Developing cultural heritage preservation databases based on Dublin Core data elements

Fenella G. France
Art Preservation Services
Tel: +1 212 722 6300 x4
ff@apsnyc.net

Michael B. Toth
R.B. Toth Associates
Tel: +1 703 938 4499
mbt.rbtoth@starpower.net

Abstract:
Preserving our cultural heritage – from fragile historic textiles such as national flags to heavy and seemingly solid artifacts recovered from 9/11 – requires careful monitoring of the state of the artifact and environmental conditions. The standard for a web-accessible textile fiber database is established with Dublin Core elements to address the needs of textile conservation. Dynamic metadata and classification standards are also incorporated to allow flexibility in recording changing conditions and deterioration over the life of an object. Dublin core serves as the basis for data sets of information about the changing state of artifacts and environmental conditions. With common metadata standards, such as Dublin Core, this critical preservation knowledge can be utilized by a range of scientists and conservators to determine optimum conditions for slowing the rate of deterioration, as well as comparative use in the preservation of other artifacts.

Keywords:
Preservation, cultural heritage, Dublin Core, conservation, deterioration, environmental parameters, preservation vocabularies, textiles, fiber, metadata standards, digital library.

1. Introduction

Preserving our cultural heritage for future generations is a critical aspect of the stewardship of historical objects. This requires careful monitoring of the state of the artifact and environmental conditions, and standardized storage of this information to support a range of professionals who play a role in the preservation, storage and display of the artifact – including conservators, scientists, curators, and other cultural heritage professionals. Global access to preservation information can increase the utility of this information in the preservation of other artifacts. Making this information available across various domains requires common standards, which are supported by the Dublin Core metadata elements. Since artifacts inevitably deteriorate, these metadata elements must support a dynamic data set, with metadata and classification systems flexible enough to define and store information on the changing conditions and parameters. The primary goal of conservation is the preservation of cultural property, with preventive conservation focusing on minimizing intervention techniques if possible. An important consideration is whether stabilized conditions alone can confer enough of a benefit to offset the requirement to treat the historic cultural artifact. If treatment is required to remove harmful contaminants, the treatment should be assessed to ensure that it confers a

https://doi.org/10.23106/dcmi.952108550

benefit by removing soiling and particulates, and does no harm through fracturing, abrasion or reducing the object's mechanical integrity. Preventive conservation can be greatly enhanced by building on data from the exhibition, treatment and storage of similar objects.

Several preservation efforts under way include the development of standardized databases. At opposite ends of the size spectrum are the Fiber Reference Imaging Library (FRIL) and the World Trade Center (WTC) Archive database. To address the needs of textile conservation and research efforts worldwide, the textile community has begun planning for the establishment of a web-accessible textile fiber database. With standardized data elements based on the Dublin Core metadata elements, the FRIL database can provide important artifact information at both the macro and micro levels for a range of microscopic textile fibers. At the extreme macro level, the 9/11 WTC Archive contains rusting steel columns, large complex composite structures, crushed vehicles, and the Last Column, the final steel support column with large amounts of ephemera attached, such as paper memorabilia, photos, and personal items. Both databases contain data on objects that are fragile, both in terms of structural integrity and vulnerability from exposure to uncontrolled environments. Linking the changing structural status of objects to support the preservation of a range of artifacts requires useful defined metadata based on Dublin Core.

2. Standardizing the Evaluation of Deterioration

Useful deterioration information requires the accurate measurement and assessment of key object characteristics for effective storage and cross-domain transfer. Whether the object is a tiny fragile organic fiber or a huge seemingly robust metal beam, specific standardized attributes can be measured, imaged or recorded, and stored as data for access by users ranging from scientists and materials experts to conservators and curators. Whether the source artifact is the fabric of the Star Spangled Banner or the structure of the former Twin Towers, standardized data elements must be used to store identification and deterioration information. The development of a fiber reference library and a database of steel objects from the WTC is dependent on standardized metadata, and both are being developed with Dublin Core metadata elements. Kebbell et. al. note that "Institutions around the world are accumulating more and more digital objects. The task of managing these objects throughout their lifecycle, especially for institutions tasked to preserve them in perpetuity, becomes more complex the more deeply it is investigated." (1) Although museum standards such as AMICO are used to catalog museum objects, a common standard is required to record changes in the object and environmental conditions over time for cross-domain use by not only museum professionals, but also experts in other fields ranging from forensics to basic science (2). As Besser noted, "Dublin Core, developed as a core set of semantic elements for categorizing Web-based resources for easier search and retrieval, has become popular in the museum and education communities. The schema is deliberately simple, consisting of fifteen optional, repeatable data elements, designed to coexist with, and map to, other semantically and functionally richer metadata standards. Dublin Core's simplicity makes it an excellent medium of exchange, and thus a basis for interoperability." (3)

https://doi.org/10.23106/dcmi.952108550

The linkage of micro sampling object data with the "macro" of observations about the object itself is a critical part of analyzing and accurately assessing the object and its preservation requirements. Standardized metadata allows the tracking of dynamic changes associated with preservation conditions and risks from access, display, storage, and movement of the object.

Figure 1. Aged Wool Fiber Micro-fractures

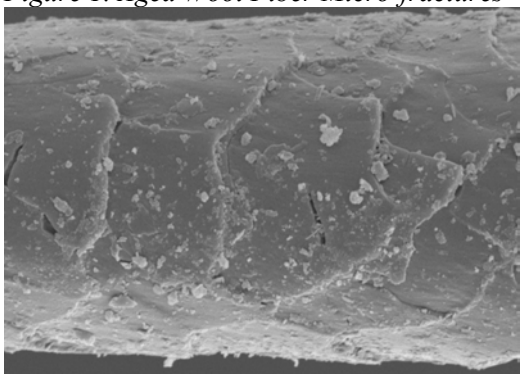


Figure 2. Fissure formation in WTC steel

Since all materials exhibit fracture surfaces indicative of deterioration, both fibers and structural steel can be monitored to assess degradation changes. Recording the results of accurate assessments and analysis of these changes using standardized metadata is critical for preservation. The use of a controlled vocabulary of deterioration is fundamental to assessing the mode of fracture, representing a fracture type associated with a cause (4).

3. The Fiber Reference Imaging Library (FRIL)

The mechanical properties of historic fibers reveal key information about the state of the textile and its degradation from environmental parameters and treatments. While subjective observations of historic textiles have traditionally been the basis for conservation evaluations, linking these assessments with objective analyses offers a more accurate reflection of the state of the textile.

3.1 Fiber Reference Image Standard

A range of organizations participating in the planning for FRIL have developed a metadata standard for a collaborative, digital, public domain, web-based textile fiber reference library. Data associated with specific textile fibers, including fiber images and associated metadata, will be accessible on a range of fibers for conservation treatment support, research, education and training by:

Curators
Conservators
Scientists (Conservation, Textile, Forensic, Anthropology)
Educators and Students
Preservation Professionals

The FRIL Standards Committee, represented a range of organizations and skills – with expertise in textile and forensic science, conservation, information systems, systems integration, education, and textile collection management – has developed the initial draft of a metadata standard based on the *Dublin Core Metadata Element Set,* combined with

https://doi.org/10.23106/dcmi.952108550

additional image and textile metadata elements. This standard has been tested with exemplar fiber images and metadata entries provided by conference participants. Based on the exemplar input and further review and discussion, the committee has proposed a final draft *Fiber Reference Image Library Metadata Standard* for review. The committee intends to host this standard for broader comment when a sponsoring organization is found.

3.2. Metadata Standard Elements

The Dublin Core Metadata Element Set provides the basic reference data sets with Categories 1, 2, 4 and 5 using Dublin Core as natural descriptors of the data or its basis, and Categories 3 and 6-8 indicating the specific extensions necessary to adequately address the requirements of dynamic preservation data. The FRIL standard includes the following metadata categories, based upon the Dublin Core metadata element set and specific ontology relevant to the image and preservation database (5).

1. *Identification Information*
   Basic information about the objects and images (the data set); domain must be in accordance with Dublin Core Metadata Element definitions.
2. *Sample Reference Information*
   Details about the original object from which the data set is derived.
3. *Imaging and Spectral Data Reference Information*
   Information describing the conditions used to acquire image objects.
4. *Data Type Information*
   Information about the technical format and protocols for the images and data elements.
5. *Object Description Information*
   Information about the textile and yarn from which the sample is derived.
6. *Fiber Information*
   Information about the fiber morphology and any changes or deterioration of the fiber sample.
7. *Metadata Reference Information*
   Information about the status of the metadata information, and the responsible party.
8. *Metadata Extensions*
   Additional specialized elements needed by the metadata producer or user.

Table 1 outlines the Dublin Core-based metadata element set used to define Category 1: The identification information in the FRIL database standard. This foundation information has served as the basis for the development of the preservation specific metadata categories, utilizing most of the fifteen optional, repeatable data elements. Elements from the DCMI element set such as: identifier, date, title, creator, subject, contributor, source, publisher and rights management are shown below. In this database, Dublin Core elements are intended to serve as the foundation for other metadata elements that offer additional semantics. Its design is intended to encourage the use of rich metadata schemes in combination with Dublin Core. This simplicity reduces the cost of creating metadata, and supports interoperability (6). As noted by Toth “System controls for an imaging database are those elements that play a critical role in guiding and defining the scope and content of the database, as well as the utility of the data and information (7).” The effectiveness of any preservation database is dependent upon good metadata to

https://doi.org/10.23106/dcmi.952108550

ensure the data elements 1) address the range of users' needs, and 2) encompass those observations critical to the preservation of the cultural objects.

Table 1. Identification Information

Identification Information

1. Data Set Identifier: a unique identification number; used to reference the image and its associated metadata.
CORE: YES TYPE: SINGLE
DOMAIN: TEXT and nnnnnnn

2. Date: date of creation of the fiber image in year, month, date.
CORE: YES TYPE: SINGLE
DOMAIN: year/month/date = nnnn/nn/nn

3. Date of Object: Historic date of creation of the textile from which the fiber was obtained, or the best approximate date.
CORE: YES TYPE: SINGLE or COMPOUND
DOMAIN: year = nnnn (CORE) month and/or date = nn and/or nn (if available)

4. Title: Name or description of textile object from which the fiber was obtained, using current collection catalog title, if available.
CORE: YES TYPE: SINGLE
DOMAIN: TEXT

4.1. Keywords: Description of textile object type from which the fiber was obtained, using standard nomenclature (with current catalog text if available).
CORE: YES TYPE: COMPOUND
DOMAIN: TEXT and DEFINED TEXT

5. Place of Origin: Location where textile from which the fiber was obtained was manufactured.
CORE: NO Type: COMPOUND
DOMAIN: DEFINED TEXT Country and/or region, and
TEXT state and other geographic detail

6. Creator Entity: Identification of entity that produced the textile from which the fiber was obtained.
CORE: NO Type: SINGLE
DOMAIN: "Individual"
"Manufacturer"
"Culture"
"Tribe / Clan"
"Unknown"

6.1. Creator: Name of maker of the textile from which the fiber was obtained
CORE: NO TYPE: SINGLE
DOMAIN: TEXT

7. Publisher: Organization sponsoring the image and associated data set
CORE: YES TYPE: SINGLE
DOMAIN: TEXT

8. Rights Management: Organization(s) holding the rights to the fiber image, associated data and images, and restrictions on data sharing
CORE: YES TYPE: COMPOUND
DOMAIN: "Public Domain"
"Link to Organization web site"
"Available upon request from rights holder"
"Not available"
"Copyright (rights holder)"

9. Contributor: Name and title of individual(s) performing fiber analysis, image collection or study
CORE: YES TYPE: SINGLE
DOMAIN: TEXT (Last Name, First Name)

10. Lineage: Acquisition and source data that contribute to the data set.
CORE: NO TYPE: SINGLE

https://doi.org/10.23106/dcmi.952108550

DOMAIN:	TEXT

3.3 Controlled Vocabulary

Integrating existing classification systems – in this case classification schemes for the textile item, fabric, yarn, and fiber types – is one of the challenges faced in defining deterioration and describing historic artifacts. Textiles are notoriously complex due to the wide range of substructures and base material – an item can be made from wool, silk, cotton, linen, metallic thread, or any combination of these fiber types.

Controlled vocabularies are used to describe content and represent concepts by assigning terms of one or more words as metadata associated with objects (8). A well-defined controlled vocabulary for each classification schema is imperative for a robust database. The retrieval of comparative data is heavily dependent on the quality of the underlying vocabulary and thesauri that support the standard. "The effectiveness of searching can be significantly enhanced through the existence of rich consistent metadata." (9). Metadata that has been well structured can facilitate the search for and integration of information, without compromising that quality of the resource. Linking the environmental conditions to the object also offers valuable conservation information about the state of the artifact. Table 2 gives examples of descriptors of fiber changes that can then be quantified to establish levels of deterioration. Clearly this is an area in which extensions of Dublin Core are needed to support specific artifact preservation metadata.

Table 2.Fiber Descriptors

Aspects of Fiber Deterioration
Fiber fractures
(Fibrillar, Smooth, Shear, Concave)
Surface soiling
(Particulates, Size, Amount)
Crushed / Flattening
Fibrillation
Broken distal ends

4. 9/11 World Trade Center Archive

The WTC Archive contains a diverse range of objects retrieved from Ground Zero during the rescue and clean-up operations. The archive is unique for its immense size and complex items including: large 30-60 ton rusting steel columns and building sections, 15-30 ton composites of a number of floors of compressed building materials and office contents, crushed emergency vehicles, and the photo and paper covered “Last Column.” These objects are both structurally and environmentally fragile.

https://doi.org/10.23106/dcmi.952108550

Figure 3. Distorted WTC Steel Beam

Figure 4. Vehicles in the WTC Archive

The various types of materials are categorized and prioritized for treatment, storage and display by *risk*. This is assessed based on the inherent characteristics and condition of the object, and preservation requirements in conjunction with storage and/or display conditions and environment.

4.1 Cross-Domain Standards

As with all museum collections, much of the WTC archive will remain in storage. This is a significant challenge for this collection due to the immense size and weight of so many unwieldy objects. A balanced and standardized set of metadata elements are needed to meet a range of requirements, including environmental conditions for storage and exhibition, as well as condition and display potential based on fragility and treatment. Preservation data can be an effective tool in determining the best environment for an item and reduce the level of risk. The WTC Archive is developing a metadata standard based on the *Dublin Core Metadata Element Set* to support a cross-section of users – from architects and structural engineers to curators, conservation scientists and conservators – who need this data.

While AMICO has established fields for cataloging objects, it does not allow the flexibility of Dublin Core to create the required metadata schema for specific ontology relevant to the preservation database. The difficulty in choosing the most appropriate standard is in selecting one with the flexibility to meet the broad needs of the archive. In this case, the metadata definitions must address the needs of the proposed 9/11 museum, while also meeting a range of preservation and loan requirements over the years. The use of metadata crosswalks (semantic mapping) will be critical for interoperability of the database with others, especially for data collected from multiple sources.

4.2 Standardization of Categorization

Each category of data includes more specific metadata about the risk associated with each piece, which then dictates the conservation and preservation requirements. This is depicted in Figure 5. Well-defined metadata for object location is critical for tracking and moving objects. Metadata is also needed for material not going to museums to adequately assess the possibilities for storage or display to meet the numerous requests for loans, based on the condition of the piece. Location categories include the following data about museum, indoor and outdoor environments:

Outdoor monuments

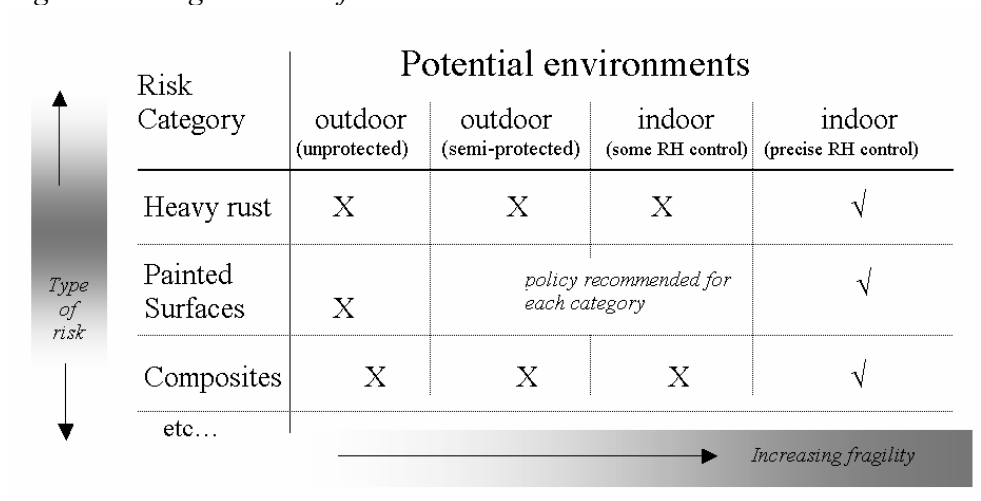
- open to weathering
- some weathering protection

https://doi.org/10.23106/dcmi.952108550

Indoor controlled environment
- some control
- precisely controlled environmental specifications

The risk elements are dependent upon regularly updated data from assessments of the changing state of the piece, and the potential availability, applicability and amount of research information about surface treatments. For example: Has the rust level stabilized or is it increasing? Have environmental changes increased the risk to the piece? Has the structural integrity decreased? Are significant markings at risk?

Figure 5. Categorization of Risk versus Environment



With the creation of microclimate areas for specific items in museum spaces, the level and control of relative humidity will be based on the fragility and risk of the object. Relative humidity fluctuations will increase the risk for the object because of its response to changes in moisture levels, which will cause expansion and contraction of various layers in the composite with decreased stability, and increased rust and detachment in such integral items as the Last Column and composites. This standardization of all metadata is also important for potential loan requirements. Any potential loan agreements would need to be based upon and include information regarding: analysis and stability of the piece, specifics of moving and transporting, relative humidity control, and the environment required for its preservation. With well-created metadata, the analysis and interpretation of the changing state of the artifacts will also modify what protective measures can be taken to increase their longevity, and the feasibility of treatment. Good metadata has been defined as:

appropriate to the materials in the collection, users and intended use
supporting interoperability
using content standards such as controlled vocabularies to describe the objects
requiring statement of conditions and terms of use of the object, including legal rights
supporting the long-term management of the collections
having good quality meta-metadata

https://doi.org/10.23106/dcmi.952108550

It has also been noted that indicators of "*goodness*" must emphasis factors contributing to interoperability, reusability, persistence, verification, documentation, and support for intellectual property rights (10).
The use of controlled vocabulary has been particularly problematic for the WTC database due to the large amounts of text currently used to describe disparate objects. Table 3 illustrates the need for clearly defining core and non-core data identifiers (as in Table 1), as well as a hierarchical schema for levels of object information. Non-standard text and metadata elements will greatly reduce the utility of this data for searches, data management and access. Standards for searchable metadata are now being developed with definitions based on Dublin Core Data Elements to support a database for a broad range of users to ensure the preservation, storage and display of critical historic objects marking the events of 9/11 2001.

Table 3. Example of Complex Object Descriptors for WTC Archive Database

Description	Material	Notes
Column	Steel	Due to axial overloading, this column bent into a "U" shape. A built-up steel section, no buckling of steel on compression side of the bend, one tear on tension side of the bend.
North Tower Antenna	Steel	This part of the antenna is the highest accessible point of the North Tower. A steel pipe stair ladder extended through the interior of the antenna and ended at a hatch at the top of the 3.5 ft. x 3.5 ft. platform 325 ft. above roof.
Structural Triad	Steel	North Tower façade structure. Located at an elevation of approx. 70 ft above the concourse level. Sections of stainless steel mullions attached to the sides of the column.
Exterior Column with Spandrel	Steel	North Tower façade structure. Located at an elevation approx. btwn first and second floors. The following text is painted on the steel: "SAVE" and "5 NORTH" in orange spray paint.
Beam	Steel	Steel beam with connection plates at one end. The other end has been cut for removal. The beam has folded near its mid-point. No visible identification markings on the steel.
Beam	Steel	Beam has been slightly bent. At one end, the web and one flange have been cut away. No visible identification markings on the steel.
Tower Exterior Wall Steel	Steel	One of three columns that were part of a typical tower exterior wall prefab panel. Both ends of the column have been cut for removal. No visible identification markings on steel.

5. Conclusion

Dublin Core metadata elements can serve as critical elements in the foundation of preservation databases. The value and utility of preservation databases relies solely on the effective creation, maintenance and management of standardized metadata elements. While deterioration is inevitable for cultural heritage objects, it can be mitigated and controlled through careful analysis and monitoring of artifact conditions and environmental parameters. The database requirements of both the FRIL and the WTC archives demonstrate the need for precise, well-defined and carefully managed metadata to support useful analyses with a preservation database. Preservation metadata supports a range of object requirements including:

- Control of the environment
- Storage and movement
- Exhibition and display
- Loan requirements

The development of these two databases is a challenging and ongoing process. While other conservation object identification databases exist, there are no other current efforts

https://doi.org/10.23106/dcmi.952108550

to develop and establish deterioration data, so critical to cultural preservation. The creators of these databases recognize that with development work still underway, standardized metadata elements and vocabularies are essential to creating a workably and interoperable worldwide preservation resource, as new ground continues to be broken in this area of artifact conservation.

The ease of web accessibility allows participants in international collaboration efforts to gain global access to preservation information. This ensures that artifacts of similar materials, time periods, and conditions can be studied through a comparative database of images and associated data to allow informed preservation and conservation decisions to be made. With these projects, the main issues have been establishing an extensive controlled vocabulary to allow ease of analysis and integration of preservation data, and adapting this to amend and accurately document changes in the state of the artifact and environmental information. The continued development and refinement of Dublin Core elements for the WTC archive will be critical to the utilization and management of this large and complex database. While library and general museum database schema tend to be used for cataloging purposes, the expanded focus of preservation databases requires the integration of dynamic environmental preservation information. This much richer and complex system effectively capitalizes on the flexibility of the *Dublin Core Metadata Element Set*. These principles all point strongly to the use of Dublin Core in terms of its ability to be adapted to the needs of museums and preservation databases.

Establishing preservation databases to store images and textual information about deterioration rates and processes requires common information standards, such as Dublin Core, for the storage, management, access and sharing of critical preservation information. With these standards we can ensure the effective stewardship of our cultural heritage for future generations.

References:

1. A. Kebbell and D. Campbell. Managing digital objects and their metadata: challenges and responses. http://www.slais.ubc.ca/PEOPLE/faculty/tennisp/dcpapers2004/Paper_09.pdf DC_2005 Conference, Madrid, 12-15 September 2005.
2. AMICO Data Specification: Data Dictionary (Version 1.3) http://www.amico.org/AMICOlibrary/dataDictionary.html Art Museum Image Consortium, 23 December 2004.
3. H. Besser. Selecting a Metadata Schema. *Introduction to Imaging*. USA: Getty Information Institute, 2003, pp. 47-53.
4. F.G. France. Creating a standard vocabulary for defining levels of deterioration. *Development of a Web-Accessible Reference Library of Deteriorated Fibers.* USA: National Park Service, 2003, pp. 77-86.
5. DCMI Metadata Terms, http://www.dublincore.org/documents/dcmi-terms/ DCMI Usage Board, June 2005.
6. ANSI/NISO Z39.85-2001. The Dublin Core Metadata Element Set, NISO, Sep 2001.
7. M.B. Toth. Textile Database Systems Integration. *Development of a Web-Accessible Reference Library of Deteriorated Fibers.* USA: National Park Service, 2003, pp 121-127.

https://doi.org/10.23106/dcmi.952108550

8. Guidelines for the Construction, Format and Management of Monolingual Controlled Vocabularies http://www.niso.org/standards/resources/Z39-19-2005.pdf National Information Standards Organization, 25 July 2005.
9. A. Gilliland-Swetland. Defining Meta data. *Introduction to Metadata, Pathways to Digital Information*. USA: Getty Information Institute, 1998, pp.1-8.
10. NISO Framework Advisory Group, A Framework of Guidance for Building Good Digital Collections, 2nd Edition, NISO Press, Maryland, USA, 2004.

https://doi.org/10.23106/dcmi.952108550