Applying DCMI Elements to Digital Images and Text in the Archimedes Palimpsest Program

Michael B. Toth
Walters Art Museum –
R.B. Toth Associates, USA
mbt.rbtoth@gmail.com

Doug Emery
Walters Art Museum –
Emery IT, USA
doug@emeryit.com

Abstract

The digitized version of the only extant copy of Archimedes' key mathematical and scientific works contains over 6,500 images and 130 pages of transcriptions. Metadata is essential for managing, integrating and accessing these digital resources in the Web 2.0 environment. The Dublin Core Metadata Element Set meets many of our needs. It offers the needed flexibility and applicability to a variety of data sets containing different texts and images in a dynamic technical environment. The program team has continued to refine its data dictionary and elements based on the Dublin Core standard and feedback from the Dublin Core community since the 2006 Dublin Core Conference. This presentation cites the application and utility of the DCMI Standards during the final phase of this decade-long program. Since the 2006 conference, the amount of data has grown tenfold with new imaging techniques. Use of the DCMI Standards for integration across digital images and transcriptions will allow the hosting and integration of this data set and other cultural works across service providers, libraries and cultural institutions.

Keywords: Dublin Core; metadata standards; archiving; imaging; manuscript; Archimedes Palimpsest; cultural heritage; digital library

1. Introduction

Effective metadata standards are required to efficiently handle the large amounts of data collected in imaging and scholarly studies of the earliest known copy of Archimedes' work. The *Dublin Core Metadata Element Set* is being utilized to provide key identification information, with additional metadata extensions to ensure the imaging and scholarly information can be readily integrated in a Web 2.0 environment. Applying the Dublin Core Metadata Initiative (DCMI) Metadata Element Set and additional elements from the DCTerms namespace to a variety of images containing different texts in a rapidly changing technology environment has posed a unique set of challenges. These challenges include linking together and integrating data from different sources and formats: Digital images from advanced cameras in numerous spectral bands, and digitally encoded texts in varied fonts from a team of scholars. With increased focus on data management and explosive growth in data with advanced imaging techniques, the application of the *DCMI Metadata Element Set* provides a robust data set that will meet worldwide metadata standards.

2. Archimedes Palimpsest Program

The Archimedes Palimpsest Program is a 10-year effort to produce digital images of Archimedes' text as originally written on parchment in the latter half of the tenth century. In the early thirteenth century, this text was scraped off and overwritten, or "palimpsested," to create a prayer book. A team of scientists and scholars has been digitally imaging and studying the 174 parchment leaves that currently make up the Archimedes Palimpsest. Since the 2006 Dublin Core Conference the program has developed new imaging techniques that have yielded over two terabytes of data. This includes images of the only copies of Archimedes treatises *The Method* and *Stomachion*; the only copy in Greek of *On Floating Bodies;* and copies of the *Equilibrium of Planes*, *Spiral Lines*, *The Measurement of the Circle*, and *Sphere and Cylinder*. Imaging has also

revealed ten leaves of text by the fourth century B.C. Greek orator Hyperides; six leaves of commentaries on Aristotle; four liturgical leaves; and twelve leaves from two unidentified books.

2.1. Imaging

At the time of the 2006 Dublin Core Conference, the imaging team had imaged the entire palimpsest with three spectral bands of light, yielding three images for each leaf and processed "pseudocolor" images in what was then considered to be a large data set of about 240 Gigabytes of data. Since the 2006 conference, the imaging team developed new imaging techniques to yield more information with more advanced cameras and lighting in 12 spectral bands. These yielded 16 images of each leaf and more refined processed images with a total of about 2,400 GB of data. (See Figure 1) Managing all this data required careful metadata logging and data management based on the *Dublin Core Metadata Element Set*.
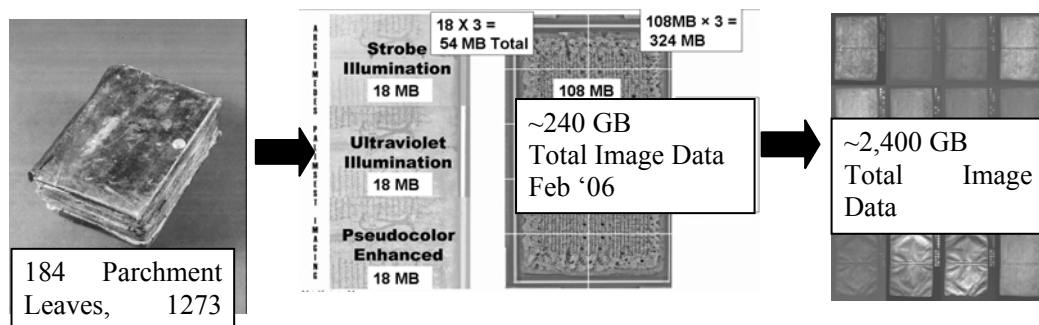

Fig. 1. Archimedes Palimpsest Data Growth © Images Copyright Owner of the Archimedes Palimpsest.

The Archimedes Palimpsest team also created images of key leaves at the Stanford Synchrotron Radiation Laboratory using X-ray fluorescence. This required an extensive range of metadata extensions to capture the broad range of metadata on energy levels and system parameters. They also imaged original prints of photographs of the Archimedes Palimpsest taken almost 100 years earlier at the direction of John Ludwig Heiberg in Constantinople, and photographs of one leaf taken in Chicago in the 1930's. These images of the photographs offered standardized images of text that has since been lost, and one leaf that has been lost in its entirety.

2.2. Metadata

With 6,797 digital images and 130 pages of transcriptions of the Archimedes Palimpsest, metadata has proved to be essential for 1) accessing images and integrating spectral bands for digital processing and enhancement, 2) managing transcriptions from those images for study by scholars around the world, and 3) linking and integrating the images and the transcriptions. This work required extensive identification metadata to ensure the data was manageable, as well as spatial metadata to line up and register the various images.

The *Dublin Core Metadata Element Set* offers the key identification elements required for image storage, management and retrieval, with additional spatial and spectral information added as extensions. The Archimedes Palimpsest Metadata Standard incorporates the DCMI Standards with six types of metadata elements:

1. Identification Information
2. Spatial Data Reference Information
3. Imaging and Spectral Data Reference Information
4. Data Type Information
5. Data Content Information
6. Metadata Reference Information

The "Identification," "Data Type" and "Data Content" metadata elements incorporate the *Dublin Core Metadata Element Set*. The "Spatial Data Reference" and "Imaging and Spectral Data Reference" elements are extensions to the DCMI Standards, using metadata elements detailed in the Federal Geographic Data Committee *Content Standard for Digital Geospatial Metadata*. The standard is hosted on the www.archimedespalimpsest.org website.

In keeping with the project goals of long term data and metadata accessibilty, the program follows the DCMI's principble of simplicity. We have created records that are machine-readable with very little effort and easily intelligible by a human reader. Each image metadata record is a series of simple name-value pairs, employing Dublin Core and project-specific metadata elements (See Table 1).

TABLE 1. Image Metadata Elements

Identifier	60000
Date	2008-03-03T08:20:56-05:00
Creator	Christens-Barry, Bill
Creator	Easton, Roger
Creator	Knox, Keith
Subject	Euchologion Image
Subject	Archimedes Palimpsest Image
Subject	Palimpsest Image
Subject	Multispectral Image
Subject	Digital Image
Subject	Greek Manuscript Image
Subject	Byzantine Manuscript Image
Subject	Private Collection
Publisher	Owner of the Archimedes Palimpsest
Contributor	Noel, Will
Contributor	Toth, Michael
Contributor	Auer, Kevin
Contributor	Emery, Doug
Contributor	Gerry, Kate
Contributor	Potter, Daniel
Contributor	Quandt, Abigail
Contributor	Tabritha, Ariel
Contributor	Tilghman, Ben
Contributor	Stokes, John R.
Type	Image
Source	Processed from image with Identifier 15380,0000100r_Arch53v_ Sinar_LED445 _01_raw.tif
Source	Processed from image with Identifier 15383, 0000-100r_Arch53v_Sinar_LED 530_01_raw.tif
Source	Processed from image with Identifier 15386, 0000-100r_Arch53v_Sinar_LED 625_01_raw.tif
Coverage	Walters Art Museum
Coverage	2007-08-06 to 2007-08-26
Coverage	Baltimore, MD
Coverage	USA
license	http://creativecommons.org/licenses/by/3.0/legalcode
license	Licensed for use under Creative Commons Attribution 3.0 Unported
accessRights	Copies of any articles published must be sent to William Noel, Walters Art Museum, Baltimore, MD.
ID_File_Name	0000-100r_Arch53v_Sinar_true_pack8.tif

Since the 2006 Dublin Core Conference, individuals with standards experience in OCLC and other organizations have provided input on the best application of the Dublin Core Standard to the Archimedes Palimpsest Metadata Standard. A range of organizations have also provided guidance on the use of standards for archival purposes, including guidance for the best use of standards in the digital data set to ensure users years hence will have access to the actual versions of the standards used in creating the data set. Input from the Library of Congress, the British Library, NASA and Google proved fruitful in defining the application of standards not only to the Web 2.0 environment, but the range of possible digital environments possible in decades to come.

2.3. Transcriptions

Scholars have been transcribing the Greek text since the initial digital imaging, revealing new information about the origins of mathematical theories and science. The integration of these scholarly transcriptions in digital form with the digital images has taken on greater impetus since the 2006 Dublin Core Conference, with the digital tagging and encoding of text in various forms and formats, including handwritten, MSWord Symbol font, and various other custom fonts. A team of scholars and students is encoding the transcribed text into XML tagged Unicode following the Text Encoding Initiative standards (See Figure2).

```
<seg TEIform="seg" n="17v1" part="N" type="folio">
   <seg TEIform="seg" n="1" part="N" type="line">
      <supplied TEIform="supplied" reason="lost">
         <expan TEIform="expan">ὅτι</expan> τὸ ΦΑ</supplied> μέγε<supplied
         TEIform="supplied" reason="lost">θος</supplied> τῶι βάρει πρὸς </seg>
   <seg TEIform="seg" n="2" part="N" type="line">
      <supplied TEIform="supplied" reason="lost">τὸ ὑγρ</supplied>ὸν τὸ ἰσόογκον
      τοῦτον ἔχει</seg>
   <seg TEIform="seg" n="3" part="N" type="line">τὸν λόγον, ὃν τὸ Α <expan
         TEIform="expan">πρὸς</expan> τὸ Φ<unclear TEIform="unclear"
         >Α</unclear>.</seg>
</seg>
```

FIG. 2. XRF Tagged Transcriptions

Header information is provided for each folio in the encoded text, with cross-walked Dublin Core Identification and Data Content metadata elements mapped to the TEI format (Figure 3). These encoded texts are then hosted with the images, with the Dublin Core elements providing a common structure for image and transcription metadata.

```
<teiHeader>
   <fileDesc>
      <titleStmt>
         <title>Transcription of fols. 17v-16r of the Archimedes Palimpsest
               (= Archimedes fol. 7v, On Floating Bodies)</title>
         <respStmt>
           <resp>Responsible for primary transcription (Dublin Core creator)</resp>
            <name>Reviel Netz</name>
         </respStmt>
         <respStmt>
            <resp>Contributor</resp>
            <name>Mike Toth</name>
         </respStmt>
<publicationStmt>
         <idno>5021</idno>
         <publisher>Owner of the Archimedes Palimpsest</publisher>
         <date>2008</date>
</publicationStmt>
   </fileDesc>
   <profileDesc>
      <langUsage>
        <language id="grc-c">accented ancient Greek in Unicode-C Greek
characters</language>
      </langUsage>
```

FIG 3. Sample Dublin Core Header Information in Encoded Transcription Headers.

3. Integrated Product

The images and transcriptions are linked through metadata in the Archimedes Palimpsest Data Product, enabling common searches, access and study. The standard use of the Dublin Core Metadata Element Set across the products of the image scientists and scholars enables linkage between these two disparate data sets for further study (Figure 2). Integrating metadata of various types tailored to meet a range of users' needs has proven critical to making integrated data available across domains and disciplines amidst ever changing technologies. Building on the Archimedes Palimpsest application, the DCMI Standards are being used to integrate hyperspectral imaging of the Waldseemuller 1507 Map at the Library of Congress. The DCMI Standards serve as the basis for information discovery in the Web 2.0 environment, and hopefully for decades to come in future formats and technologies. This information will advance the study of the original manuscript by individuals around the world with ubiquitous access via the Internet.
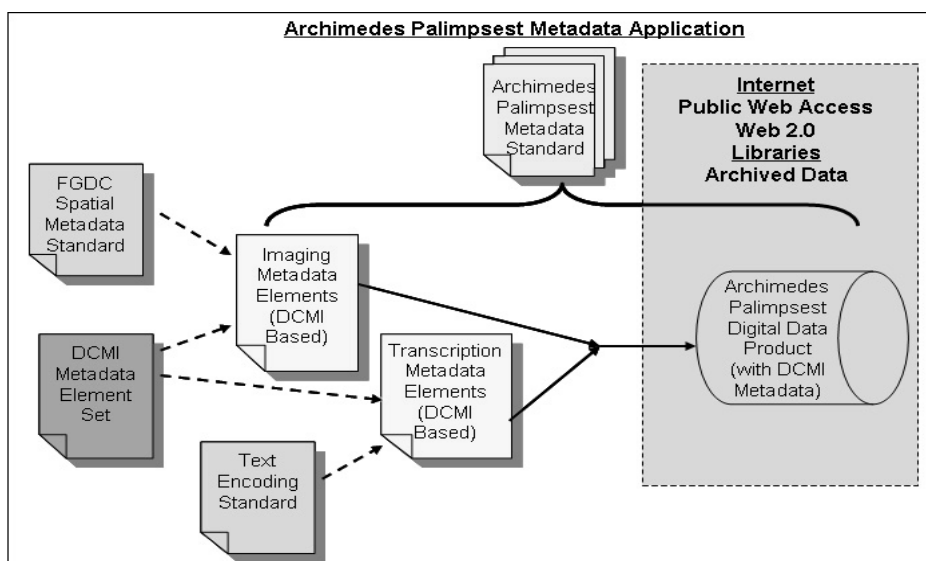

Fig. 4. Archimedes Palimpsest Metadata Application Architecture.

DCPAPERS

2008 Proc. Int'l Conf. on Dublin Core and Metadata Applications

Acknowledgements

References

Archimedes Palimpsest Program. (2006). *Archimedes Palimpsest Metadata Standard 1.0, Revision 5.* Baltimore, Maryland: Walters Art Museum (WAM).

Dublin Core Metadata Initiative. (2000-2008). *Dublin Core Metadata Element Set, Version 1.1: Reference Description*. Retrieved from http://dublincore.org/documents/dces/.

Dublin Core Metadata Initiative. (2000-2008). *DCMI Metadata Terms*. Retrieved from http://www.dublincore.org/documents/dcmi-terms/.

Federal Geospatial Data Committee. (2002). *Content Standard for Digital Geospatial Metadata: Extensions for Remote Sensing Metadata.* FGDC-STD-012-2002**.** Washington, DC: Federal Geospatial Data Committee.

Knox, Keith T., Roger L. Easton Jr., and William A. Christens-Barry. (2003). Multispectral imaging of the Archimedes Palimpsest. *2003 AMOS Conference.* Maui, Hawaii: Air Force Maui Optical & Supercomputing Site.

Netz, Reviel. (2000). The origin of mathematical physics: New light on an old question. *Physics Today*, 32-37.

Noel, Will, Roger L. Easton Jr., and Michael B. Toth. (2006). *The Archimedes Palimpsest.* California: Google Inc. Retrieved, March 7, 2008, from http://www.youtube.com/watch?v=S19Xyjxl4fI.

Toth, Michael B., William A. Christens-Barry, and Roger L. Easton Jr. (2006). Dublin Core based metadata supports the Archimedes Palimpsest Manuscript Imaging Program. *International Conference on Dublin Core and Metadata Applications, Colima, Mexico, October 3-6, 2006.*

WAM. (2008). *Archimedes - The Palimpsest*. Retrieved, March 20, 2008, from http://www.archimedespalimpsest.org/.