

MAXIMIZING METADATA: EMBEDDED METADATA TOOLS

Marcia Focht

Department of Art History
Binghamton University
Binghamton, New York, USA
mfocht@binghamton.edu

INTRODUCTION

Archives, libraries, museums, businesses, and individuals have fully embraced the digital era and a proliferation of digital assets abound. The challenge is creating, managing, storing, and sharing these assets. Metadata is information that helps overcome this challenge. It is data that provides definition, descriptions, and context, all set into a structure that describes how the data itself is arranged. Metadata helps us understand and organize our digital assets. It can answer questions, such as (but not limited to) who, what, where, when and how. Metadata can be used to add value to your work by adding standardized information that will aid in structuring, organizing, discovering, and sharing your digital assets.

WHAT IS METADATA?

Metadata is a structured way of describing an information resource. It can be simple or complex, depending on your needs. The metadata discussed in the following paper pertains to cultural heritage objects, but metadata can and should be created for anything that can be digitized or classified or organized: online catalogs, search engines, books, magazines, newspapers, video, music, and so on.

The following table obtained from the American National Information Standards Organization (NISO) summarizes the various types of metadata succinctly. [Table 1] NISO explains it thus: Descriptive Metadata is used to find or understand a resource. Technical, Preservation, and Rights metadata (all considered “administrative” metadata) are primarily used for the management of files. Structural metadata details the relationships of parts of resources to one another, and markup languages integrate metadata for structural or semantic features within content. [1] [Figure 1]

Metadata has an underlying structure to it. There are ELEMENTS (such as title, subject, etc.). In the example above, “Creator” has a value of “Green, Ruzzie, United States, American, 1892-1956” and the element “Date” has a value of “1941” and so on. STANDARDS, (a group of elements approved by professionals) are developed in response to the group need. For example, the Dublin Core Metadata Element Set (DCMES) is a schema used to describe digital and physical resources that has become a minimum standard in many archives and libraries. [2] Other metadata standards, like the Darwin Core, are specifically geared to items that relate to biology and can be viewed as an extension of the Dublin Core for biodiversity information. [3] The VRA Core (the standard chosen for this example), developed by the Visual Resources Association and hosted by the Library of Congress, contains other expanded elements to more adequately describe works of visual culture and the images that document them. [4] Those who work with collections of digital assets should choose the standard that most closely aligns with the nature of their materials, or customize, if need be. A SCHEMA conveys the underlying structure, that combination of elements and standards, the relationships between elements, and the rules for use and syntax for the values. The SCHEMA is often the markup language that the end users do not see, but allows the metadata to be machine-readable. For example, XML and RDF are common markup languages. [5]

WHAT IS EMBEDDED METADATA?

We understand that Metadata helps describe our digital assets. What then, is embedded metadata? Embedded metadata is created using a process that converts text into code and adds that information about an asset directly into the digital file itself. These metadata bytes are stored just like the image bytes are stored. Many of the digital files we use every day already have metadata embedded in them. Cameras and scanners automatically embed metadata such as date, time, and device settings. Music files, e-mails, texts, web pages, and more are all packed with metadata; the ‘key to functionality of the systems holding the content, enabling users to find items of interest, record essential information about them, and share that information with others’. [6]

Researchers, archivists, curators, and educators often have knowledge and data that go far beyond the automatically stored metadata attached to their digital images, but need to find efficient ways to add this value. Depending on the nature of the materials and the schema a person wishes to follow, there are different types of data that can be added to describe those digital files. This additional information about the objects, people, and places described in an image or audio file can be embedded. It is important to take the time to decide what information you want to add, and then to be consistent in your application. [7] Additionally, administrative metadata can be quite useful to manage images, such as telling users who owns the image and how it can be used. This information can also be embedded.

Not only digital images can have embedded metadata inserted; it can be done to digital videos, audio files, and PDF’s. Embedding metadata into a file allows the descriptive information to travel with the file across file formats and software applications. When you embed data into your digital file, it will sit alongside the already automatically generated embedded information, for example technical metadata from a camera, and you will be able to search for it and view it on your computer. Improving your digital files discoverability and accessibility is a key advantage to embedding metadata. Conducting a search on your own computer, in a database, or other online collection can typically involve the data embedded in the files.

Embedded metadata helps reduce the difficulty of finding an image or audio file on your computer, no need to recall a lengthy URL or scratch your head trying to remember the original source to cite when using or publishing your images. Ensuring that those who might wish to use your images are informed about copyright constraints and provided with contact information is easier than you might think, since this type of data can be easily embedded. Embedded metadata enables functionality that streamlines and enhances your research and presentations by making images, video, and audio files searchable and sharable, with easy to identify content, sources, and rights.

TOOLS FOR EMBEDDED METADATA

There are many tools out there that can be used to embed metadata: Corel Paintshop, Apple Photos, ACDSee, Lightroom, Photoshop Elements, and more. They were all developed to work with an image caption standard called IPTC, which is the global standards body for the news media providing the technical foundation to allow for sharing information about images in a structured, consistent way. [8]

There now are tools available that will allow you to improve workflow, build upon work that has already been done, and eliminate the need to slavishly enter information more repetitively, saving time and money. Three such tools were developed by the Visual Resources Association, a multi-disciplinary organization dedicated to furthering research and education in the field of image and media management. [9] These tools were designed to work in Adobe Bridge and can be used to create and manage embedded metadata, making your workflow much more efficient. Although most Adobe products are now only available by paying for a subscription called Adobe Creative Cloud, Bridge can be downloaded for free once you create an account. [10]

VRA METADATA PALETTE

The first tool is a custom XMP palette, which allows you to embed data into your digital files. Knowing that those who deal with material culture have a need for more description, especially when data about the object depicted is needed rather than just data about the image itself, the Visual Resources Association developed a tool that will allow for richer data using IPTC and the VRA Core. [11] We call it the VRA metadata palette. This tool allows you to easily enter data into a digital file, carefully separating out information about the artwork or object depicted from the image data, while also allowing for the addition of administration data (which is useful for archivists and librarians, or anyone handling projects with multiple workers and roles). Data can be entered one digital file at a time, or in batches when some information holds true for all in the group, such as source or copyright. [Figure 2]

The palette utilizes the Getty Vocabularies Linked Data Service, which includes: the Union List of Artist Names, the Thesaurus of Geographic Names, and the Art and Architecture Thesaurus. [12] Using such authorities, structured terminology compliant with international standards, helps create metadata that is easily searchable and more discoverable. Institutions use metadata standards, authorities, and controlled vocabularies for two primary reasons; quality control and interoperability. The controlled vocabularies most commonly used for image and cultural object cataloging are those provided by the Getty Vocabulary Program. In the example above, the user can search the Getty ULAN for the architect “Leon Alberti”, select his name, and the rich metadata (name, nationality, role, dates, etc.) are automatically added, along with authority reference ID’s, which are useful for preparing your data for use in linked data environments. [Figure 3]

I’ve used the VRA palette to quickly record data for images taken from books and journals on loan. It could just as easily be used by researchers in the field or anywhere order needs to be imposed on files before information is lost and identification becomes difficult. Students with minimal training can search for names and terms and then copy accurate information into the file. The information stays with the file until it is uploaded into a database, added to a website, sent to a publisher, or so on.

Once this detailed data is entered and ordered using the VRA Core standard, the summary section will fill this data into IPTC and Dublin Core fields, manually, or automatically (see below). Populating the IPTC metadata allows for almost all systems to discover and display the embedded metadata. [Figure 4]

VRA METADATA EXPORT/IMPORT TOOL

The second VRA tool allows for importing or exporting digital file data to and from spreadsheets (perfect for all those digital humanities projects). Using the VRA Export/Import tool allows the user to export all of the data from single files or groups of files into spreadsheets. These spreadsheets can be manipulated and put into data sets, used to populate databases, and so on. By the way, the reverse is true, you can also take spreadsheets of data, import them, and embed thousands of files in one fell swoop. I’ve used the import/export tool many times. When I purchased high quality images from Archivision, they came with a spreadsheet of rich and complete descriptive metadata. I embedded that data into the images using the Import tool before I loaded them into my online database. Now when a user downloads one of the images, the data travels with the image. This makes it easier for my patrons to search for specific images they have downloaded, as well as keep track of the detailed information about each file, with no effort on their part. Here is another example; I scanned over 9,000 slides from Binghamton University Theatre Department productions. Using a spreadsheet they had produced with information about each production, I was able to import and embed that information into all the files in just minutes. Then, using the VRA palette, we added copyright and contact information, as well as standardized subject terms. Next, we exported the improved data into a spreadsheet that we used to populate data in the JStor Forum, where the images

were then stored and published. [13]

The beauty of these tools can be found in their back-and-forth utility and flexibility: creating data, exporting data, importing data, updating data, and more. This gives you the freedom to work in whatever sequence or on whatever platform you prefer and, it allows you the option to change your mind down the road.

VRA METASHOTPPT

The last VRA tool is called the MetaShotPpt. It quickly allows for the creation of Powerpoint presentations in which digital images and embedded metadata automatically load into the slides. Simply open the tool, select a folder of embedded images, select desired options, and then click “Create PowerPoint”. It just takes moments. Nothing has to be retyped, saving time and ensuring accuracy. [Figure 5]

CONCLUSION

Embedded metadata is so useful, and the time and effort it takes to create and embed data yields returns to the end user—saving time, increasing accuracy, and providing long term accessibility. The VRA tools are easy to use and free to download from <http://metadatadeluxe.pbworks.com/w/page/20792238/FrontPage>, along with guides for embedded metadata and easy to follow YouTube tutorials. [14]

All of this is made possible because of the power of embedded metadata, which makes digital images independent, self-defining, sharable, flexible and portable.

FIGURES AND TABLES

Metadata Type	Example Properties	Primary Uses
Descriptive metadata	Title Author Subject Genre Publication date	Discovery Display Interoperability
Technical metadata	File type File size Creation date/time Compression scheme	Interoperability Digital object management Preservation
Preservation metadata	Checksum Preservation event	Interoperability Digital object management Preservation
Rights metadata	Copyright status License terms Rights holder	Interoperability Digital object management
Structural metadata	Sequence Place in hierarchy	Navigation
Markup languages	Paragraph Heading List Name Date	Navigation Interoperability

Table 1. Table obtained from the American National Information Standards Organization (NISO)

Item Details	
	Institution Collections
Creator	Green, Ruzzie, United States, American, 1892-1956
Culture	American
Title	Colgate Palmolive Ad
Date	1941

Figure 1. Example of an image and associated metadata elements

WORK

Agent	Francesco Talenti (Italian architect, ca. 1300-1370); Giovanni Rucellai (Ita
Title	Santa Maria Novella
Date	1458-1470 (façade) (alteration); ca. 1246-1399 (creation)
Style/Period	Medieval; Renaissance
Cultural Context	Italian
Work Type	architecture
Material	marble
Technique	construction (assembling)
Measurements	
Location	
Description	The second design commissioned by Giovanni Rucellai, the spectacular gr
Subject	buildings; religious buildings; churches
Inscription	
State/Edition	
Relation	
Text Reference	
Source	Grove Art Online; http://www.groveart.com/

Figure 2. Screenshot view of Adobe Bridge with VRA palette open

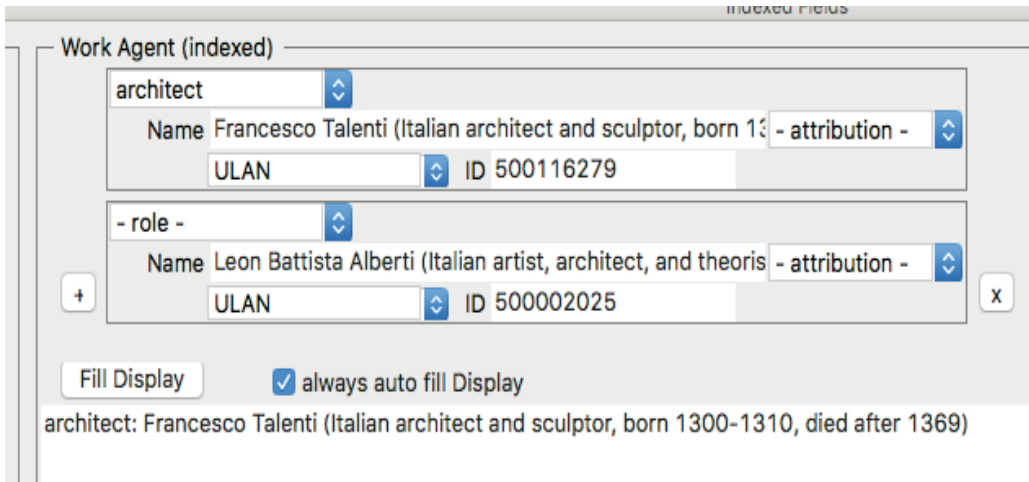


Figure 3. Screenshot of the VRA palette showing the links to the controlled vocabularies

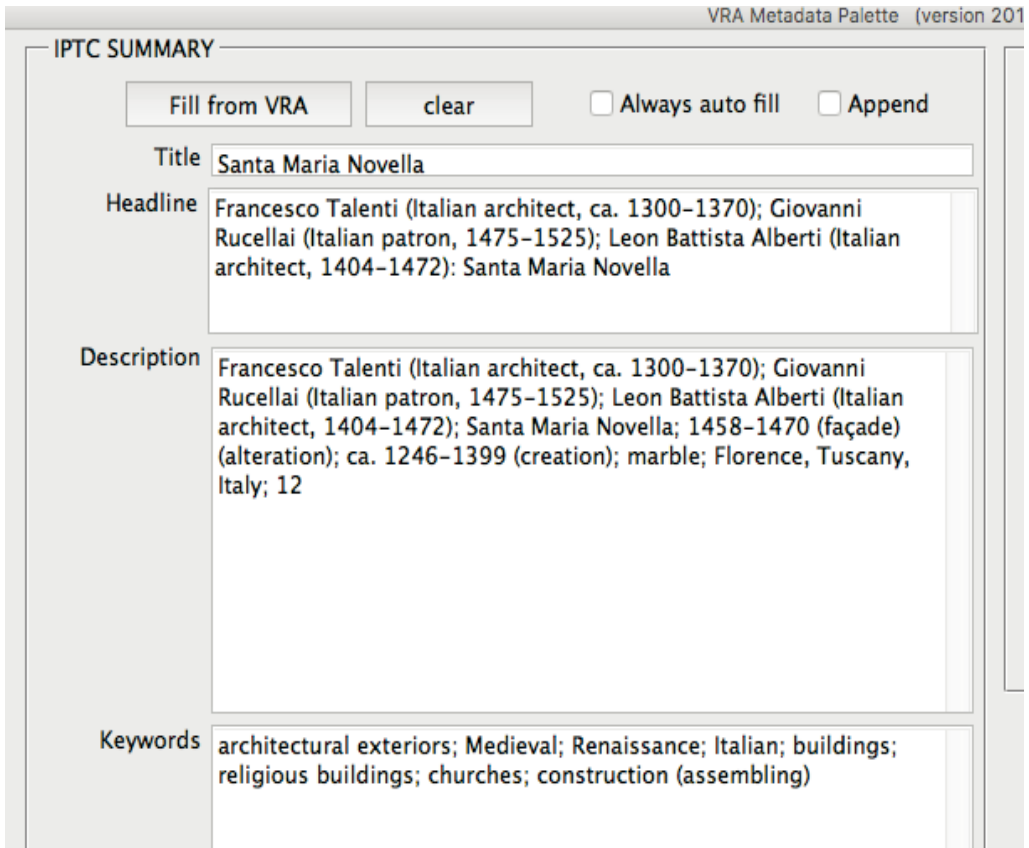


Figure 4. Screenshot of populating IPTC from VRA Core



Figure 5. Screenshot of MetaShotPpt Tool

References

- [1] More information about NISO standards can be found at: “Understanding Metadata.Pdf,” accessed March 7, 2018, https://groups.niso.org/apps/group_public/download.php/17446/Understanding_Metadata.pdf
- [2] Dublin Core Metadata Initiative (DCMI) information can be found at: “DCMI: Home,” accessed March 7, 2018, <http://dublincore.org/>
- [3] This biodiversity information standard can be found at: Darwin Core Task Group, “Darwin Core,” accessed March 7, 2018, <http://rs.tdwg.org/dwc/>
- [4] For more information on the VRA Core, see: “VRA CORE - a Data Standard for the Description of Works of Visual Culture: Official Web Site (Library of Congress),” accessed March 7, 2018, <https://www.loc.gov/standards/vracore/>
- [5] More about Standards and Schemas can be found here: “Metadata Standards - CCO Commons - Cataloging Cultural Objects,” accessed March 7, 2018, http://cco.vrafoundation.org/index.php/toolkit/metadata_standards/
- [6] “Understanding Metadata.Pdf.” https://groups.niso.org/apps/group_public/download.php/17446/Understanding_Metadata.pdf
- [7] Greg Reser, “What Not to Embed: Is It Better Not to Embed Certain Cultural Heritage Metadata in Images?,” *VRA Bulletin* 39, no. 1 (December 10, 2012), <https://online.vraweb.org/vrab/vol39/iss1/5>
- [8] For more information on IPTC see: “Home Page,” IPTC, accessed March 7, 2018, <https://iptc.org/>
- [9] Information about the Visual Resources Association can be found at: “About | Visual Resources Association,” accessed March 7, 2018, <http://vraweb.org/about/>
- [10] For more information about Adobe Bridge, and to download a free copy, see: “Digital Asset Management Software | Download Free Adobe Bridge CC Trial,” accessed March 7, 2018, <https://www.adobe.com/products/bridge.html>
- [11] The Data Standards Committee oversees the VRA Core standard and Embedded Metadata Working Group, see: “Data Standards Committee | Visual Resources Association,” accessed March 7, 2018, <http://vraweb.org/about/committees/data-standards-committee/>
- [12] More information about the Getty vocabularies can be obtained at: “Getty Vocabularies (Getty Research Institute),” accessed March 7, 2018, <http://www.getty.edu/research/tools/vocabularies/>
- [13] See the Binghamton University Theatre Collection here: Artstor, “Artstor,” Artwork, accessed March 7, 2018, <http://library.artstor.org/-/collection/87731951>
- [14] The VRA tools are free to download and can be downloaded from “Metadata Deluxe / FrontPage,” accessed March 7, 2018, <http://metadatadeluxe.pbworks.com/w/page/20792238/FrontPage>