

Keeping the Promises Associated with Your Digital Assets: Is it Possible?*

Marie E. Gallagher; Lister Hill National Center for Biomedical Communications, National Library of Medicine; Bethesda, Maryland, USA

Abstract

The Profiles in Science® Web site [1] features digital reproductions of historical papers, photographs, audios and videos selected from the archival collections of the National Library of Medicine as well as collaborating institutions. The Web site's digital items are derivatives of high quality digital masters residing in a repository designed for long term survival. The need for significant infrastructure to guarantee the persistence of these digital files was expected, and our desire to ensure their persistence influenced our choices and system design decisions from the start. Choosing file formats carefully, making certain that our digital assets were not lost due to media degradation or obsolescence, and planning for maximum flexibility in order to take advantage of rapidly changing technology were among the strategies we employed from the beginning [2].

However, the items selected for Profiles in Science include both public domain materials and materials restricted for copyright or other reasons. The need for additional infrastructure to keep copyright and restriction-related promises associated with individual digital assets became apparent. Situations that called for sharing items from our repository required different strategies for different cases depending on the recipients and their intended use of the items. In all cases, we decided to prominently associate the property rights information with the digital artifacts in order to protect all stakeholders. Rights information embedded in digital items can be overlooked or ignored. Rights information stored separately from digital items can be forgotten or lost. In this paper we share our experiences and examine linking property rights information with digital items.

Background

The title of this paper suggests two challenges: 1) living up to the implicit and explicit promises underlying our digital assets, and 2) keeping those promises from becoming separated from our digital assets. The former depends on the latter. Should the promises become separated from our digital assets, the promises would become lost, forgotten or ignored in the future. In order to live up to the promises, we need to take steps to ensure that these promises are not lost because of our actions or inactions.

Library and archives patrons expect that paper-based materials will not substantively change or disappear, and they may expect the same for digital materials. Implicit promises are made whenever digital information is made available for the long term [3]. Long term availability requires ensuring the integrity and persistence of digital items. Methods for keeping these types of promises have been and continue to be actively discussed, developed and implemented by the digital preservation and library science communities and others [4, 5, 6, 7, 8, 9].

Individuals or institutions who grant permission to post items on Profiles in Science expect us to respect copyright laws. Our project's copyright researcher determines whether items are subject to copyright, identifies copyright holders, requests permission to digitize and post their materials on Profiles in Science, and obtains written permission from the copyright holder. At the conclusion of copyright research, each item is assigned to one of the seven categories listed in Table 1. Only items in the top four categories are posted on Profiles in Science [10]. The copyright status of the Profiles in Science items is displayed in the metadata record associated with each digital item. Respecting copyright laws, rules and agreements associated with the Profiles in Science digital items is one of the explicit promises we make.

Table 1: Profiles in Science rights categories and the number of items assigned to each category as of December 2006

Public domain	10,339
Permission granted unconditionally	10,820
Permission granted: conditions acceptable	1,553
Rights owner indeterminable: post with disclaimer	2,114
Rights owner indeterminable: do not post	79
Permission granted: conditions unacceptable	221
Permission denied	282

In the footer of each page of the Profiles in Science Web site is a link to the National Library of Medicine notice [11] alerting the public that they may encounter items protected by copyright laws and that they must request permission from the copyright owner to transmit or reproduce these items beyond "fair use." The Profiles in Science site also displays specific copyright information in the rights section of each metadata page associated with each item subject to copyright. Based on the requests for information that we receive about contacting copyright owners to republish items, these notices and the rights metadata are being seen and are working as intended. The strategy we are using on the Web site has been effective in helping us keep our promises to respect copyright and in helping others to do the same.

Our project has been asked to make copies of the Profiles in Science digital repository to share with collaborating researchers and institutions on several occasions. Aside from the copyright issues, providing a copy of all of the digital files would be technically easy. This could also put us and our collaborators at risk. We could violate copyright by retransmitting protected items beyond the uses allowed by the previously obtained copyright agreements. Or, our collaborators could unwittingly violate copyright because the copyright issues that are visible on the Profiles in Science Web site would not be obvious to a collaborator examining a disk full of files.

* In: Archiving 2007 Final Program and Proceedings

The role of metadata

We record descriptive, administrative and technical metadata for all of the Profiles in Science digital items. We store these metadata in a database, and link the metadata to their corresponding digital items through a unique identifier. We use the metadata for a variety of purposes, including generating the Web pages seen on Profiles in Science [12]. Rights-related metadata are part of these administrative metadata. The promises we make regarding copyright are recorded in the metadata records, rather than embedded in the digital files. We rely on the metadata now and in the future to help us keep those promises. While working within our project's software or browsing the Profiles in Science Web site, the metadata are as visible as the digital items, so the promises associated with individual items are hard to ignore. But when the metadata and digital items become separated, as they could when provided to researchers and collaborators out of context, the fact that promises have been made is lost.

There are tradeoffs associated with either storing metadata externally or embedding them within items. An RLG survey suggests that, in the case of technical metadata, users are divided about whether it is best to embed the metadata, keep them external, or do both: embed the metadata in the associated digital files and keep an external copy of the metadata [13]. These three possibilities and their advantages and disadvantages apply to the situation with our rights metadata.

Table 2: Advantages and disadvantages of embedding metadata in digital items

Method	Advantages	Disadvantages
Embedded metadata	Metadata and item are a single package	Migration to other file formats may exclude the metadata
External metadata	Metadata is readily available for computation and updates	Metadata and item may become separated
Embedded + external metadata	Reduces risk of separation of metadata from item, and retains ability to perform computations	Synchronization issues between the embedded and external metadata

Another challenge regarding our rights metadata (as well as other types of administrative and descriptive metadata) is that the metadata are occasionally subject to change. Although the digital items in our repository do not change over time, sometimes we discover new information about them. For example, copyright ownership may be transferred from one individual or organization to another. Or the identity or whereabouts of a previously unknown photographer may become known. Changing metadata is one of the reasons we choose to keep our metadata external to our digital items. This also means that when we provide our collaborators copies of the metadata associated with the digital images, we make sure to note that the metadata are current only as of the date we provide them.

Protecting our collaborators

Several times we have been asked to provide copies of the Profiles in Science digital items to different groups for different purposes. So far, we have used two different methods for doing so, based on the needs of the group. In order to protect our collaborators and ourselves, we either make the copyright status of the digitized items very prominent, or we share only public domain items.

Case 1: collaborating with researchers

Collaborating researchers have asked us to provide copies of the Profiles in Science master digital files (mostly TIFF files) for their experimentation. Only the files, not the metadata, were of interest to them. Without the metadata, the researcher might not be aware that copyright issues were associated with the files they were using. This could lead to the researcher inadvertently violating copyright by republishing protected images in a paper or on a Web site, or distributing them as part of a test set. In order to enable researchers to use the images, and avoid putting our collaborators at unnecessary risk, we exclude copyrighted items and provide only public domain items for research and experimentation (see Figure 1).

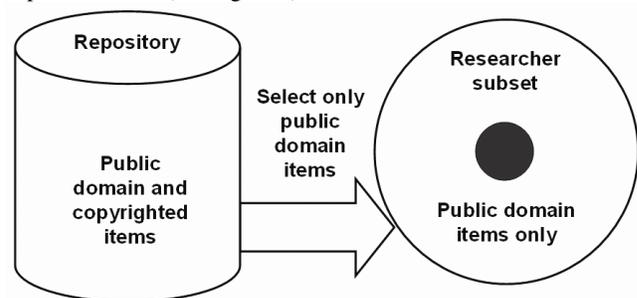


Figure 1. Researcher subset contains only public domain items

Case 2: collaborating with institutions

On a few occasions collaborating institutions have loaned us items for inclusion on Profiles in Science and asked us to provide them copies of the digital reproductions and metadata we create for Profiles in Science. In all cases, this arrangement was known before digitization or copyright research commenced, so permissions were secured on behalf of Profiles in Science as well as our collaborating institution. We knew that the person receiving the digitized files and metadata understood copyright law as well as the rights issues associated with the individual items in their collection. However, in the future, the disks could be passed along to students, volunteers, Web masters, information technology staff, or other collaborating researchers who might not be so familiar with copyright issues and might unwittingly violate them. In our attempt to protect all stakeholders, we sought a solution that would prominently display, like a warning label, the copyright status of the items so that this information could not be easily missed by anyone handling the disks we provided. Currently our "warning label" solution is to segregate the files into separate directories (folders) named according to the permission status of the items in that directory. A person accessing our disks of files and metadata must traverse this short list of directory names in order to reach the digitized files. Hopefully this provides sufficient information "at a

glance" to alert the user that the digital items on the disk should not be freely reproduced or distributed. The user also sees immediately which items are in the public domain and freely available for use. For files of interest that are not in the public domain, more detailed copyright information is available in the metadata records corresponding to those files. Figure 2 shows sample directory names corresponding to the categories listed in Table 1. If files are digitized that are not to be made available to the public under any circumstances, they could be further segregated by storing them on a different disk from public items.

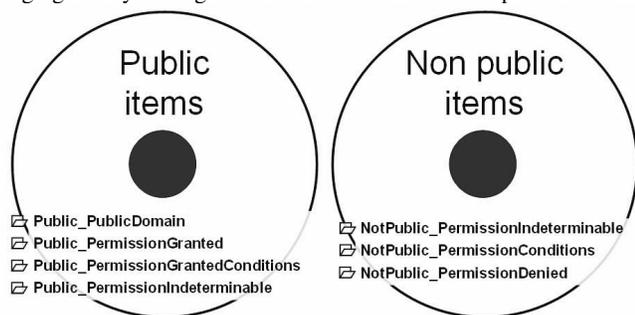


Figure 2. Directory names can increase visibility of basic copyright information

Summary

It is impossible to guarantee that others will keep the promises we have made, but it is possible to take actions to increase the chances that they will. Keeping the promises underlying our digital assets seems a minimum requirement for their long term survival. A key factor in keeping our promises is ensuring that metadata remain associated with their corresponding digital items. Although we consider all of our metadata important and useful, there could be a greater cost for overlooking our rights metadata than other types of metadata. We have considered our own project's needs and found that, so far, rights metadata are the only category of metadata that warrant "warning label" status. That is, in order to keep our promises, rights metadata need to be highlighted whenever we share our digital items. Not keeping these promises could endanger our project's ability to secure permission from copyright holders in the future, and could damage the reputations of our institution and collaborating institutions. Others may find that other types of metadata need to be highlighted in order to keep their promises.

We considered the advantages and disadvantages of embedding metadata in digital items vs. storing metadata externally. For now, we will continue to store our metadata separately from their digital items. We will need to continue to find methods for keeping the metadata and digital items from becoming separated permanently.

We outlined two relatively inexpensive technical solutions that we have implemented to tide us over until permanent comprehensive solutions are developed. One solution is to avoid risk by providing only public domain images in cases where rights issues are likely to be ignored or forgotten. The other solution requires the user to navigate through a very simple form of our rights metadata before he or she reaches the digital items. Although our solutions cannot guarantee that our promises will be

honored, they may raise awareness of the fact that promises exist, and may increase the chances that these promises will not be ignored.

Acknowledgments

The author thanks her dedicated and creative Profiles in Science colleagues at the National Library of Medicine: Meghan Attalla, Michael Flannick, Erica Haakensen, Walter Hickel, Xiaohui Ma, Gregory Pike, Edwin Rivera, Susan Speaker, Elizabeth Suckow, Michele Tourney and Karl Wolf. Thanks especially to Digital Manuscripts Program manager Christie Moffatt and copyright researcher Sandra Brauksieck Taylor for consultations and review of this paper. Essential encouragement and guidance has been provided by project leaders Elizabeth Fee, Donald King, Clement McDonald, Aaron Navarro and Paul Theerman, as well as by Alexa McCray who led the Profiles in Science program from its inception through the year 2004. Thanks to the collaborators and donors of the Profiles in Science collections for their willing and active participation.

This research is supported by the Intramural Research Program of the NIH, National Library of Medicine. The Library, the world's largest library of the health sciences, is a component of the National Institutes of Health, U.S. Department of Health and Human Services.

References

- [1] Profiles in Science (National Library of Medicine, Bethesda, MD) <http://profiles.nlm.nih.gov/>. Accessed Feb 15, 2007.
- [2] Alexa T. McCray and Marie E. Gallagher, "Principles for Digital Library Development." *Communications of the ACM* 44, no 5 (May 2001).
- [3] Margaret M. Byrnes, "Permanence Levels and the Archives for NLM's Permanent Web Documents." *NLM Tech Bull.*, 343, e4 (Mar-Apr 2005). http://www.nlm.nih.gov/pubs/techbull/ma05/ma05_archive.html. Accessed Feb 15, 2007.
- [4] NISO Z39.87-2006 Data Dictionary--Technical Metadata for Digital Still Images (National Information Standards Organization, Bethesda, MD, December 18, 2006) <http://www.niso.org/standards/resources/Z39-87-2006.pdf>. Accessed Feb 15, 2007.
- [5] PREMIS: Preservation Metadata Maintenance Activity (Library of Congress, Washington, DC) <http://www.loc.gov/standards/premis/>. Accessed Feb 15, 2007.
- [6] OCLC/RLG Working Group on Preservation Metadata, *Preservation Metadata and the OAIS Information Model: A Metadata Framework to Support the Preservation of Digital Objects* (OCLC Online Computer Library, Inc., Dublin, OH, 2002) http://www.rlg.org/en/pdfs/pm_framework.pdf. Accessed Feb 15, 2007.
- [7] RLG/OCLC, *Trusted Digital Repositories: Attributes and Responsibilities* (RLG, Mountain View, CA, May 2002) <http://www.rlg.org/en/pdfs/repositories.pdf>. Accessed Feb 15, 2007.
- [8] *Digital Preservation: The National Digital Information Infrastructure and Preservation Program* (Library of Congress, Washington, DC) <http://www.digitalpreservation.gov/>. Accessed Feb 15, 2007.
- [9] *Digital Preservation* (Digital Library Federation, Washington, DC) <http://www.diglib.org/preserve.htm>. Accessed Feb 15, 2007.
- [10] Marie E. Gallagher and Christie Moffatt, *Surviving Change: The First Step toward Sustaining Your Digital Library*, in J. Trant and D. Bearman (eds.). *Museums and the Web 2006: Proceedings*, Toronto:

Archives & Museum Informatics, published March 1, 2006 at <http://www.archimuse.com/mw2006/papers/gallagher/gallagher.html>. Accessed Feb 15, 2007.

- [11] NLM Copyright Information (National Library of Medicine, Bethesda, MD) <http://www.nlm.nih.gov/copyright.html>. Accessed Feb 15, 2007.
- [12] Alexa T. McCray, Marie E. Gallagher, and Michael A. Flannick, Extending the Role of Metadata in a Digital Library System, in Proceedings of the IEEE Forum on Research and Technology Advances in Digital Libraries, pp. 190-199. (1999).
- [13] Automatic Exposure: Capturing Technical Metadata for Digital Still Images (RLG, Mountain View, CA, 2004)

http://www.rlg.org/en/pdfs/projects/automatic_exposure/ae_whitepaper_2003.pdf. Accessed Feb 19, 2007.

Author Biography

Marie E. Gallagher, a computer scientist in the National Library of Medicine's Lister Hill National Center for Biomedical Communications since 1990, is the project leader of the Digital Library Research and Development team. The team investigates systems and develops the software underlying Profiles in Science. Ms. Gallagher earned her B.S. degree in Computer Science and Mathematics from the College of William and Mary in Virginia.