

**Archimedes Palimpsest Transcription Metadata Standard**  
**22 October 2007**  
**DRAFT**

## **INTRODUCTION**

1. Objectives. This standard is intended to provide a common set of terminology and definitions for annotations of the transcriptions of the manuscript known as *The Archimedes Palimpsest*. This standard establishes the names and definitions of data elements, specifies the types and ranges of values these elements may contain, and defines requirements for the entry of metadata for these elements.

The primary uses of these metadata are to:

- a) maintain a standard for transcriptions to be linked to digital imagery or scanned data;
- b) provide information about transcriptions for external data catalogs and service providers;
- c) provide information needed to process and interpret transcriptions to be transferred to an external source.

2. Scope. This standard is intended to support the transcription metadata. It is intended for use by scholars studying the Archimedes Palimpsest, data managers and hosts of the transcriptions. The standard is not intended to reflect an implementation design nor software tools. Its development draws from existing standards that allow four criteria provided by metadata to be determined:

- a) availability: data needed to determine the sets of data that exist for a given location;
- b) fitness for use: data needed to determine if a transcription meets a specific need;
- c) access: data needed to acquire a transcription and its associated data set;
- d) transfer: data needed to process, share and use transcriptions and the data set.

This standard establishes the names of data elements and compound elements to be used for these purposes, definitions of these data elements and compound elements, and information about the values that are to be provided for the data elements. This standard does not specify how this information is organized in a system, nor how it is stored, accessed, disseminated, or presented to the users.

4. Related Standards. These standards were developed to allow the transfer of digital transcription data sets between data holdings. They draw from the standards used in related fields having features and goals in common with the Archimedes Palimpsest Project, in particular:

- a) *Dublin Core Metadata Initiative Element Set* (ISO Standard 15836)<sup>1</sup>;
- b) *Content Standard for Digital Geospatial Data* (FGDC-STD-001-1998)<sup>2</sup>;
- c) *Archimedes Palimpsest Metadata Standard 1.02*

## **DEFINITION OF METADATA**

Metadata are data elements about the content, quality, condition, and other characteristics of the

data sets that make up the digital holdings. Metadata records are to be produced according to rules and definitions governing several subtypes, which are defined in conjunction with the metadata elements used for imaging of the Archimedes Palimpsest:

1. Identification Information
2. Spatial Data Reference Information
3. *Not Used for Transcription Metadata Elements*
4. Data Type Information
5. Data Content Information
6. Metadata Reference Information

## **METADATA PRODUCTION RULES**

Sections 1, 2 and 4-6 define the production rules used for constructing metadata records from these subtypes. Conventions used in these rules are:

CORE: required metadata

TYPE: single or multiple (compound) entries may be present

DOMAIN: permissible metadata formats and values

1. Identification Information: basic information about the data set; domain must accord with external standard standards (e.g. Dublin Core).

1.1 Resource Identifier: A unique data identification number used to reference the data set.

CORE: YES TYPE: SINGLE

DOMAIN: TEXT

1.1.1 File Name: Unique identifier for the data file or files assigned to the transcription.

CORE: YES TYPE: SINGLE

DOMAIN: TEXT

1.2. Date: Date of creation of transcription.

CORE: YES TYPE: SINGLE

DOMAIN: TEXT

1.3. Author or Creator: The individual or individuals responsible for creating the transcription data set.

CORE: YES TYPE: SINGLE

DOMAIN: TEXT

1.4 Subject and Keywords: Description of transcription object type

CORE: YES TYPE: COMPOUND

DOMAIN: TEXT

1.5. Publisher: Organization sponsoring the transcription data set

CORE: NO TYPE: SINGLE

DOMAIN: TEXT

1.6. Other Contributor(s): Other contributors to the content of the transcription resource, data set or data study

CORE:NO TYPE: SINGLE

DOMAIN: TEXT

1.7. Resource Type: Type of data set based on the content of the data

CORE:YES TYPE: SINGLE

DOMAIN: "Transcription"

"Annotation"

"Marked up Image"

1.8 Source Lineage: information about the events, parameters, and source data that constructed the data set, including the source image, and information about the responsible parties.

CORE: YES TYPE: COMPOUND

DOMAIN: TEXT

1.9. Rights Management: Information about rights held in or over the resource

CORE:YES TYPE: COMPOUND

DOMAIN: TEXT

2. Spatial Data Reference Information or Coverage: the spatial extent or scope of the data set

2.1 Coordinate Units: unit type used in quantitative spatial metadata on the source image

CORE:YES TYPE: SINGLE

DOMAIN: "pixel"

"centimeters"

"millimeters"

"micrometer"

*2.2 Not Used for Transcription Metadata Elements*

*2.3 Not Used for Transcription Metadata Elements*

2.4 Upper Left XY Coordinates per line of transcription: upper left coordinate on the source image for each line of the transcription expressed in coordinate units and line number (and column if appropriate).

CORE: YES TYPE: COMPOUND

DOMAIN:  $x = n$  and  $y = n$  and line number =  $n$  (and column "left" or "right" if appropriate)

2.5 Lower Right XY Coordinates per line of transcription: lower right coordinate on the source image for each line of the transcription expressed in coordinate units and line number (and column if appropriate).

CORE: YES TYPE: COMPOUND

DOMAIN:  $x = n$  and  $y = n$  and line number =  $n$  (and column "left" or "right" if appropriate)

2.6 Bounding Coordinates: the limits of coverage of the complete data set for the entire

transcription expressed by Cartesian coordinate values for the entire source image in the order upper left, lower right.

CORE: YES TYPE: COMPOUND

DOMAIN: x left = n and y upper = n, and x right = n and y lower = n

2.8 Grid Coordinate System: Orientation of and definition of Cartesian coordinate so that spatial positions can be readily transformed to and from plane coordinates.

CORE: YES TYPE: COMPOUND

DOMAIN: "Archimedes text UP, running horizontally starting at 0:0 at upper left corner of image"

2.9 X-Positional Accuracy: an estimate of accuracy of the horizontal positions of the lines of transcriptions.

CORE: YES TYPE: SINGLE

DOMAIN: REAL NUMBER

2.11 Y-Positional Accuracy: an estimate of accuracy of the vertical positions of the lines of transcriptions.

CORE: YES TYPE: SINGLE

DOMAIN: REAL NUMBER

### 3. NOT USED

## 4. Data Type Information

4.1 Data File TYPE: File type of the data file.

CORE: YES TYPE: SINGLE

DOMAIN: "Unicode" Unicode Character Encoding Standards  
"ASCII" American Standard Code for Information Interchange

4.2 Format Version Number: Version number of the data file format.

CORE: NO TYPE: SINGLE

DOMAIN: TEXT

4.3 Format Version Date: Date of the version of the data file format.

CORE: NO TYPE: SINGLE

DOMAIN: TEXT

4.4 File Size: The size of the data set in Kilobytes.

CORE: YES TYPE: SINGLE

DOMAIN: REAL

4.5 File Processing Techniques: Description of transcription tools and techniques that contributed to the creation of the data file.

CORE:NO TYPE: COMPOUND  
DOMAIN: TEXT

5 Content Description Information: a general assessment of the content of the data set.

5.1 Content Keyword: word or phrase used to describe the name of the text in the transcription data set.

CORE:YES TYPE: SINGLE  
DOMAIN: "Archimedes"  
"Euchologion"  
"Hyperides"  
"Aristotle Commentary"  
"Unknown"

5.1.1 Transcription Convention: Convention used to render the transcription

CORE:YES TYPE: SINGLE  
DOMAIN: TEXT

5.2 Source Information: Source images that served as the basis for the transcription.

CORE:YES TYPE: COMPOUND  
DOMAIN: "Palimpsest images"  
"Heiberg photograph images"  
"Both"

5.3 Foliation Scheme: Reference for the type of foliation used.

CORE:YES TYPE: COMPOUND  
DOMAIN: "Prayer book foliation"

5.3.1 Secondary Foliation Scheme: Reference for the type of foliation used.

CORE:NO TYPE: COMPOUND  
DOMAIN: "Under text"

5.4 Foliation Number: Reference number for the location of the original source for the image or scanned data in the source object

CORE:YES TYPE: SINGLE  
DOMAIN: NUMBER AND TEXT

5.4.1 Secondary Foliation Number: Secondary reference number for the location of the original source for the image or scanned data in the source object

CORE:NO TYPE: SINGLE  
DOMAIN: NUMBER AND TEXT

5.5 Source Citation: Reference for transcription source document.

CORE:NO TYPE: COMPOUND

DOMAIN: TEXT

5.6 Language: Language of Transcription Text

CORE:NO TYPE: COMPOUND

DOMAIN: TEXT

6 Metadata Reference Information: Information on the validity and state of the metadata information, and the responsible party.

TYPE: compound

6.1 Metadata Status: flag for presence and validity of all core metadata.

CORE:YES TYPE: SINGLE

DOMAIN: "Valid"

"Incomplete"

6.2 Metadata Date: The date that the metadata were created or last updated.

CORE:YES TYPE: SINGLE

DOMAIN: TEXT

6.3 Metadata Review Date: The date of the latest review of the metadata entry.

CORE:NO TYPE: SINGLE

DOMAIN: TEXT

6.4 Metadata Future Review Date: The date by which the metadata entry should be reviewed.

CORE:NO TYPE: SINGLE

DOMAIN: TEXT

6.5 Metadata Contact: The party responsible for the metadata information.

CORE:YES TYPE: SINGLE

DOMAIN: TEXT

6.6 Metadata Standard Name: The name of the metadata standard used to document the data set.

CORE:YES TYPE: SINGLE

DOMAIN: "Archimedes Palimpsest Transcription Metadata Standard"(Including ISO 15836:2003, Dublin Core Metadata Elements)"

6.7 Metadata Standard Version: Identification of the version of the metadata standard used to document the data set.

CORE:YES TYPE: SINGLE

DOMAIN: TEXT

6.8 Metadata Extensions: a reference to extended elements to the standard which may be defined by a metadata producer or the academic community. Extended elements are subjective elements outside the Standard, but needed by the metadata producer or user, sometimes referred to as dynamic metadata. If extended elements are created to the Archimedes Palimpsest Transcription metadata, they must follow the Dublin Core guidelines for Creating Extended Elements and this

document format.

CORE:NO TYPE: COMPOUND

DOMAIN: TEXT

## **Appendix A. References**

### 1. Dublin Core Metadata Initiative

<http://dublincore.org/documents/2001/04/12/usageguide>

### 2. Federal Geographic Data Committee. FGDC-STD-001-1998. Content standard for digital geospatial metadata (revised June 1998),

[http://www.fgdc.gov/standards/status/csdgm\\_rs\\_ex.html](http://www.fgdc.gov/standards/status/csdgm_rs_ex.html)

## **Appendix B. Further Guidelines (TBD)**