some limited access to semi-dark archives as a means to identify problems.

Topic 4.

From the InterPARES experience of examining dynamic, experiential, and interactive records, websites do not themselves pose hugely different problems than the other types of records we have studied. Website preservation has the same requirements as noted above: websites must be managed throughout their lifecycle as records. Also, point-in-time copies must necessarily be made to properly document the history of a webpage.

Websites are technological constructs and as such it cannot be assumed that authorship resides in a single individual or that it is necessarily possible to manage separately the intellectual property rights of the individuals participating in the creation of the website. For example, rights adhering to photographs used in a website belong to the photographer(s) that created them, while the rights adhering to the text belong to the author. Where the website belongs to an institution or a single individual, it may be possible to consider them as entities to which intellectual property rights can be viably established and maintained. Where websites are created by collective or collaborative efforts, this would be difficult or impossible.

While Section 108 should enable trusted custodians to exercise professional judgement in access and copying decisions that support preservation, as noted under topic 1, above, it should also acknowledge that, in the case of digital records, trusted custodians will need to make key decisions about preservation at much earlier parts of a records' life, including the collection stage.

It is with this in mind that InterPARES reiterates the principles it has outlined above. Records must be managed throughout their lifecycle, and any limitation on the number of copies a preserver is allowed to make is artificial and becomes a barrier to the preservation of a record. In addition, InterPARES once again states its position that if the policies and functions of an organization show that it is a trusted custodian, such an organization should be allowed to carry on its good-faith preservation activity.