Eureka! Dublin Core Based Metadata Supports the Archimedes Palimpsest Manuscript Imaging Program

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

William A. Christens-Barry
Equipoise Imaging
Tel: +1 410 750 6656
Fax: +1 410 750 6158
equipoise1@verizon.net

Roger L. Easton, Jr.
Chester F. Carlson Center for Imaging Science, Rochester Institute of Technology
Tel: +1 585 475 5969
rlepci@cis.rit.edu

Abstract:
Digital imaging of the Archimedes Palimpsest offers a complex set of metadata challenges. The thousand-year-old manuscript contains the earliest known copies of some of Archimedes unique mathematical works, overwritten with a book of prayer. Digital imaging of this manuscript is yielding a large and rich volume of data. The Archimedes Palimpsest team has developed the *Archimedes Palimpsest Metadata Standard* using the *Dublin Core Metadata Element Set* as the core metadata for this effort. The need for spatial metadata to support a complex range of image data has required the integration of metadata standards based on those originally developed for geospatial imaging from space. The result is a unique manuscript imaging standard to support digital "scriptospatial" data accessible by a range of users on the Internet as part of this dynamic, ongoing program.

Keywords:
Dublin Core, metadata standards, spatial metadata, imaging, manuscript, Archimedes Palimpsest, accessibility, cultural heritage, digital library, museum, DC2006

1. Introduction

Effective metadata standards are required to efficiently handle the hundreds of gigabytes of data that make up thousands of individual and stitched digital images of the Archimedes Palimpsest taken in visible and ultraviolet light and multiple X-Ray Fluorescence elements, as well as the resulting processed images. The *Dublin Core Metadata Element Set* is being utilized in this ongoing program to provide key identification and data information, with additional metadata elements incorporated to ensure the spatial, imaging and spectral information can be readily stored, managed and accessed. Applying the Dublin core

https://doi.org/10.23106/dcmi.952108455

metadata standards to a variety of images containing different texts and integrating additional spatial standards to the effort has offered a unique set of challenges.

2. Archimedes Palimpsest Program

The Archimedes Palimpsest Program is an ongoing, multi-year effort to produce digital images of the Greek text of seven treatises of Archimedes, 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," with Greek text to create a prayer book [1]. Beginning in 1999, under the auspices of the Walters Art Museum in Baltimore, Maryland, USA, a team of scientists, conservators and scholars has been disbinding, conserving, imaging, analyzing, transcribing, translating and studying the 174 fragile parchment folios that make up the Archimedes Palimpsest. Using a range of imaging techniques, the imaging team has produced over 400 Gigabytes (GB) of digital data, yielding images of unique diagrams of Archimedes, and the only copies of Archimedes' treatises *The Method* and *Stomachion*, the only copy in Greek of *On Floating Bodies,* and copies of the *Planes in Equilibrium*, *On Sphere and Cylinder, Spiral Lines*, and *Measurement of the Circle* [2]. They have also imaged ten folios of text by the fourth-century B.C. Attic Greek orator Hyperides; six folios from a Neo-Platonic philosophical text; four folios from a liturgical book; and twelve folios from two other books, the text of which has yet to be deciphered [3].

In an ironic technical complement, the Archimedes Palimpsest team also was able to image prints of original photographs of the Archimedes Palimpsest taken almost 100 years earlier in Constantinople at the direction of Johann Ludwig Heiberg. These photographs offer standardized images of some text that has since been lost to mold or other damage, and of one leaf that has been lost in its entirety. They also posed an interesting metadata contextual challenge because they refer not to the original object, but to data derived from the object many years prior to the current digital imaging task.

The advanced imaging effort required to effectively tap the unique mathematical information hidden in the Archimedes Palimpsest has required support over the years from a multidisciplinary team. Conserving the parchment, imaging the manuscript in specific spectral bands and at various energy levels, storing the digital data in a standardized format, processing the data to yield useful information, and displaying the available information for academic researchers and the public has required the integration of a wide range of tasks and technical capabilities. This requires effective integration of Dublin Core metadata elements with additional metadata elements to meet the needs of a range of disciplines across domains in this ongoing effort.

2.1. Imaging Program

Central to the collection of information from the original Archimedes text is the multispectral digital imaging of the ancient manuscript. To read the underlying ancient Greek text that was overwritten orthogonally with Greek prayer text called the "Euchologian" in similar iron gall ink, the imaging scientists identified different spectral signatures between the two inks, as well as the underlying parchment. The Red/Green/Blue (RGB) images were taken under Ultraviolet (UV) and visible illumination to produce the spectral differences needed to digitally reveal the underlying Archimedes' text (Fig. 1) [4].

https://doi.org/10.23106/dcmi.952108455

The original images were taken in all spectral bands, and then digitally processed to produce false color, or "pseudocolor" images of the underlying Archimedes text. The images were taken with a spatial resolution of 600 dpi, which required ten sectional images of each page, with each image stored as an 18 Megabyte (MB) Tagged Image File Format (TIFF) file. The images were then digitally stitched to create full-size images of the manuscript leaf, with a file size of approximately 108 MB.

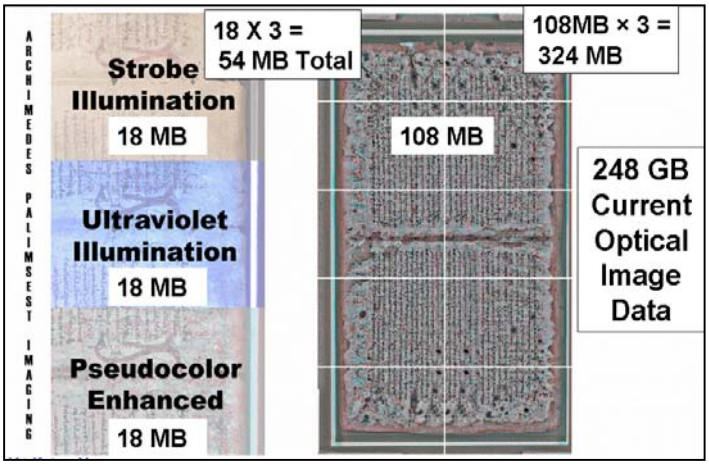

Figure 1 – Archimedes Imaging and Data Holdings

2.2. X-Ray Fluorescence Imaging

Four of the palimpsested manuscript leaves are also overlaid with twentieth-century forged paintings of the type commonly found in illuminated manuscripts, and the original text on other leaves was obscured by mold or excessive palimpsesting techniques. A combined team of imaging scientists and X-Ray physicists conducted X-Ray Fluorescence (XRF) studies of these leaves. The studies of these few leaves resulted in almost 300 data sets with up to twelve channels of data from some collection sessions. Each scan of an area only 40 mm × 20mm produces an ASCII data file of approximately 32 MB. Each ASCII file is converted into twelve TIFF images (one from each elemental detector), each approximately 3 MB in size. Scans of adjacent areas on the page are then stitched to form larger images of critical portions of the forgery leaves with underlying Archimedes text.

3. Image Metadata

With over 5,000 digital images of the Archimedes Palimpsest as of this writing, use of the Dublin Core based metadata elements is proving critical to efficient digital processing and enhancement, digital stitching of the smaller images into larger images, and transcription and study by scholars around the world. This work requires extensive identification metadata to ensure the data remains manageable and accessible, and that a point on any image can be registered to corresponding points on other images [5]. With the public data hosting just beginning in 2006, the metadata has also proven to be somewhat dynamic, with revisions to the standard as new user needs, spatial formats and data hosting challenges have been encountered. While the Dublin Core metadata elements offered the key identification elements required for image storage, management and retrieval, additional spatial information was critical to the process of registering and overlaying images, stitching smaller images to yield larger, high-quality images of full manuscript leaves, and correlating specific image content across various media.

https://doi.org/10.23106/dcmi.952108455

4. Metadata Standard Development

One of the primary goals of the Archimedes Palimpsest Program is to provide ubiquitous and ready access to the images of the Archimedes text for further study by scholars, scientists and the general public. Identification, rights and source information are critical to the effort to ensure all information remains in the public domain for future study.

The Archimedes Palimpsest team has developed its own Archimedes Metadata Standard to meet the unique requirements of digital manuscript imaging [6]. With a small group of data creators and users, the standard development and review effort proved fairly straightforward, as drafts were distributed via the Internet for comment and coordination among the team members. As the program progressed, the standard was refined to keep pace with new data needs, with the better understanding of both the task and the data, and to account for new imaging and processing methods. During the multiyear program, advancing technologies for data processing, storage and access have changed the priority and need for some data elements over time, requiring review and revision of the metadata standard and determination of mandatory use at key milestones in the program. Once stored, all data elements have been retained throughout the program. The team is currently hosting this and other standards on the website archimedespalimpsest.org, and welcomes reviewer comments from any interested parties [7]. The standard has been shared with other parties, including the British Library and Google, for informal comment and review as standard development continues prior to disseminating data through external hosts.

4.1. Archimedes Palimpsest Metadata Standard

The *Archimedes Palimpsest Metadata Standard 1.0* cites 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 information on "Identification," "Data Type" and "Data Content" is based on Dublin Core elements. The "Spatial Data Reference" and "Imaging and Spectral Data Reference" comprises metadata unique to the Archimedes Palimpsest Program, which builds on geospatial metadata elements detailed in the Federal Geographic Data Committee *Content Standard for Digital Geospatial Metadata* [8]. Finally the "Metadata Reference Information" addresses metadata information about the standard itself, as well as any metadata extensions for metadata required for specific aspects of the Archimedes Program. Each of these element sets posed challenges in defining the unique data elements of the Archimedes Palimpsest, but together they offered the necessary foundation structure for integrating the broad range of data at this program continues in the data hosting phase.

4.2. Dublin Core Elements

Beginning in 2000, the Archimedes Palimpsest team began working with the *Dublin Core Metadata Element Set* to incorporate the Dublin Core metadata elements for use in the Archimedes Palimpsest imaging program [9]. A comparison of the *Archimedes Palimpsest Metadata Standard* with the Dublin Core metadata elements is included in Table 1.

https://doi.org/10.23106/dcmi.952108455

Table 1 – Comparison of Archimedes Palimpsest and Dublin Core Metadata Elements

Archimedes Palimpsest Metadata	Dublin Core Metadata Elements
1.1. **Resource Identifier:** A unique data identification number used to reference the data set. *(Computer assigned ID Number)*	**Element Name: Identifier** **Label: Resource Identifier** An unambiguous reference to the resource within a given context.
1.2. **Date:** Date of creation of resource. *(Date of imaging)*	**Element Name: Date** **Label: Date** A date of an event in the lifecycle of the resource.
1.3. **Author or Creator:** The individual or individuals responsible for acquiring content for the data set. *(Names of Imagers or Imager Processors)*	**Element Name: Creator** **Label: Creator** An entity primarily responsible for making the content of the resource.
1.4 **Subject and Keywords:** Description of object type. *("Archimedes Palimpsest")*	**Element Name: Subject** **Label: Subject and Keywords** A topic of the content of the resource.
1.5. **Publisher:** Organization sponsoring the data set. *(Organization responsible for imaging)*	**Element Name: Publisher** **Label: Publisher** An entity responsible for making the resource available
1.6. **Other Contributor(s):** Other contributors to the content of the resource or data study. *(Individuals contributing to the images)*	**Element Name: Contributor** **Label: Contributor** An entity responsible for making contributions to the content of the resource.
1.7. **Resource Type:** Type of data set based on the content of the data. *("Image" or "Annotation")*	**Element Name: Type** **Label: Resource Type** The nature or genre of the content of the resource.
1.8 **Source Lineage:** Information about the events, parameters, and source data that constructed the data set, and information about the responsible parties. *(Images/Data that contributed to the image)*	**Element Name: Source** **Label: Source** A reference to a resource from which the present resource is derived.
1.9. **Rights Management:** Information about rights held in or over the resource. *("Copyright Owner of the Archimedes Palimpsest")*	**Element Name: Rights** **Label: Rights Management** Information about rights held in and over the resource.

https://doi.org/10.23106/dcmi.952108455

Definitions were applied based on the specific requirements of the Archimedes Palimpsest Program, within the broader context of the Dublin Core definitions [10]. Defining the various data elements within the context of the Archimedes Palimpsest program was critical to the effective use of the Dublin Core based metadata. Key to this was agreement that each image was a data set, not the source manuscript. This ensured the metadata provided unique information for the lowest possible data element, and not just higher level data for the entire Archimedes Palimpsest. While common data elements existed for some data sets, such as "Subject and Keywords," and "Rights Management," the unique data for each image accurately identifies the thousands of images, including original, stitched and annotated images. The "Data Type" and "Data Content" metadata have been just as important as the basic "Identification" metadata (if not more so) to place the individual images in context. The "Data Content" metadata elements followed the same Dublin Core format, construction and concepts, offering an additional layer of derived information about the text, the specific leaf from which the image was taken (including the side of the leaf), other images that could be linked to the image, and image processors and others contributing to the image (Table 2).

Table 2 – Archimedes Palimpsest Spatial Metadata Elements

5.1 Content Keyword: Word or phrase used to describe the name of the undertext in the data set.
CORE: YES TYPE: COMPOUND
DOMAIN: "Archimedes" "Euchologion" "Hyperides" "Unknown"

5.2 Source Information: Source document that served as the basis for the data element.
CORE: YES TYPE: COMPOUND
DOMAIN: "Palimpsest leaf" "Heiberg photograph"

5.3 Foliation: Reference for the location of the original source for the image or scanned data in the source object
CORE: YES TYPE: COMPOUND
DOMAIN: TEXT

5.4 Foliation Scheme: Reference for the type of foliation used.
CORE: NO TYPE: COMPOUND
DOMAIN: "Palimpest foliation" "Netz name"

5.5 Source Citation: Reference for a source document.
CORE: NO TYPE: COMPOUND
DOMAIN: TEXT

5.6 Data Set Credit: recognition of those who contributed to the data set.
CORE: NO TYPE: COMPOUND
DOMAIN: TEXT

5.7 Cross Reference: Information about other, related or derived data sets; includes stitched or partially assembled image or scanned data.
CORE: NO TYPE: COMPOUND
DOMAIN: TEXT

https://doi.org/10.23106/dcmi.952108455

Consideration was given to using some of the "Data Content" information as the basic Dublin Core metadata elements, but the decision was made to include these in a section devoted to contextual information to ensure that the metadata is based on the images themselves, instead of the source document. As additional content was identified in the Palimpsest as a result of the imaging, this proved fortuitous, as the original source text proved to be the product of a number of different authors and not just Archimedes.

4.3 Spatial Elements

Unique to document imaging efforts is the requirement for spatial metadata to allow establishment of linkages across domains to specific images and to specific locations within images. This is necessary not only to register locations on the same section of a manuscript leaf in UV and RGB images, but also to link XRF images, conservation drawings, Heiberg image data (if available) and ultimately the transcriptions to the original images. Building on the analogy that a camera collecting images over the manuscript is similar to a satellite collecting geospatial data over the Earth, the Archimedes Palimpsest team used the *Standard for Digital Geospatial Metadata* as a basis for their spatial standard [11]. This required the establishment of a coordinate grid system for the manuscript leaves, and an imaging standard that the Archimedes text would always run horizontally on the image from left to right. Using this coordinate system, the team collected spatial metadata for the upper left and lower right coordinates of the image on the manuscript. An excerpt of the complete record of identification and spatial data elements collected during an XRF imaging session in a collection database are shown in Figure 2.

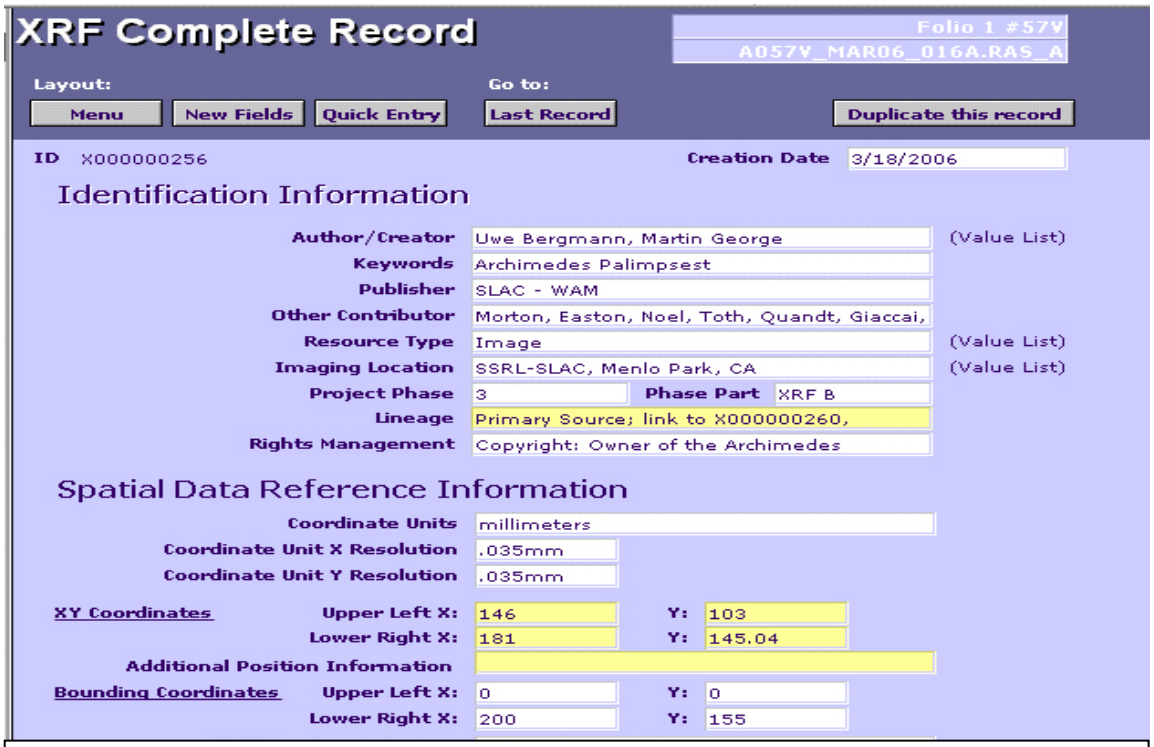

Figure 2 – XRF Metadata Elements for single image recorded in collection database

https://doi.org/10.23106/dcmi.952108455

Since the spatial elements changed most frequently, a "quick table" entry form was also prepared for the database, in which the imagers could readily enter only data elements that changed frequently ("Lineage" and "XY Coordinates" data elements in Figure 2).

4.4 Metadata Extensions

The XRF imaging of the Archimedes Palimpsest required a unique set of metadata extensions to capture the new data elements for different imaging techniques, energy levels and data formats. With no scanning XRF imaging standards available in the XRF literature, the team created a draft set of metadata extensions. These were incorporated into a collection database prior to the first XRF imaging session. The XRF imaging team then refined the additional metadata elements in a dynamic process, while entering data during 24-hour imaging sessions lasting up to 12 days. The numerous variables in position around the manuscript, high energy levels and elemental information had not been anticipated during the early optical imaging phases of the program, requiring significant additional metadata in the XRF Extensions (as excerpted from the collection database in Figure 3). The XRF team reviewed the metadata elements and updated the current draft of the *Archimedes Palimpsest Metadata Standard XRF Extensions* at the end of the imaging session [12]. This standard is also available for broader review on the archimedespalimpsest.org website.

XRF Metadata Extensions

X-Ray Beam Source	wiggler
Tube Voltage	
Tube Current	90-100 mA
Source Bandwidth	1ev
Source Intensity for SLAC	3x10x10 photons per second
Scan Matrix Dimensions X by Y	35 x38
Beam Dimension	50 Micron
Beam Spacing	40 Micron
Scanning Mode	continuous bidirectional
Scanning Speed	493157.481
Beam Distance from Specimen	14 mm
Beam Angle to Specimen	45 degrees
Lens Type	
Primary Beam Filter	

Angle to Beam Longitude in °
Angle to Beam Latitude in °
Distance to Specimen (mm)
Amplifier Time Constant in milliseconds
Counting Equipment Make
Counting Equipment Model

Detectors

Detector	Long	Lat	Dist	Make	Model	Bias	Amp	Count Make	Count Model
I0(Beam Front)	0	0	360	Ion Chamber	SLAC-seeBelow	1500		Keithley	
1 (Archie Front)	270	0	80	Canberra 13	Germanium	-500	.0001	Canberra	2026X
2(Archie Back)	90	0	54	Radiant Vortex	Silicon Drift		.0001	Canberra	2026X
I1 (BeamBack)	180	0	80	Ion Chamber	SLAC-seeBelow	1500		Keithley	

Chamber Atmosphere	Air
Chamber Temperature	72 Degrees
Chamber Humidity	50%
Pixel Acquisition Time (dwell time)	10 milliseconds
Post Processing Module	
Second Post Processing Module	
Time for scan	4 hrs

Figure 3 – XRF Metadata Extension Elements in collection database

https://doi.org/10.23106/dcmi.952108455

5. Scriptospatial Linkage

Access to original images of Archimedes text in this ongoing effort is dependent on fully integrated metadata – both Dublin Core for intelligent information discovery, and spatial for discovery within the context of the original image. Cross-domain linkage and integration of Archimedes Palimpsest information from all sources is the ultimate goal of the continuing program. Just as geospatial data can be linked to images from earth resource satellites, the Archimedes Palimpsest team is working to link a range of "scriptospatial data" with images from optical cameras and XRF detectors [13]. With standardized metadata, data points from the original manuscript in these images can be linked to information derived from the images, including conservation, scientific and scholarly information. This is dependent on common standards for all data elements, and linkage of all six types of metadata elements with common "Identification Information" based on the *Dublin Core Metadata Element Set*. Once the metadata and scriptospatial data is distributed to information providers for hosting on the Internet, the source images and related information can be accessed by a larger group than the initial team of scholars and researchers. Making the source data accessible to a broad range of interested scholars will ensure the information derived from this data – including transcriptions and translations – can be validated from the original images by a larger pool of specialists. With the source data, scientists from a range of disciplines will have the opportunity to apply new image processing algorithms and refine the images with new and more powerful software and hardware. The origins of modern science and mathematics can then be studied from the original text, without risking damage to the fragile manuscript from handling and exposure.

6. Conclusion

Integrating various types of metadata tailored to meet the needs of the users is critical to making unique data available across domains and disciplines. The *Dublin Core Metadata Elements* offer standardization for information access and discovery, and flexibility to support the inclusion of specialized metadata. With this standard as the basis, metadata elements can be added depending on the complexity of the data collection effort, from spatial metadata for images to metadata extensions for a broad range of XRF data elements. With accurate and validated metadata elements, original data elements can be linked to a range of derived data across multiple domains to offer rich layers of information, such as transcriptions and translations, spatially linked to specific locations on images. This information will in turn advance the study of source material by a much larger virtual team with ubiquitous access via the Internet. With advanced information technology tools available to a broad range of information providers, a growing number of specialized metadata elements could support a broad range of applications across a wide and varied range of domains. The *Archimedes Palimpsest Metadata Standard* combines Dublin Core metadata elements with needed new elements from other standards, as well as unique elements not standardized previously. This and other efforts utilizing a range of new techniques and technologies – such as the Archimedes Palimpsest Project, the International Dunhuang Project [14], and various papyri studies – could benefit from an on-line repository of Dublin Core based metadata standards as an initial source for unique metadata element definitions.

https://doi.org/10.23106/dcmi.952108455

7. Acknowledgements

The authors wish to thank the owner of the Archimedes Palimpsest for making this all possible and Will Noel for his leadership as program director. They would also like to thank their colleagues on the Archimedes Palimpsest Program team: Imaging teammate Keith Knox for the many years of optical imaging and processing; Abigail Quandt and Jennifer Giaccai for the conservation and conservation science; Doug Emery for the database and metadata support; Uwe Bergman, Bob Morton and Gene Hall for the XRF imaging; and Reviel Netz, Nigel Wilson and Natalie Tchernetska for the scholarly study. They would also like to thank Bob Toth for the standards support, and other museum and conservation science professionals for their advice and guidance. Archimedes Palimpsest images are Copyright the owner of the Archimedes Palimpsest and used with permission.

8. References

1. Walters Art Museum, *EUREKA! Archimedes Palimpsest at the Walters Art Gallery* [sic], "The Archimedes Palimpsest". <http://www.thewalters.org/archimedes/frame.html> (5 March 2006)
2. Reviel Netz, "The Origin of Mathematical Physics: New Light on an Old Question," *Physics Today*. June (2000) 32-37
3. Walters Art Museum (WAM), *Archimedes – The Palimpsest*. <http://www.archimedespalimpsest.org/> (2 March 2006).
4. Keith T. Knox, Roger L. Easton, Jr., and William A. Christens-Barry, "Multispectral Imaging of the Archimedes Palimpsest," *2003 AMOS Conference.* Maui, Hawaii: Air Force Maui Optical & Supercomputing Site, Sept. (2003)
5. Karin Kneöaurek, et. al., "Medical image registration," *Europhysics News* [online]. Vol. 31 No. 4 (2000) <http://www.europhysicsnews.com/full/04/article1/article1.html> (27 March 2006)
6. Archimedes Palimpsest Program "Archimedes Palimpsest Metadata Standard 1.0X," Revision 3. Baltimore, Maryland: Walters Art Museum, 20 March (2006)
7. WAM, *Archimedes – The Palimpsest,* <http://www.archimedespalimpsest.org/programmanage_documents.html > (4 April 2006).
8. Federal Geospatial Data Committee (FGDC) "Content Standard for Digital Geospatial Metadata: Extensions for Remote Sensing Metadata," FGDC-STD-012-2002. October (2002)
9. Dublin Core Metadata Initiative, *Dublin Core Metadata Element Set, Version 1.1: Reference Description.* <http://www.dublincore.org/documents/dces/>(2000-2006)
10. Dublin Core Metadata Initiative, *DCMI Metadata Terms.* http://www.dublincore.org/documents/dcmi-terms/ (2000-2006)
11. FGDC, "Content Standard for Digital Geospatial Metadata" (2002)
12. Archimedes Palimpsest Program "Archimedes Palimpsest Metadata Standard XRF Extensions," Draft. Baltimore, Maryland: Walters Art Museum, 20 March (2006)
13. Will Noel, Roger L. Easton, Jr., and Michael B. Toth, "The Archimedes Palimpsest," Mountain View, California: Google Inc., 7 March (2006)
14. IDP International Dunhuang Project, *Technical Infrastructure.* <http://idp.bl.uk/pages/technical_infra.a4d > (13 July 2005)

https://doi.org/10.23106/dcmi.952108455