

December 2012

What Not to Embed: is it better not to embed certain cultural heritage metadata in images?

Greg Reser

University of California - San Diego, greser@ucsd.edu

Follow this and additional works at: <https://online.vraweb.org/vrab>

Recommended Citation

Reser, Greg (2012) "What Not to Embed: is it better not to embed certain cultural heritage metadata in images?," *VRA Bulletin*:Vol. 39: Iss. 1, Article 5.

Available at: <https://online.vraweb.org/vrab/vol39/iss1/5>

This Feature Articles is brought to you for free and open access by VRA Online. It has been accepted for inclusion in VRA Bulletin by an authorized editor of VRA Online.

What Not to Embed: is it better not to embed certain cultural heritage metadata in images?

Abstract

While many people agree that embedding descriptive metadata in image files, there is a lot of disagreement about how much of it to embed, particularly when it comes to images of cultural heritage works. The "embed everything you know" camp see more value in having flawed metadata than no metadata. The "embed only metadata that never changes" camp see too much risk in passing along data that is apt to change due to scholarly review. This article joins the debate and presents arguments for joining the "embed everything you know" camp.

Keywords

Adobe XMP; embedded metadata; cataloging; digital asset management; IPTC

Embedding metadata in digital images is generally considered to be useful, but one question still generates debate: which metadata should be embedded? The question is not about the images' technical metadata (such as file creation date or camera settings), but rather about the metadata that describes the images' content, especially when the content is a work of art.

If you have rich sets of descriptive information about cultural heritage works, how much of it should you embed in surrogate digital images?

Digital images are created in many different environments with specific uses in mind. Embedded metadata should enhance those uses, not subvert them. This article addresses digital images that are shared with others - often times end users - not images that are kept within the confines of a single institution. In-house images that are actively managed by a single database or digital asset management system (DAMS) should have only minimal metadata that ties them to their master record. After all, authority control and data maintenance, is the main benefit of a centralized system. On the other hand, images that are widely distributed outside the institution need to carry more metadata with them because they live on their own as individual files in a user's local drive.

The point of embedding metadata in an image file is to make it and its content identifiable wherever the image may travel. If a user opens an image you sent to them and there is no metadata, they have a mystery to solve. For example, if a user has 1,000 mysterious images, they would need to do a lot of research before the images are truly useful. Embedded metadata would not only tell them what has been captured in those images, it would allow them to search and sort the image files in a system browser or a photo organization application. The usefulness of digital media files with metadata is widely appreciated: imagine using music files on your desktop or mobile device without track information!

What is the problem then? Why not embed everything you know about a work of art shown in an image and send it on its way? The problem stems from the fact that that cultural heritage works are subjects of scholarly examination and what is known about them evolves over time. This is, in fact, what makes these works so interesting - they are objects with complex ties to people, places and events. Curious and dedicated people from all fields spend a lot of time and energy to uncover these ties. Why then, some people ask, would you embed what is known about a work today in a digital image where it will live on as long as the image exists? If research reveals that a painting was created in 1678 by a follower of Rembrandt rather than in 1633 by the master himself, old images with that outdated information will misinform everyone who sees them.

The Federal Agencies Still Image Working Group's document, "Basic Guidelines for Minimal Descriptive Embedded Metadata in Digital Images" [1] explains the case for embedding only minimal metadata:

"Embedded metadata by its nature is part of the image file and therefore travels with the file throughout the entire lifecycle from creation to archiving or deletion. As a result, data embedded should be information that will remain valid as long as the file exists which, for a museum, is theoretically forever."

The true question is: which information about a cultural heritage work will remain valid as long as an image file exists? It is safe to say that most of the technical metadata will remain valid, but even something as straightforward as who created an image is vulnerable to change. While people rarely change their names, institutions do. For example, The Smithsonian American Art

Museum has had many names over the years - National Gallery of Art of the Smithsonian Institution, National Collection of Fine Arts, and National Museum of American Art. It has been known by its current name since only October 2000 [2]. Does this mean that elements that use names, such as Image Creator, Source, and Provider should not be embedded?

There are some metadata elements that are extremely unlikely to change—so unlikely they could be considered valid forever. For example, the date the image was created is a fact. Other elements are intended, by policy and best practice, to never change: e.g. a work of art's accession number. Out of the ten required core elements in the Federal Guidelines, only three could be considered valid for life: Document Title (defined as a unique identifier), Date, and Job Identifier. The other seven seem too vulnerable to change during the lifetime of the image. Let us consider rights information: it is not uncommon for works once designated “copyrighted” to be changed to “public domain”. It is also possible for a work once considered to be “orphaned” [3] later to be attributed to a living artist who wishes it to be copyrighted. Even with the real possibility of the Copyright Notice changing, most organizations would not be comfortable publishing images without it.

If seven out of ten of the Federal Guidelines required core elements fail the “valid forever” requirement, why are they included in the standard? Because they are useful. We have to accept that some “likely to change” information is too valuable to leave out. That is why the Federal Guidelines includes the Keywords element. Keywords are extremely useful for identifying, finding and sorting images, but they are not valid for life. Subject vocabularies to alter and deprecate terms change quite often in response to community practice, but this is not a good enough reason to exclude them from embedded metadata. The same is true of the required elements Copyright Notice, Source, Creator, Description, Credit/Provider, and Headline. The information in these fields cannot be guaranteed forever, but they are crucial pieces of information. If an exception to the “valid for life” rule can be made for these, then why not allow the embedding of descriptive information in image files of cultural heritage works? At the least, basic “tombstone” information such as Artist, Title, Date, Material, Measurements, and Site should be included - their usefulness should override an expectation that they are perfect forever. If other information such as Work type, Technique, Style, Culture, or Subject are known, they should also be included.

When you share a digital image--you are—in a sense—publishing it. If you were publishing a printed article with an image of a sculpture you would probably include a caption beneath it even though that information might be outdated someday. There are certainly many millions of printed images of artworks with invalid captions, but that does not mean it was a bad idea to publish them nor does it render them useless today. Even if information, such as the current location of an artwork, is no longer true, it is still better than no information at all. If you were researching a box of old photo prints, which would you prefer to find: a photo with notes written on the back, or a photo with nothing written on the back? Assuming you would prefer the photo with the notes, would you take everything you read as fact or would you take it as a head start that needs to be confirmed before you build your thesis around it?

A frequently suggested solution to avoid out of date embedded metadata is to only embed identifiers or links that take users to a database with the most current metadata available. This approach is appealing because the user would always have access to rich, accurate data. However, embedding only URIs and leaving out the textual descriptive metadata would be missing a good portion of the usefulness of embedded metadata: desktop searching which is done through text. Also, if “valid forever” is the criteria, can you count with 100% certainty on an identifier or its linking website to be available in the long term?

Why not embed both unique identifiers and full descriptive metadata? Textual description allows local search and identification, even offline, while the identifiers allow users to confirm that the information is still true. Identifiers encoded as linked data [4] could also allow applications to refresh embedded metadata or to display richer, up to date information, such as the fact that the painting shown is currently in a travelling exhibition. As useful as identifiers are, it is not realistic to expect them to be valid forever just as textual metadata attached to any object is not valid forever. Both approaches—minimal identifiers and full descriptive text—might be considered flawed, but, in the end, are too useful to be excluded from embedded metadata practices.

References:

1. Embedded Metadata Working Group – Smithsonian Institution. 2010. Basic Guidelines for Minimal Descriptive Embedded Metadata in Digital Images
http://www.digitizationguidelines.gov/guidelines/digitize-core_embedded_metadata.html
2. National Collection of Fine Arts (U.S.) <http://id.loc.gov/authorities/names/n80033904.html>
“Upon creation of a new National Gallery of Art in 1937, the name of the earlier National Gallery of Art of the Smithsonian Institution was changed to National Collection of Fine Arts.”

National Museum of American Art - <http://id.loc.gov/authorities/names/n81062959.html>
“Aspects of collecting the modern, 1981: t.p. (National Museum of American Art, formerly National Collection of Fine Arts) text (remains a part of Smithsonian Institution)”

Smithsonian American Art Museum - <http://id.loc.gov/authorities/names/no2001011043.html>
“Smithsonian American Art Museum) press release, Oct. 27, 2000 (President Clinton signed bill renaming the National Museum of American Art on Oct. 27, 2000; new name is: Smithsonian American Art Museum”
3. Borgman, Christine L. 2007. Scholarship in the digital age information, infrastructure, and the Internet. Cambridge, Mass: MIT Press, 108.
4. Bizer, Christian, Tom Heath, and Tim Berners-Lee. 2009. "Linked Data - The Story So Far". International Journal on Semantic Web and Information Systems. 5 (3): 1-22.