

METADATA MADNESS & MDID

Jeanette Mills
Director of Visual Services
University of Washington, School of Art

I remember hearing the word "metadata" for the first time in the 1990s and thinking that it was a somewhat pretentious word. Now it is a word that I use almost every day. Somewhere in between was a transition that took place in my thinking. That transition was prompted by the adoption of MDID as our digital image database and the start of what became metadata madness for the three staff, including myself, who are the Visual Services team in the University of Washington School of Art.

It all began in early 2004 when we decided that something really had to be done about developing a digital image database for long-term use. We had been making low-resolution scans for student review through the University of Washington Libraries' electronic reserves system since autumn 2001, but we knew that this was not a long-term solution because these images would not work well for classroom projection. I started to make noise in 2002 about wanting to use MDID since I had seen it demonstrated and thought it would fit our needs. The price tag for the software—FREE—was also an enticement since our budget did not allow for purchasing anything like Luna Insight. Unfortunately, the School of Art did not yet have the infrastructure for supporting MDID or the storage of large image files. By 2004 that had changed, so during that Spring Quarter the Visual Services team met with the School of Art's IT specialist to make plans.

The goal was to have the first course taught with MDID in Spring Quarter 2005, one year away. Our IT specialist installed the database during autumn 2004 and got it running. I then set about creating a collection, using fields from our Filemaker Pro image cataloging database as a guide. I figured out how to export the data, convert it, and upload it to MDID. This was somewhat complicated by the fact that we operate in a Mac environment but, with help from Virtual PC, we got it to work.

The first course using MDID images did happen in spring 2005, but the professor ended up downloading the medium-sized images from MDID to use in PowerPoint rather than using MDID's ImageViewer software. At the time, the ImageViewer software was not functioning reliably with the professor's Apple laptop computer, so we felt we had to go with a more predictable presentation tool. That problem was later solved by an MDID upgrade in summer 2005. Since then we have had a professor and her four teaching assistants successfully use MDID and the ImageViewer software for a large survey course.

Around the same time in 2004 that we were planning for the implementation of MDID, we were also involved with the application for and, ultimately, the receipt of a University of Washington Fund for Innovation and Redesign grant. This effort was spearheaded by Denise Hattwig at the University of Washington Bothell campus and brought together the three University of Washington visual resources collections, the University of Washington Libraries, and faculty from the University of Washington Tacoma campus, which has had no campus image collection. The goal of the grant was to develop a centralized image collection, known as the Image Bank, that could be shared by all three campuses. By late spring/early summer 2005 the group was making decisions about what digital images to license. Along with that came the task of developing metadata standards for the collection. Members of the University of Washington Libraries' Metadata Implementation Group met with our group, and we created an

Image Bank data dictionary that was then mapped to both the Dublin Core and VRA Core. Figure 1 is the first page of the data dictionary.

UW Image Bank Data Dictionary April 17, 2006						
Collection title: UW Image Bank Metadata contact: Denise Hattwig						
Field Name	VRA Core 3.0	Dublin Core	Display (Y/N)	Search-able (Y/N)	Authority	Comments
Title	Title	Title	Y	Y		<p>Title of work, provided by Image Source.</p> <p>Use English-language Title, except in the case of dedications (architecture) or proper names.</p> <p>Capitalize all words in Title, except prepositions, articles, conjunctions. Avoid abbreviations.</p> <p>Do not include information that is captured in other fields, unless necessary for clarity.</p> <p>For architecture, include a descriptor of the building's function. For churches and cathedrals, include the city and dedication for clarity. Generally, use City Cathedral (Dedication) for cathedrals and Church of Dedication in City for churches. For manuscripts and illuminations, use the common Title of the manuscript with the shelf mark in (). Common usage should generally determine preferred Title, with Grove Art Online*** and museums and repositories as resources. Alternate (or non-English) titles may be included in () after the preferred title.</p> <p>Examples: Edgar J. Kaufmann House (Fallingwater); Lake Shore Drive Apartments (860 Lake Shore Drive); Portland Public Services Building; Chartres Cathedral (Notre Dame); Church of Sant'Ivo della Sapienza in Rome; Madame X (Madame Pierre Gautreau); Farnese Hours (MS M. 69)</p> <p>In the case of a work that is part of a larger work, include larger work title and smaller work title in the format "Larger Work Title. Smaller Work Title". If applicable, the larger work should be the site or complex for architecture; and the series, program, or site-specific work for art. The smaller work should be the building for architecture, and should be the smallest discrete work for art. The named Creator can often serve as a guide – the program created by the named Creator is the larger work, and the smaller named works within it are the smaller works.</p>

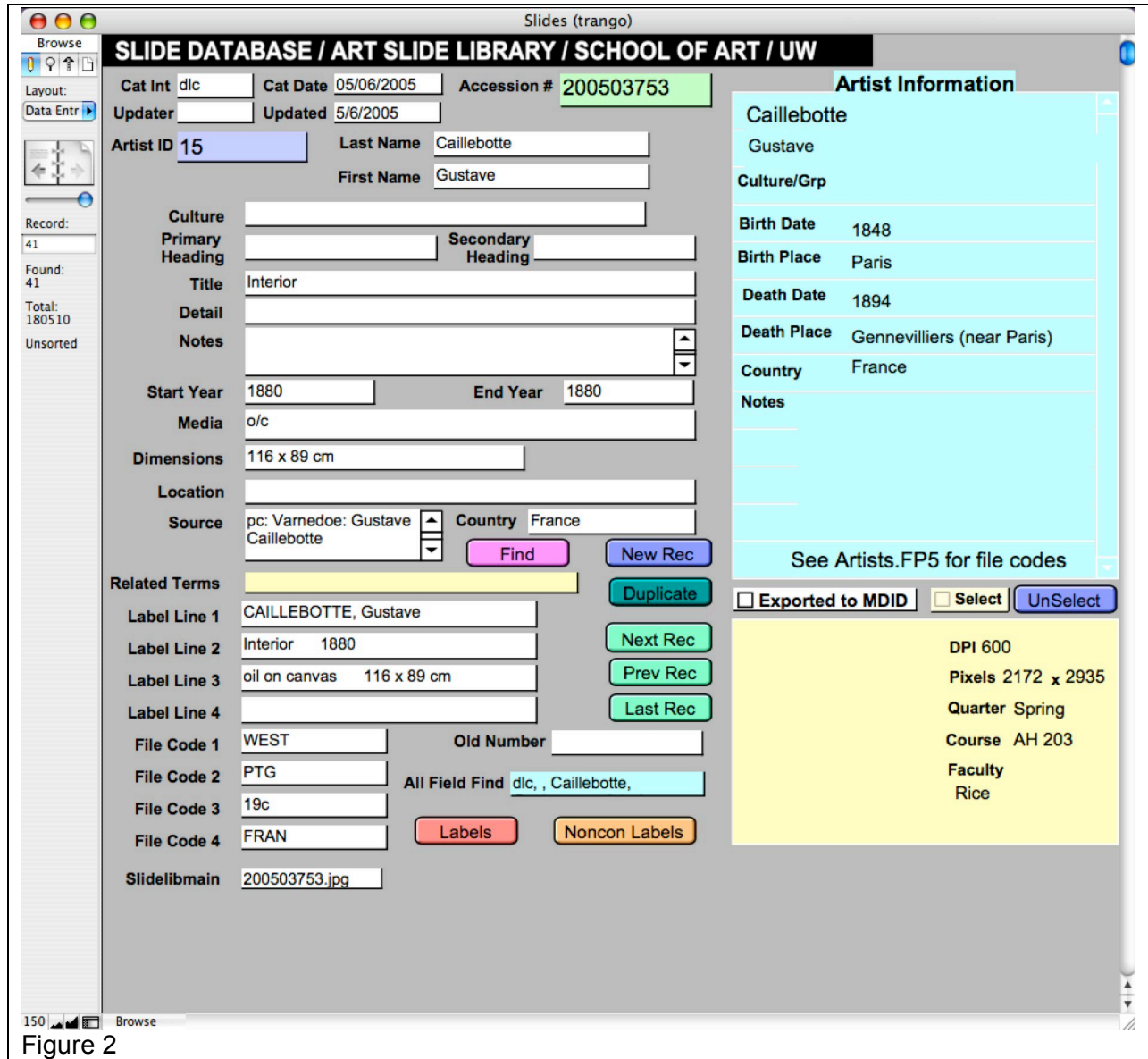
Figure 1

Up to the point when the data dictionary discussion began, I was still hiding my head in the sand when it came to metadata. I knew our Filemaker Pro database and, therefore, our MDID database were not ideally designed, but I was not quite able to get myself to face the work that would be needed to fix it. Preparing for the Image Bank data dictionary development meant that I finally had to really look at the VRA Core and Dublin Core, and that was my wake-up call.

Our database was originally developed using FoxPro back in 1988-89 by Jo Nilsson, who was then the Director of Visual Services, with programming done by staff in another University of Washington department. About ten years later we migrated to Filemaker Pro, which gave us, as non-programmers, much more flexibility. It also gave us the cross-platform functionality that was required. We quickly started adding fields, but I never took the time to compare our metadata with the VRA Core, which was being developed around the same time. Over the next few years we also added two associated databases—one for artist metadata and one for metadata about the scans made from slides.

In summer 2005 we carefully went through our databases, comparing them with the VRA Core, Dublin Core, and the Image Bank data dictionary, once it was completed. We also used Cataloging Cultural Objects (CCO) as a guideline for some of our work, but we were not

attempting to be CCO-compliant. The end result was the addition of eleven fields to the slide database (now called the image database). Just to give you a quick visual comparison, figure 2 shows what the data entry screen looked like for the old database. The two colored blocks on the right are portals to our other databases. Figure 3 shows what the image database data entry screen looks like now.



The layout and the number of fields have changed dramatically. A few fields were also added to the artist database (now called the creator database). Some of the new fields combined data from two fields in the older versions of the databases. For example, we had separate last name and first name fields, and these were combined into one creator field. Other fields separated out information that had once been combined in one field. An example of this is that one location field now became three fields: location: repository, location: site, and location: original. Some fields were completely new, such as work type, work style/period, and century.

IMAGE DATABASE / VISUAL SERVICES / SCHOOL OF ART / UNIVERSITY OF WASHINGTON

Cat Int: jcm Cat Date: 08/04/2005 Accession #: 200505573 Creator ID: 10543

Updater: jcm Updated: 8/19/2005 Old #: _____

Last Name: Gold First Name: Pat Courtney Creator: Gold, Pat Courtney

Culture: Warm Springs/Wasco

Primary: _____ Secondary: _____

Headings: _____

Title: Round bag

Detail or View: _____

Work Notes: Burke #2000-122/1; humanoid faces, dogs, star; inspired by Sally bag collected by Lewis & Clark in 1805-1806 (see 200108847 & _____)

Work Type: Fiber art, Containers Wk Style/Period: Native North Century: 20c

Start Yr: 2000 End Yr: 2000 Date: 2000

Material: Hungarian hemp, cotton yarn, chenille yarn, synthetic suede

Technique: weaving, twining

Dimensions: 11.5 x 7"

Location: Repository: Burke Museum, Seattle, WA, USA

Location: Site: _____

Location: Original: _____

Country: _____

Image Source: scan: Burke Museum slide

Slidelib: 200505573.jpg

IB Identifier: soa200505573.tif

Label Line 1: _____
 Label Line 2: _____
 Label Line 3: _____
 Label Line 4: _____

File Code 1: _____
 File Code 2: _____
 File Code 3: _____
 File Code 4: _____

Creator Information

Last Name: Gold
 First Name: Pat Courtney
 Creator: Gold, Pat Courtney
 Culture: Warm Springs/Wasco
 Artist: Native North American--Plateau
 Style/Period: _____
 Birth Year: 1939
 Birth Place: _____
 Death Year: _____
 Death Place: _____
 Country: USA

Creator Notes: _____

SEE ARTISTS .FP7 FOR FILE CODES AND ADDITIONAL INFORMATION

Qtr Scanned: Summer 2005 File Size: 17.9M DPI: 600

Dimensions: 1979 x 3168

Professor: Wright
 Course: AH 331

SEE SLIDE IMAGES .FP7 FOR COMPLETE INFORMATION

Figure 3

Working with the heavily changed image database was a challenge at first. We quickly discovered that we were having to re-learn how to catalog since there were so many new and revised fields. We also had to develop a work type list and to start developing a style/period list, both based heavily on the Art and Architecture Thesaurus. For the first several months our primary cataloger, Debra Cox, and I spent time every day discussing the minute details of what should and should not go into particular fields. Faculty or other staff would sometimes walk into our office in the middle of one of these conversations and look at us as if we were speaking a foreign language. We often found ourselves re-working records multiple times until we were satisfied with the results. To put it mildly, we had developed full-blown cases of metadata madness. Thankfully, over time, the detailed metadata discussions have diminished to the point of happening only once a week or so. Normalcy, such as it is, is beginning to return.

After all the Filemaker Pro database revisions were done, I then had to face MDID. We realized that one image collection, into which everything had been thrown up to that point, was not going to be viable in the long term. It made the most sense for me to re-work everything at the same time. I started by creating a main collection that would have the broadest access and then several additional collections, which would have varying levels of restricted access. For example, one of the courses that was scheduled to be taught with MDID was going to have a significant number of scans from personal slides belonging to the professor and her predecessor. While some of these images could be shared widely, the majority of them had been shot in museum collections or during ceremonies, and permissions did not exist for widespread dissemination. After discussion with the professor, we decided that these images could be made available to any School of Art professors and teaching assistants who would be teaching in this subject area. To make such a large number of images available in this way, it

was easiest for us to create a special collection in our MDID rather than set up a personal collection for her.

Next we completely re-thought what fields were appropriate for MDID. Clearly, not every field in the image and creator databases was necessary for MDID, so we winnowed the list until we felt we had the essential information that users would want. I made a crosswalk (figure 4). When I then created the field definitions in MDID, this made it possible for me to map each field to Dublin Core with relative ease. In our image and creator databases, I colored the fields that go into MDID—they are the pink fields—so it would be easier for us and others that we train in cataloging to quickly be sure that the most important fields had been filled in. I also documented the information and settings for each of our MDID collections in a Word document. In a way, I am recording the metadata about the database.

University of Washington
School of Art
Visual Services

DATABASE FIELD MAPPING

Image DB Field	SoA MDID	Dublin Core	VRA Core	Image Bank
Accession Number	<u>AccessionNumber</u>	Identifier	ID Number	Contributor Accession Number
Creator	<u>Creator</u>	Creator	Creator	Creator (Display)
Creator Style/Period	<u>CreatorStyleorPeriod</u>	Subject	Style/Period	Style/Period
Creator Notes	<u>CreatorNotes</u>	Description	Description	Notes
Culture	<u>Culture</u>	Coverage	Culture	Culture
Title	<u>Title</u>	Title	Title	Title
Detail/View	<u>DetailorView</u>	Description	Title	Detail or View
Work Style/Period	<u>WorkStyleorPeriod</u>	Subject	Style/Period	Style/Period
Work Type	<u>WorkType</u>	Type	Type	Work Type
Work Notes	<u>WorkNotes</u>	Description	Description	Notes
Century	<u>Century</u>	Coverage. Temporal	Date	Date (Century)
Date	<u>Date</u>	Date	Date	Date (Display)
Material	<u>Material</u>	Format. Medium	Material	Material
Technique	<u>Technique</u>	Format	Technique	Material
Dimensions	<u>Dimensions</u>	Format	Measurements	Measurements
Location: Repository	<u>Location:Repository</u>	Coverage. Spatial	Location. Current Repository	Current Location – Repository
Location: Site	<u>Location:Site</u>	Coverage. Spatial	Location. Current Site	Current Location – Site
Location: Original	<u>Location:Original</u>	Coverage. Spatial	Location	Creation Location
Headings	<u>Headings</u>	Description	Description	Related Terms
Related Terms	<u>RelatedTerms</u>	Description	Description	Related Terms
Image Source	<u>ImageSource</u>	Source	Source	Image Source

Figure 4

Even though we mapped all our MDID fields to Dublin Core, we initially made the decision to limit user searching to the keyword field. We felt that most users would tend to use only the keyword field anyway and that other fields, such as work style/period, might puzzle them. In practice, however, we have discovered that users sometimes received confusing search results when using an artist name since some artists are mentioned in other artists' creator notes and both types of records would be retrieved with a keyword search. Because of this, we have made the creator field searchable for all collections. Users do have to be told that names searched in the creator field must be entered with surname first since word order in this field must match the original data entry in order to get accurate results. At one point we thought it also would help to make the culture field searchable, but, as soon as one searches across more than one collection, the label for this field defaults to the Dublin Core label, which is the word "coverage." While the Visual Services staff might know that coverage is the same field as culture, most users would not know that, so we have dropped this field. If a capability is added to MDID at some point to allow for different labeling of fields when searching multiple

collections, then we will add that field again to the search options. This issue was not a problem for the creator field since Dublin Core also uses the term "creator" for that category.

Once the Filemaker Pro and MDID database redesign was near completion in late summer/early autumn 2005, a truly painful reality had to be faced. More than 1,400 records were already in MDID and roughly another 2,000 records were waiting to be loaded. This was because, in addition to scanning one full course every other quarter or so (with that scanning outsourced), we also had been doing high-resolution scanning in-house for any new course-review images, with the intention that those would go into MDID as well. A bit of a backlog was developing, and it was going to get worse in the short term since all those records had to be re-edited to our new metadata standards before they could be imported or re-imported into MDID. Not a pleasant thought, but our primary cataloger has tackled this a bit at a time and has made significant progress. We also want to train graduate student employees to do more cataloging, but it has not been easy to find people who truly understand the importance of metadata standards. Either way, we are sincerely hoping that by the end of summer 2006 we will be able to get ahead of this backlog.

As we have added records to MDID, we have discovered quirks that affect the way we enter metadata. One example is that, unlike our Filemaker Pro databases, MDID does not search for partial terms. If you search on the word "photo" and the work type in a record is the word "photographs," then you will not get that record. In cases where we feel that people may easily search on either form of a word, we have added the word form not used in the major fields in our related terms field. A more significant issue for us is that words that contain an apostrophe are not searchable. For most English words, this is not a problem. However, it becomes a very noticeable problem for Native American words. During Autumn Quarter 2005, I cataloged over 1,000 images of Northwest Coast Native art for MDID. Many of the words I used in cataloging contained apostrophes, such as the name of one of the tribal groups: Kwakwaka'wakw. I had uploaded a number of records with this word before I realized that searching on it in MDID brought no results. I then had to go back and enter the word without punctuation in the related terms field and tell the professor and teaching assistants using that material to search it that way in order to get results. Note that I did not use the underlined letter "a" when entering the name in Filemaker Pro because our experience has been that many diacritics are turned into gibberish during the metadata conversion process for MDID. Even smart quotes have caused problems in conversion, so I have shut off smart quotes in all our installations of Filemaker Pro. Because many tribal groups have multiple name forms, I use the primary form in the culture field and other forms in the related terms field. For example, the Kwakwaka'wakw have also been known as Kwakiutl and Kwagiulth. Providing this level of metadata for users does require some specialized knowledge, and it is not necessarily typical. We do the best job we can with the knowledge we have available to us, but we try not to get too wrapped up in making perfect records.

There is one interesting side note for database administrators about working with MDID. There are times when I am logged into MDID as a faculty member, such as when I help someone set up their first series of lectures, and I also want to have a separate browser tab or window logged in as an administrator in order to make record corrections or other adjustments as I see the need. I have discovered that, at least on Macintosh computers, this is impossible. As soon as you open the second window or tab and try to log in, you are immediately taken to the same place you were in the other window or tab. I have tried this in Safari, Firefox, and Internet Explorer. The only way around this that I have found is to open the database in two different browsers, such as Firefox and Safari. This certainly is not a big issue, but it is worth noting.

MDID has performed very well for us overall, especially after we wrestled with the metadata bear. I will not pretend that the process was loads of fun, but it definitely was a learning experience. I feel we are much better equipped to deal with metadata issues now and

that our users are being much better served. Becoming more compliant with metadata standards at first appeared to be a nearly insurmountable task, but we faced metadata madness and we have prevailed. I only hope that others can learn from our journey and do it right the first time, so they do not have to face some of the struggles that we have endured.

I have a Web site that I originally designed to keep faculty informed about our digital imaging progress, but it has also become a way for me to share some of the materials we have gathered or created. One group of items there is our workflows, an area about which a whole paper could be written, and they already need to be updated again. Feel free to take a look at this Web site: <http://staff.washington.edu/jcmills/Digital/Digital.html>. I hope to update and expand it in the near future.

POSTSCRIPT (late March 2006): One point I did not make clear when presenting this paper is that our Filemaker Pro database contains image records, not work records. In an ideal world, I would have made that change in our database during summer 2005 when I did the other work described above. However, with over 180,000 records in the database, it would set back our digital-imaging progress even more if I had chosen to do that. A compromise had to be made somewhere, so this is where it happened. Since MDID is currently a flat-file database (and I prefer that it stay that way for a variety of reasons), there is no point in putting the effort into making our Filemaker Pro image database more complex than it is at this point. Every visual resources professional will have to make decisions like this based on local needs and expectations.

POST-POSTSCRIPT (February 2007): Our Filemaker Pro and MDID databases and the way we enter metadata are continuing to evolve. To see more current examples of Figures 3 & 4 as well as other related documents, please visit the following web site: <http://staff.washington.edu/jcmills/Digital/Digital.html>