DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010

The One-To-One Principle: Challenges in Current Practice

Steven J. Miller
University of Wisconsin--Milwaukee, USA
mll@uwm.edu

Abstract

The DCMI One-to-One Principle holds that related but conceptually different entities, such as a photograph and a digital image of that photograph, should be represented by separate metadata records. In practice, however, large numbers of practitioners do not adhere to this principle and commonly mix elements representing two related entities in a single metadata record. This paper explores reasons why this is the case, why it is problematic, how the principle itself would benefit from greater clarity, some practical options for maintaining the principle in current systems, with advantages and disadvantages of each, and suggests a possible compromise option. The paper focuses on the widespread application context of small to medium-sized cultural heritage institutions digitizing unique local resources, creating metadata using digital collection software packages such as CONTENTdm, and exposing only simple Dublin Core metadata for OAI harvesting and aggregating.

Keywords: one-to-one principle; 1:1 principle; metadata; Dublin Core; content versus carrier; multiple versions; manifestations; CONTENTdm

1. The One-To-One Principle and Challenges in Its Practical Application

The One-to-One (or 1:1) Principle is a long-standing principle within the Dublin Core and other cultural heritage metadata communities (Han, et al., 2009, p. 230; Hillmann, 2005, sec. 1.2; Hutt and Riley, 2005, pp. 262, 264, and 268; Riley, et al., 2008, 231-236; VRA 3.0 Introduction; VRA 4.0 Introduction, p.1; Woodley). Concisely put, it is "The principle whereby related but conceptually different entities, for example a painting and a digital image of the painting, are described by separate metadata records" (Woodley). In practice, however, large numbers of cultural heritage institutions creating metadata for digital collections today do not adhere to this principle. Instead, they commonly mix elements representing two manifestations together in a single metadata record, most often elements describing both an original analog resource and digitized version of that resource. This paper is written primarily from the perspective of the small to medium-sized cultural heritage institutions digitizing unique local resources, creating metadata using digital collection software packages such as CONTENTdm, and exposing only simple Dublin Core metadata for OAI harvesting and aggregating.

There are several reasons why institutions do not maintain the One-To-One Principle in practice. Firstly, some implementers simply lack an awareness of the Principle or an understanding of the conceptual issues and potential problems of mixing elements for two different but related entities in the same record. Quite frequently, new or seasoned staff members in libraries, archives, museums, and other institutions are assigned to implement digitization projects, but they have little or no background in cataloging, resource description, or metadata. They may have little time or opportunity to attend workshops, conference sessions, or do professional reading. Even when people are aware of the distinction between metadata elements relating to the original versus the digital resource, they may not be thinking of future viability, sharing, harvesting and aggregating of their metadata outside of their immediate closed system application. But even when these things are not the case, the following four reasons still often hold true.

Secondly, many local database and user interface systems do not have the capacity to adequately link separate records and to display them together in a clear and meaningful way for

end users. This becomes a usability issue, and may drive the creation of single records instead of separate records for original and digital versions of a resource.

Thirdly, many practitioners, including those who are be well aware of the One-to-One principle, come to their digital collection projects with the intent to create records only for their digital resources. They are creating metadata for an online collection of digital resources, not a database or catalog of both their analog holdings and their digitized files. It often does not make sense to practitioners to create separate records for original physical archival photographs and slides, for example, when the database is for a purely online, digital collection.

Fourthly, commonly-used digital collection management software packages such as CONTENTdm are designed to link one metadata record to one digital object. This is the model they were built on and it drives the metadata creation process for large numbers of institutions.

Fifthly, for those who do use CONTENTdm and similar systems and models, it is not uncommon to design a separate metadata scheme or application profile for each separate collection, creating many non-standard local element names customized specifically for each collection. An option in these cases is to create customized local element names that clearly distinguish between which metadata values pertain to the original analog resource and which to the digitized version of the resource. These local elements are mapped to standard Dublin Core behind the scenes for interoperability, cross-collection searching, and exposure of the metadata for OAI harvesting.

Table 1 offers a hypothetical example that illustrates a somewhat common current practice in the CONTENTdm environment of using local field names mapped to Dublin Core elements, and mixing elements related to two manifestations in a single record.

TABLE 1: Single Record Mixing Elements for Original and Digital; Non-Standard Local Elements Mapped to DC.

Element Name	Dublin Core Mapping	Element Value
Title	Title	Menominee Reservation Sawmill
Photographer	Creator	Brown, David M.
Date of Photograph	Date [Created]	1910
City/Region	Coverage [Spatial]	Neopit
State/Province	Coverage [Spatial]	Wisconsin
Description	Description	Sawmill on a Menominee Indian reservation with numerous logs in the foreground.
Subject	Subject	Menomonie Indians
Subject	Subject	Sawmills--Wisconsin
Subject	Subject	Indian reservations--Wisconsin
Resource Type	Type	StillImage
Original Size	Format [Extent]	8.0 x 10 in.
Original Medium	Format [Medium]	Photographic prints
Original ID	Identifier	DMB-124-18-F7
Original Collection	Relation [isPartOf]	David M. Brown Photograph Collection
Original Repository	Contributor	Greenfield University Archives
Digital Copyright	Rights	Copyright 2008 Greenfield University Digital Library
Digital Collection	Relation [isPartOf]	Wisconsin Historical Images
Digital Publisher	Publisher	Greenfield University Digital Library
Digital File Format	Format	image/jpeg
Digital File ID	Identifier	WHI-02475
Date Digitized	Date [Created]	2010-04-07

This clearly violates the One-to-One Principle. But from a usability perspective, within the local closed-system context, we can ask to what extent this practice is likely to cause confusion or problems for either resource identification or retrieval for the majority of end users when searching and viewing the metadata. We should also keep in mind that in local systems using

◦DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010

CONTENTdm and similar software packages, and many others that do not, the metadata is displayed along with the digital object itself, especially digital images, and search results are also usually presented that way, thereby presenting users with further visual clues about the resource being described by each metadata record.

2. Problems for Resource Discovery

The failure to adhere to the One-to-One Principle in practice does, however, present several potential problems for resource discovery. One is that, even in the context described above, some users might still be confused about whether a metadata record is describing or representing a physical resource housed in a library, archive, or museum, or a digital resource that they can access online.

Another more significant problem for Dublin Core users is that the Dublin Core Element Set itself does not have a way to distinguish between metadata that applies to different manifestations of a resource when it appears within a single record. This is a problem for those who use "pure" Dublin Core without customized local elements. But it is also a problem for those who do use customized elements, but who almost certainly use standard Dublin Core as their base element set for searching across all of their collections. At that level, the customized element names are no longer present, and there is no way to distinguish between dates, creators, rights statements, and the like, that apply to the original versus the digital resources.

Perhaps most importantly, however, this practice presents significant potential problems when exposing only simple, unqualified Dublin Core metadata for harvesting via the OAI Protocol for Metadata Harvesting (OAI-PMH). This is probably the most commonly used method for harvesting and aggregating metadata today. Although the protocol is capable of harvesting richer metadata formats, it mandates the inclusion of a set of simple DC metadata records, and in practice many service providers harvest and aggregate only simple DC. Table 2 illustrates the metadata record from Table 1 after harvesting as simple Dublin Core, with the elements sorted alphabetically by DC element name to better illustrate the problem.

TABLE 2: Record After Harvesting As Simple Dublin Core.

Dublin Core Element	Element Value
Contributor	Greenfield University Archives
Coverage	Neopit
Coverage	Wisconsin
Creator	Brown, David M.
Date	1910
Date	2010-04-07
Description	Sawmill on a Menominee Indian reservation with numerous logs in the foreground.
Format	8.0 x 10 in.
Format	image/jpeg
Format	Photographic prints
Identifier	DMB-124-18-F7
Identifier	WHI-02475
Publisher	Greenfield University Digital Library
Relation	David M. Brown Photograph Collection
Relation	Wisconsin Historical Images
Rights	Copyright 2008 Greenfield University Digital Library
Subject	Indian reservations—Wisconsin
Subject	Menomonie Indians
Subject	Sawmills—Wisconsin
Title	Menominee Reservation Sawmill
Type	StillImage

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010

The result is multiple Date, Format, Identifier, and Relation elements in the same record, with no indication of what they apply to. But here also we can ask, when this type of metadata record is explicitly presented to users within the context of a specifically online *digital* resource collection, how likely are users to be confused about whether an individual record represents an analog object or a digital object? Taking the example in Table 2, how likely would users be to misunderstand that the values in some of the Format elements refer to a digital image file while others refer to an original physical object from which it was digitized? Even the two Relation elements, while it is no longer clear that one refers to a physical collection and the other to a digital collection, may not present terribly significant problems for most users. This assumption would need to be tested in order to be verified.

Multiple dates present much greater challenges, however, not only for resource identification, abut also especially because they are so often indexed and used as search limits and browse choices for information retrieval. Some metadata records include several kinds of dates, more than the two included in Table 2, and this presents a definite problem. Table 3 illustrates such a case.

TABLE 3: Multiple Dates Before and After Harvesting.

Local Record In CONTENTdm	
Date of Original Photograph	1914
Date Published in Newspaper	1964-05-06
Date Copyrighted	2005
Date Digitized	2005-11-30
Record After Harvesting	
Date	1914
Date	1964-05-06
Date	2005
Date	2005-11-30

3. Articulation and Conceptual Clarity of the One-to-One Principle

As practitioners work though issues of how best to represent attributes of analog and digitized versions of resources in their metadata, some further guidance may be helpful. Is the One-to-One Principle sufficiently well-articulated for most practitioners? Have its conceptual foundations been made sufficiently clear? A prime question is whether it possible to have a thoroughly-developed One-to-One Principle without making conceptual distinctions between intellectual/artistic content, on the one hand, and physical and digital carriers or manifestations, on the other hand. When attempting to work out One-to-One in practice, this distinction seems to inevitably arise. It is a distinction that has been made within the library cataloging community for some time, and is related to the long-standing "multiple versions" and "content versus carrier" discussions and developments. The four-tired FRBR conceptual model is perhaps the most significant outcome of these developments (FRBR).

The full, four-level FRBR model may not be maximally useful per se for most collections of digitized versions of unique, unpublished analog items, because there is usually no need to distinguish between work and expression or between manifestation and item. But a two-tiered model would be very useful, and arguably necessary. The FRBR term "manifestation" seems especially useful in this context.

Table 4 illustrates one way to distinguish between these aspects using the metadata from Table 1. The values listed under "Content" would seem to apply equally to the original and the digital manifestations of the resource.

TABLE 4: Content versus Carrier.

CONTENT		
Title	Title	Menominee Reservation Sawmill
Photographer	Creator	Brown, David M.
Date of Photograph	Date [Created]	1910
City/Region	Coverage [Spatial]	Neopit
State/Province	Coverage [Spatial]	Wisconsin
Description	Description	Sawmill on a Menominee Indian reservation with numerous logs in the foreground.
Subject	Subject	Menomonie Indians
Subject	Subject	Sawmills--Wisconsin
Subject	Subject	Indian reservations--Wisconsin
Resource Type	Type	StillImage
ANALOG CARRIER / MANIFESTATION		
Original Size	Format [Extent]	8.0 x 10 in.
Original Medium	Format [Medium]	Photographic prints
Original ID	Identifier	DMB-124-18-F7
Original Collection	Relation [isPartOf]	David M. Brown Photograph Collection
Original Repository	Contributor	Greenfield University Archives
DIGITAL CARRIER / MANIFESTATION		
Digital Copyright	Rights	Copyright 2008 Greenfield University Digital Library
Digital Collection	Relation [isPartOf]	Wisconsin Historical Images
Digital Publisher	Publisher	Greenfield University Digital Library
Digital File Format	Format	image/jpeg
Digital File ID	Identifier	WHI-02475
Date Digitized	Date [Created]	2010-04-07

Some questions arise if making this kind of distinction between content and carrier. First, if creating separate records for each manifestation, should all of the "content" values be included in both records? Do they apply to the original resource any more than to the digital resource? Second, does it make sense to say that the person who digitized the image is the Creator of the digital resource? Is the Dublin Core Creator element intended for this purpose, or is it intended to designate the person chiefly responsible for creating the intellectual/artistic content of the resource, whether in its analog or digital manifestation? Would that not be akin to saying that the Creator of the original photograph was the person who processed the negative and produced photographic print, but not the person who took the photograph (since they could be different)? The processor created the physical artifact, but is their role a significant one? Would this not likewise be equivalent to saying that the Creator of a printed book was the typesetter or the compositor or the person who pressed the button on the printing equipment to print the physical copies? This would seem to parallel the role of the person digitizing a book. We can also ask whether the person who digitized the text is of importance or interest to end users. Should that person really designated as the Creator in the metadata record? A clear distinction between intellectual/artistic content and analog/digital carrier could help resolve these questions.

Table 5 offers a different example, not an uncommon one, for digital image collections. In this instance, the object depicted in the image has its own creator and date of creation, making for two creators and three creation dates that could be included in the metadata record. In many cases such as this, the object depicted in the image also has Format characteristics of its own that may be included, such as the construction materials and dimensions of the building. In many such cases the photographer and the photograph as an object per se may be of little interest to end users, and there may be no need to include them in the metadata. In other cases, such as the hypothetical example in Table 5, the photographer is considered important, and the local analog collection even bears the name of the photographer. This further raises challenging issues about

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010

what "resource" is being described or represented by the metadata record, since the building, the photograph, and the digital image are all resources that have attributes represented in the metadata.

TABLE 5: Example of Three Related Resources, Creators, and Dates.

Element Name	Dublin Core Mapping	Element Value
Title / Name of Building	Title	Quadracci Pavilion, Milwaukee Art Museum, Milwaukee, Wisconsin
Architect	**Creator**	Calatrava, Santiago, 1951-
Date of Construction	**Date Created**	2001
Building Type	Subject	Galleries & museums
City/Region	Coverage Spatial	Milwaukee
State/Province	Coverage Spatial	Wisconsin
Description	Description	View of the Quadracci Pavilion pedestrian bridge and the Burke Brise Soleil over Windhover Hall
Resource Type	Type	StillImage
Photographer	**Creator**	Zajic, Tereza, 1964-
Date of Photograph	**Date Created**	2004-08-15
Original Size	Format Extent	35 mm.
Original Medium	Format Medium	Color slide
Original ID Number	Identifier	235-77c-56
Original Collection	Relation IsPartOf	Tereza Zajic Slide Collection
Original Repository	Contributor	Greenfield University Archives
Digital Copyright	Rights	Copyright 2008 University Digital Library
Digital Collection	Relation isPartOf	North American Architecture Collection
Digital Publisher	Publisher	Greenfield University Digital Library
Digital File Format	Format	image/jpeg
Digital File Number	Identifier	NAA004785
Date Digitized	**Date Created**	2008-07-15

4. Options for Maintaining the One-to-One Principle

This section of the paper briefly looks at three options that current practitioners could use to maintain a pure One-To-One principle in digital collection systems and notes some advantages and disadvantages of each option.

4.1. Option 1. Separate Linked Records

Table 6 illustrates the "classic" or "purest" form of application of the Principle: having separate records for each manifestation of the resource and linking them together.

TABLE 6: Option 1: Separate Linked Records.

One Record for the Original Photograph		
Title	Title	Menominee Reservation Sawmill
Photographer	Creator	Brown, David M.
Date of Photograph	Date [Created]	1910
City/Region	Coverage [Spatial]	Neopit
State/Province	Coverage [Spatial]	Wisconsin
Description	Description	Sawmill on a Menominee Indian reservation with numerous logs in the foreground.
Subject	Subject	Menomonie Indians
Subject	Subject	Sawmills--Wisconsin
Subject	Subject	Indian reservations--Wisconsin
Resource Type	Type	StillImage
Original Size	Format [Extent]	8.0 x 10 in.
Original Medium	Format [Medium]	Photographic prints

◈DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010

Original ID	Identifier	DMB-124-18-F7
Original Collection	Relation [isPartOf]	David M. Brown Photograph Collection
Original Repository	Contributor	Greenfield University Archives
Record ID	[Identifier]	**GUA-1111**
Digital Image	[Relation [hasFormat]]	**GUA-2222**
One Record for the Digital Image		
Title	Title	Menominee Reservation Sawmill
Digital Copyright	Rights	Copyright 2008 Greenfield University Digital Library
Digital Collection	Relation [isPartOf]	Wisconsin Historical Images
Digital Publisher	Publisher	Greenfield University Digital Library
Digital File Format	Format	image/jpeg
Digital File ID	Identifier	WHI-02475
Date Digitized	Date [Created]	2010-04-07
Record ID	Identifier	**GUA-2222**
Original Photograph	Relation [isFormatOf]	**GUA-1111**

A critical aspect of creating separate records for each manifestation is to include unambiguous, reciprocal links between the two records. The table illustrates one method: the use of record ID numbers. Perhaps another option could be to simply use the Original ID and Digital File ID identifiers as these reciprocal record links. These types of links function within the local context, but may break down in a consortial or third-party aggregator context because record numbers are usually not harvested, and neither these nor the local Identifiers can be guaranteed to be unique. Within the larger, anticipated linked data / semantic Web environment, practitioners would need to move to unique URIs in registered namespaces. As valuable as this practice is, and presumably will become even more so in the future, it is beyond the scope of large numbers of smaller and medium-sized institutions at the current time.

In the particular example in Table 6, the information relevant to the intellectual content of the resource has been included only in the record for the original object, with the exception of the title, which has been included in both. This follows a pattern established in VRA 4.0 (VRA 4.0 Examples). VRA strongly maintains the One-To-One principle, clearly distinguishing between *Works*, usually unique works of art owned by a museum, and *Images* of those works, be they staff documentary photographs, slides, or the like, or digital images, either digitized versions of the analog images or born-digital photographs of the Works (VRA 3.0 Introduction; VRA 4.0 Introduction). This practice largely seems to assume a local database context in which the records for the Work and one or more Images can be not only linked but also clearly collocated and displayed together in a single display for end users.

But this practice becomes problematic in contexts in which the two related sets of metadata cannot be meaningfully displayed together. This is often the case within local databases and user interfaces because of software limitations and other reasons. But it is also an issue for cross-collection searching and within harvested and aggregated collections. The question also arises as to whether, in these cases when there are to be two separate records, it could be best to include the full information about the intellectual content of the resource in both records.

4.2. Option 2. One Record for the Digital Resource, Noting the Original in a Source Element

The record in Table 7 illustrates a second option that would fully maintain the One-To-One Principle while having only one record in the database. This option would avoid potential problems with record linking and co-display, as noted above, and it would work well in systems that link a single metadata record to a single digital object.

◉DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010

TABLE 7: Option 2: Single Record for Digital Image; Information on Original Photograph in Source Element.

Record for Digital Image		
Title	Title	Menominee Reservation Sawmill
Photographer	Creator	Brown, David M.
Date of Photograph	Date [Created]	1910
City/Region	Coverage [Spatial]	Neopit
State/Province	Coverage [Spatial]	Wisconsin
Description	Description	Sawmill on a Menominee Indian reservation with numerous logs in the foreground.
Subject	Subject	Menomonie Indians
Subject	Subject	Sawmills--Wisconsin
Subject	Subject	Indian reservations--Wisconsin
Resource Type	Type	StillImage
Digital Copyright	Rights	Copyright 2008 Greenfield University Digital Library
Digital Collection	Relation [isPartOf]	Wisconsin Historical Images
Digital Publisher	Publisher	Greenfield University Digital Library
Digital File Format	Format	image/jpeg
Digital File ID	Identifier	WHI-02475
Date Digitized	Date [Created]	2010-04-07
Original Photograph	**Source**	8.0 x 10 in. photographic print. ID DMB-124-18-F7 in the David M. Brown Photograph Collection, located in the Greenfield University Archives.

In this option, the main "body" of the metadata record represents the digital resource, maintaining the logic that the resource being described in a database for a digital collection is for the *digital* object, not the original object. All of the information specific to the original analog object or manifestation has here been placed in a single Dublin Core Source element, where it clearly applies to a separate resource that is related to the one described in the body of the record. An alternative would be to use a DC Relation.isFormatOf element. There are, however, two problems with this approach.

This practice may be stretching the use of the Source element beyond what was envisioned by the DCMI. The preference for the Source and Relation elements seems to be for a value that can create a link between two related resources, whether a text string such as a title or numerical textual identifier or record number, on the one hand, or a unique URI, on the other hand. Nonetheless, Dublin Core was not originally designed for these kinds of digital collections, and the use of the DC elements in practice must sometimes be stretched to accommodate local information needs. This must of course be done without violating the meaning of the element as expressed in its official definition. This kind of use of the Source element, and of the Relation element as well, can be found in some institutional and consortial best practice guides, such as that of the Collaborative Digitization Program (CDP pp. 50-52 and 55-58).

The more substantial problem is that the information buried in the long text string in the Source element in this example cannot be separately processed for searching, limiting, browsing, collocating, hyperlinking, and so on. This is an information retrieval issue if an institution wants to offer users the ability to limit searches to photographic prints, for example, or to resources that were originally in the David M. Brown Photographic Collection. This is frequently an issue for dates as well, as in cases when a date applies to the original analog item alone, and this date is the date of primary interest to most end users for resource discovery.

4.3. Option 3. One Record for the Original Resource, Noting the Digital in a Relation Element

The record in Table 8 illustrates a third approach, which is simply the inverse of the second option illustrated in the previous example.

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010

TABLE 8: Option 3: Single Record for Original Photograph; Information on Digital Image in Relation Element.

Element Name	Dublin Core Mapping	Element Value
Title	Title	Menominee Reservation Sawmill
Photographer	Creator	Brown, David M.
Date of Photograph	Date [Created]	1910
City/Region	Coverage [Spatial]	Neopit
State/Province	Coverage [Spatial]	Wisconsin
Description	Description	Sawmill on a Menominee Indian reservation with numerous logs in the foreground.
Subject	Subject	Menomonie Indians
Subject	Subject	Sawmills--Wisconsin
Subject	Subject	Indian reservations--Wisconsin
Resource Type	Type	StillImage
Original Size	Format [Extent]	8.0 x 10 in.
Original Medium	Format [Medium]	Photographic prints
Original ID	Identifier	DMB-124-18-F7
Original Collection	Relation [isPartOf]	David M. Brown Photograph Collection
Original Repository	Contributor	Greenfield University Archives
Digital Image	**Relation [hasFormat]**	Digital image copyright 2008 Greenfield University Digital Library. Part of Wisconsin Historical Images online collection, published by Greenfield University Digital Library. Jpeg image file WHI-02475, digitzed 2010-04-07.

In this case the resource represented in the body of the record is the analog original, while the information about the digital manifestation is given in a single Relation element. Many of the same advantages and disadvantages apply to Option 3 as to Option 2. The primary advantage of this option is that the attributes of the original analog item are more often of interest and importance to end users than those of the digital object per se, and this approach allows these attributes to be processed for information retrieval because they reside in separate fields. But this means that the attributes specific to the digital object cannot be so processed. This includes digital file format and digital file size, often of importance for digital audio and moving image resources. An additional disadvantage is that this is approach runs counter to the logic that a database for a online collection of digital resources should contain records primarily representing the digital resources, not their original analog items.

4.4. The MODS <relatedItem> Element

At this point it can be instructive to look at a different metadata scheme / element set and see how it might handle the second or third options presented above. The Metadata Object Description Schema (MODS) includes a <relatedItem> element that has much greater potential information retrieval and identification power than do the Dublin Core Source or Relation elements. Table 9 takes the same metadata from previous examples and translates it into a MODS context, with a few additions allowed by the MODS scheme.

TABLE 9: MODS Record for Digital Image, with Information about the Original Photograph in <relatedItem> Element

MODS Element	Element Value
<titleInfo> <title>	Menominee Reservation Sawmill
<name type="personal">	
<namePart>	Brown, David M.
<role> <roleTerm type="text" authority="marcrelator">	photographer
<originInfo>	
<publisher>	Greenfield University Digital Library
<dateCreated>	2010-04-07
<physicalDescription>	
<internetMediaType>	image/jpeg

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010

<digitalOrigin>	reformatted digital
	Sawmill on a Menominee Indian reservation with numerous logs in the foreground.
<subject authority="lctgm"> <topic>	Menomonie Indians
<subject authority="lctgm"> <topic>	Sawmills--Wisconsin
<subject authority="lctgm"> <topic>	Indian reservations--Wisconsin
<subject authority="tgn"> <hierarchicalGeographic>	
<state>	Wisconsin
<city>	Neopit
<relatedItem type="host">	Wisconsin Historical Images
<identifier type="local">	WHI-02475
<accessCondition>	Copyright 2008 Greenfield University Digital Library
<relatedItem type="original">	
<titleInfo> <title>	Menominee Reservation Sawmill
<originInfo>	
<dateCreated>	1910
<physicalDescription>	
<form>	Photographic prints
<extent>	8.0 x 10 in.
<relatedItem type="host">	David M. Brown Photograph Collection
<location> <physicalLocation>	Greenfield University Archives

MODS has the capacity to include all of the attributes of both a digital and a related analog original resource in a single record while maintaining the One-To-One Principle. It can do this by using the <relatedItem> element, which allows every other MODS element to be included within it for a complete description of the related item. This means that it can break out each attribute into a separate element for computer processing, display, and information retrieval. Although this does not help in contexts that do not use MODS, this would seem to represent an elegant solution to maintain One-To-One in a single-record database while allowing all elements to be separately processed for retrieval and display.

In current practice, however, implementers in the Digital Library Federation Aquifer Initiative (DLF Aquifer) have found this to be too difficult for current metadata contributors, and have therefore made a pragmatic decision, considered necessary at the present time, of taking a "hybrid approach" of mixing values representing the digital and original resources in the main body of a single MODS record (Riley, et al., pp. 231-234). It is interesting that DLF Aquifer participants consist of large research libraries in the forefront of digital library development, yet even they in present practice do not adhere to the One-to-One Principle in their MODS records.

Still, the MODS <relatedItem> element shows potential promise for future systems and applications that may be able to fully implement it. This could include a cataloger-friendly metadata creation interface that helps metadata creators better understand the use of the element, clear and easily comprehensible user displays, and clearly-distinguished but full use of the main and related item metadata values for information retrieval.

4.5. Non-Standard Local Elements: Hidden Future Potential for One-To-One?

The point above about future system possibilities provides a segue into a brief digressive consideration relevant to One-to-One. In current implementations such as those using CONTENTdm, there is the potential, at least theoretically, to be able to map specific local elements to specific field and record combinations in a richer, standardized element set in the future. This is perhaps a hidden value in using non-standard customized local elements to label some values as relating to original versus digital manifestations of the resource represented in each metadata record. For example, a program could be written that could take all of the elements for a given collection that are designated as applying to the original resource and map them into a separate record from those elements designated as applying to the digital resource, and to automatically create a link between the two records. Or, a program could do the mapping into the

kind of MODS <relatedItem> structure shown in Table 9 above at such a time as systems could better process this structure. Or there might be yet new database structures and element set developments into which these elements could be machine mapped and massaged. This machine-mapping potential would not be possible if the metadata were not at present put into separate elements and explicitly labeled.

5. A Suggested "Compromise" Option

Turning from speculations about possible automated solutions in the future, what can current implementers do *now* to adhere as best as possible to One-to-One and create good quality sharable metadata? Taking into account the advantages and disadvantages of the three relatively "pure" options presented in Section 4 above, could there be a "best practice" recommendation for current metadata practitioners? It seems that there is no solution that maintains a pure One-to-One Principle without significant disadvantages. Of the three options given in Section 4, Option 2 emerges as the best method of maintaining One-to-One in a single-record implementation. It entails a single record representing the digital resource in the digital collection, with all of the information about the analog original in a DC Source element. The problem with this approach, and this is the critical problem that breaks its practical usability, is that some elements and data values of primary importance to users for searching, browsing, limiting, collocating, and navigating, cannot be so processed because they are buried in a long free-text description and are not separately marked for machine processing.

Given this situation, a modified version of Option 2 is suggested here as a possible "compromise" approach to addressing One-to-One issues in single record systems and simple DC harvesting environments. This suggested compromise option can be expressed in the following four instructions. (1) Follow Option 2 as much as possible while taking into account users' information retrieval needs. (2) Put into separate elements only those values applicable to the original manifestation that are considered necessary for processing for user resource discovery, such as searching, search limits, browsing, indexing, collocating, hyperlinking, and navigation. (3) Intentionally use the Source element to explain any ambiguities that arise from having multiple dates, names, identifiers, and the like, in the main body of the record, taking into account how the metadata will appear after harvesting in simple Dublin Core. (4) Consider which data elements and values are primarily administrative and are of little or no use to end users, and either omit them from the descriptive metadata record or suppress them from OAI harvesting. The record in Table 10 illustrates one possible application of this approach, taking the same metadata used in previous examples in this paper.

TABLE 10: Compromise Option Example A: Modified Approach to Example from Table 7.

Record for Digital Image		
Title	Title	Menominee Reservation Sawmill
Photographer	Creator	Brown, David M.
Date of Photograph	Date Created	1910
City/Region	Coverage Spatial	Neopit
State/Province	Coverage Spatial	Wisconsin
Description	Description	Sawmill on a Menominee Indian reservation with numerous logs in the foreground.
Subject	Subject	Menomonie Indians
Subject	Subject	Sawmills--Wisconsin
Subject	Subject	Indian reservations--Wisconsin
Resource Type	Type	StillImage
~~Original Size~~	~~Format Extent~~	~~35 mm.~~
~~Original ID Number~~	~~Identifier~~	~~235-77c-56~~
Original Medium	**Format Medium**	Photographic prints
Original Collection	**Relation isPartOf**	David M. Brown Photograph Collection
Original Repository	**Contributor**	Greenfield University Archives

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010

Digital Copyright	Rights	Copyright 2008 Greenfield University Digital Library
Digital Collection	Relation isPartOf	Wisconsin Historical Images
Digital Publisher	Publisher	Greenfield University Digital Library
Digital File Format	Format	image/jpeg
~~Digital File ID~~	~~Identifier~~	~~WHI-02475~~
~~Date Digitized~~	~~Date Created~~	~~2010-04-07~~
Original Photograph	**Source**	Digital reproduction of: 8.0 x 10 in. photographic print. ID DMB-124-18-F7 in the David M. Brown Photograph Collection, located in the Greenfield University Archives.

Strikethrough typography is used in Table 10 to indicate elements that can be omitted from the body of the record using this suggested compromise approach. All information applicable to the original manifestation is contained in a description-style Source element. Information about the original manifestation not needed for machine processing is not included in separate elements in the body of the record. Information about the original manifestation judged important for processing is retained in the body of the record, but its meaning is explained by the Source element. In this particular example, the institution has judged that it wants users to be able to search and browse by specific original format, such as "photographic prints." They also want users to be able to search and browse by original collection name and original repository name. The institution has also judged that the digital file ID and date of digitization are better treated as administrative metadata that may be retained in the descriptive metadata record but not included for OAI harvesting. The institution might also decide to make these administrative fields non-indexed and/or not displayed to users of the local system. After harvesting as simple Dublin Core, this record comes out as illustrated in Table 11.

TABLE 11: Modified Metadata from Table 10 After OAI Harvesting.

Record for Digital Image	
Title	Menominee Reservation Sawmill
Creator	Brown, David M.
Date	1910
Coverage	Neopit
Coverage	Wisconsin
Description	Sawmill on a Menominee Indian reservation with numerous logs in the foreground.
Subject	Menomonie Indians
Subject	Sawmills--Wisconsin
Subject	Indian reservations--Wisconsin
Type	StillImage
Format	Photographic prints
Relation	David M. Brown Photograph Collection
Contributor	Greenfield University Archives
Rights	Copyright 2008 Greenfield University Digital Library
Relation	Wisconsin Historical Images
Publisher	Greenfield University Digital Library
Format	image/jpeg
Source	Digital reproduction of: 8.0 x 10 in. photographic print. ID DMB-124-18-F7 in the David M. Brown Photograph Collection, located in the Greenfield University Archives.

Despite the fact that it does not adhere strictly to the One-To-One Principle, the metadata is likely to be relatively more sharable and usable in a harvested and aggregated context because there are fewer instances of the same element (such as Date) in the record, and when there are multiple instances (such as Relation), the content of the Source element disambiguates their meaning. This is even more evident in the example in Table 12, which provides a second illustration of this compromise approach, based on the record given previously in Table 5 of this paper.

TABLE 12: Compromise Option Example B: Modified Approach to Example from Table 5.

Element Name	Dublin Core Mapping	Element Value
Title / Name of Building	Title	Quadracci Pavilion, Milwaukee Art Museum, Milwaukee, Wisconsin
Architect	**Creator**	Calatrava, Santiago, 1951-
Date of Construction	**Date Created**	2001
Building Type	Subject	Galleries & museums
City/Region	Coverage Spatial	Milwaukee
State/Province	Coverage Spatial	Wisconsin
Description	Description	View of the Quadracci Pavilion pedestrian bridge and the Burke Brise Soleil over Windhover Hall
Resource Type	Type	StillImage
Photographer	**Creator**	Zajic, Tereza, 1964-
Date of Photograph	**Date Created**	2004-08-15
~~Original Size~~	~~Format Extent~~	~~35 mm.~~
~~Original Medium~~	~~Format Medium~~	~~Color slide~~
~~Original ID Number~~	~~Identifier~~	~~235-77c-56~~
~~Original Collection~~	~~Relation IsPartOf~~	~~Tereza Zajic Slide Collection~~
~~Original Repository~~	~~Contributor~~	~~Greenfield University Archives~~
Digital Copyright	Rights	Copyright 2008 University Digital Library
Digital Collection	Relation isPartOf	North American Architecture Collection
Digital Publisher	Publisher	Greenfield University Digital Library
Digital File Format	Format	image/jpeg
~~Digital File Number~~	~~Identifier~~	~~NAA004785~~
~~Date Digitized~~	~~Date Created~~	~~2008-07-15~~
Original Photograph	**Source**	Digital reproduction of original: 35 mm. color slide taken by Tereza Zajic on Aug. 8, 2004, Item 235-77c-56 in the Tereza Zajic Slide collection located in the Greenfield University Archives. Image of the Quadracci Pavilion, designed by architect Santiago Calatrava, construction completed in 2001.

This example is similar to the one given in Table 10, but it has an added layer of complexity: the implementers need to make statements for machine processing about the building depicted in the image, including its creator and date, as well as about the original photograph and the digital image of that photograph. It is quite understandable that the architect of the building depicted would be primary access points, and that in a collection of images focused on North American Architecture, users would want to browse and search by name of architect and date of construction. This example includes a scenario in which the photographer is of interest as well. The photographer may be locally, nationally, or internationally known, and the institution has a physical slide collection named for this photographer. In this case, the institution has judged that it does not need to separately process or display the original medium, collection, or repository name. All of this information is given in the Source element. The content of the Source element also disambiguates the two creator names and creation dates. After OAI harvesting as simple Dublin Core, the record comes out as illustrated in Table 13.

TABLE 13: Modified Metadata from Table 12 After OAI Harvesting.

Dublin Core Element	Element Value
Title	Quadracci Pavilion, Milwaukee Art Museum, Milwaukee, Wisconsin
Creator	Calatrava, Santiago, 1951-
Date	2001
Subject	Galleries & museums
Coverage	Milwaukee
Coverage	Wisconsin
Description	View of the Quadracci Pavilion pedestrian bridge and the Burke Brise Soleil over Windhover Hall

162

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010

Type	StillImage
Creator	Zajic, Tereza, 1964-
Date	2004-08-15
Rights	Copyright 2008 University Digital Library
Relation	North American Architecture Collection
Publisher	Greenfield University Digital Library
Format	image/jpeg
Source	Digital reproduction of original: 35 mm. color slide taken by Tereza Zajic on Aug. 8, 2004, Item 235-77c-56 in the Tereza Zajic Slide collection located in the Greenfield University Archives. Image of the Quadracci Pavilion, designed by architect Santiago Calatrava, construction completed in 2001.

This suggested compromise practice is obviously far from maintaining a "pure" One-to-One Principle and it is not particularly elegant from a "purist" perspective, but it does serve the dual purposes of acknowledging and maintaining the One-to-One Principle as much as possible, on the one hand, and serving the search, browse, navigation, and other information retrieval needs of users, on the other hand. It is not proposed as a definitive best practice, but is explored as one possible way of addressing a rather intractable problem in metadata practice.

6. Conclusion

If the application challenges of the One-to-One Principle are to be better met in the world of current cultural heritage metadata practice, at least two significant problems would seem to need to be addressed. First, large numbers of implementers need to be made aware of these issues, including the meaning and value of the One-to-One Principle in an OAI metadata sharing context, and the distinction between metadata elements that contain free-text description and elements that contain controlled values for processing.

Second, and related to the point above, there is a need for the weight of some kind of respected authority behind a set of general best practices that large numbers of implementers could be made aware of and have an investment in adopting. There are currently a plethora of institutional, regional, and consortial best practice documents and application profiles, but nothing on a broader national or international level intended specifically for the wide-spread community of practice creating collections of digitized resources in the types of systems described in this paper. Could there be some kind of relatively authoritative Dublin Core Best Practices for Digital Collections (or Digitized Resources, or Digital Reproductions, or the like) that comes from some respected group or organization? In this author's experience, there is a great deal of interest and desire for such guidance among current practitioners, especially those relatively new to digital collection projects, who want to create good quality and sharable metadata but usually do not know where to turn for such guidance.

Who would create and maintain such a best practices document? This may no longer be within the purview and scope of the DCMI. The focus of the DCMI seems to have moved on to Application Profiles and RDF/Linked Data/Semantic Web developments (DCMI Metadata Basics). The current "Using Dublin Core" usage guide (Hillmann, 2005) is excellent, but it is dated and does not seem for the most part to address the specific challenges of online collections of digitized resources, especially digitized images, made from unique analog originals. Potentially promising is a recent announcement of work on a revision of "Using Dublin Core" (Dekkers, 2010). Another promising development is work being done on a "'Best Practices' for CONTENTdm users creating shareable metadata" by the CONTENTdm Metadata Working Group facilitated by OCLC (Chapman, 2010). Yet, while some aspects the One-To-One problem as covered in this paper are tied to the particular characteristics of CONTENTdm and other similar systems, many aspects are equally true of other application contexts in which OAI service providers harvest and aggregate only simple Dublin Core. This can be the case, for example, with metadata originally created in a rich MODS structure but harvested only as simple Dublin Core.

DCPAPERS

Proc. Int'l Conf. on Dublin Core and Metadata Applications 2010

Could there be warrant for a full-fledged Dublin Core Application Profile (DCAP) for digital collections that follows the Singapore Framework (Nilsson, 2008) and conforms to the DCMI Abstract Model (Powell, 2007)? Can the DCAM's definition of a "description set" encompass descriptions of both a digitized resource and its analog original? A formal DCAP would develop a clear set of Functional Requirements and a Domain Model. Perhaps the most valuable component of such a DCAP for a majority of metadata practitioners would be a set of Usage Guidelines that would provide practical guidance on metadata creation for collections of digitized images and other resources.

The purpose of this paper has been to raise once again the long-standing topic of the challenges of applying the One-to-One Principle in practice. It has explored some reasons why this continues to be a problem in the world of metadata applications, how the principle itself might benefit from greater clarity, three options for wholly maintaining the principle in current systems, with advantages and disadvantages of each, and a possible compromise option for partially maintaining it so as to resolve the most critical deficiency in one of the purist options. It has concluded with an acknowledgement of the value to be gained from greater practical guidance in the form of best practice guidelines, or possibly a formal application profile, emanating from a widely acknowledged and respected source.

References

CDP (2006). CDP Metadata Working Group. Dublin Core metadata best practices, Version 2.1.1, September 2006. Retrieved April 12, 2010. http://www.bcr.org/dps/cdp/best/dublin-core-bp.pdf.

Chapman, John (2010). Promoting shareable metadata. Retrieved April 24, 2010. http://community.oclc.org/metalogue/archives/2010/04/promoting-shareable-metadata.html.

DCMI metadata basics. Retrieved July 15, 2010. http://dublincore.org/metadata-basics/.

Dekkers, Max (2010). Dublin Core Metadata Initiative - Status report March 2010. Retrieved July 2, 2010. http://dublincore.org/news/communications/statusreports/2010/03/

DLF Aquifer. Retrieved April 14, 2010. http://www.diglib.org/aquifer/index.htm.

FRBR (1998). Functional requirements for bibliographic records, Final report / IFLA Study Group on the Functional Requirements for Bibliographic Records. München : K.G. Saur. Current text available at http://www.ifla.org/publications/functional-requirements-for-bibliographic-records

Han, Myung-Ja, Christine Cho, Timothy W. Cole, Amy S. Jackson (2009). Metadata for special collections in CONTENTdm: How to improve interoperability of unique fields through OAI-PMH. Journal of Library Metadata, 9(3-4), 213-238.

Hillmann, Diane (2005). Using Dublin Core. Retrieved March 15, 2010. http://dublincore.org/documents/usageguide/.

Hutt, Arwen, and Jenn Riley (2005). Semantics and syntax of Dublin Core usage in Open Archives Initiative data providers of cultural heritage materials. In Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries, Denver, CO, June 7-11, 2005, 262-270. New York: ACM Press.

Nilsson, Mikael, Thomas Baker, Pete Johnston (2008). The Singapore Framework for Dublin Core Application Profiles. Retrieved July 15,2010. http://dublincore.org/documents/2008/01/14/singapore-framework/

Powell, Andy, Mikael Nilsson, Ambjörn Naeve, Pete Johnston, Thomas Baker (2007). DCMI Abstract Model. Retrieved April 12, 2010. http://dublincore.org/documents/abstract-model/.

Riley, Jenn, John Chapman, Sarah Shreeves, Laura Akerman, and William Landis (2008). Promoting shareability: Metadata activities of the DLF Aquifer Initiative. Journal of Library Metadata, 8(3), 221-248.

VRA 3.0 Introduction. Retrieved April 12, 2010. http://www.vraweb.org/projects/vracore3/index.html.

VRA 4.0 Examples. Retrieved April 7,2010. http://www.archivision.com/outgoing/vracore4/examplescat04.html.

VRA 4.0 Introduction. Retrieved April 12, 2010. http://www.vraweb.org/projects/vracore4/VRA_Core4_Intro.pdf.

Woodley, Mary (2005). DCMI glossary. Retrieved March 15, 2010. http://dublincore.org/documents/usageguide/glossary.shtml.