5-9-2019

The Galen Palimpsest and the Modest Ambitions of the Digital Data Set

Doug Emery
University of Pennsylvania, emeryr@upenn.edu

The Galen Palimpsest and the Modest Ambitions of the Digital Data Set

Abstract

The digital Syriac Galen Palimpsest (SGP) data set is an archive built on the model of the digital Archimedes Palimpsest. As with Archimedes, the SGP data set is meant to promote the long-term preservation of and access to the digitized palimpsest. The SGP data set follows archiving best practices and uses the Archimedes Palimpsest Metadata Standard for spectral imaging metadata. The data is released under a Creative Commons Attribution 3.0 Unported license (CC BY 3.0). The SGP project used custom software to manage its data and metadata from the time of capture to final data set publication. In the years since initial publication, newly discovered leaves of the manuscript have been discovered, imaged, and added to the on-line archive. Since the publication of the SGP data set, subsequent projects have built on and refined the methods established by the SGP team by moving away from content-based file naming, establishing formal quality assurance practices, increasing automation in the creation and management of data and metadata, and including full bit-depth capture images in the digital product.

MANUSCRIPT STUDIES

A Journal of the Schoenberg Institute for Manuscript Studies

The Schoenberg Institute
for Manuscript Studies
UNIVERSITY *of* PENNSYLVANIA LIBRARIES

Manuscript Studies

A Journal of the Schoenberg Institute for Manuscript Studies

VOLUME 3, NUMBER 1

Special Issue:
The Syriac Galen Palimpsest
Editors: William Noel and Ralph M. Rosen

Articles

The Galen Palimpsest and the Modest Ambitions of the Digital Data Set

Doug Emery
University of Pennsylvania

The digital Syriac Galen Palimpsest (SGP) data set is an archive built on the model of the digital Archimedes Palimpsest. As with Archimedes, the SGP data set is meant to promote the long-term preservation of and access to the digitized palimpsests. The work was accomplished through careful design of the archive following recommended best practices and by adopting a liberal open license to ensure the easy use of the archive and its contents. It is designed to be a self-contained, self-documenting, verifiable data set, amenable to both human and machine access. The data is released under a Creative Commons Attribution 3.0 Unported license (CC BY 3.0), which allows anyone to use the data in any form for any purpose—even commercial use—provided the user gives attribution to the owners of the documents therein. The design of the archive and the openness of its licensing contribute equally to the digital information's ease of use and sustainability, both of which are vital to the data's preservation.

The SGP project was the first time the Archimedes data-capture methods and data model were adapted to a new project. The SGP established a pattern of practice used by the project team for subsequent spectral imaging projects. The effort was focused on efficient and low-cost production of spectral imaging products. This article will look at the purpose of the data

set and its content, the process of its creation, its life since initial release in 2010, and how the team's practice of building spectral data sets has changed in the years since.

Preservation and Access

Books, especially parchment books, can last for centuries. They are not, relatively speaking, fragile or volatile. Digital information—bits stored by altering the charge of microscopic dots on electromagnetic platters that spin at thousands of revolutions per minute—is changeable by design. While some books do degrade and sometimes the text is forcibly removed from their pages, to a large extent, short of completely destroying pages, that information is hard to eradicate completely, as the results of spectral imaging have shown. Digital files become corrupt. Hard drives fail. The hardware needed to read certain media become difficult to find. File formats become outdated and sometimes unreadable when the programs that created them cannot be found or, when found, run on existing hardware.

Codex books are readily comprehensible. The conceptual order and structure of a manuscript's content are an inherent part of its physical order and structure. The physical mechanics of a book and the process of its manufacture help to ensure order and sense. Not so with digital data. A computer hard drive does not ensure readily comprehensible order. A computer's presentation of folders and files that humans use to understand its content is a kind of projection of otherwise incomprehensible and discontinuous clusters of ones and zeros distributed on a hard drive's platters, stored on solid-state chips, or taking up transient space in computer memory. What's more, the structure of folders and files is entirely up to the creator of the content. Computers do not enforce order and structure in the way the codex format does.

The vulnerability of the parchment manuscript as storage medium is not its durability, but rather its relative scarcity. Books that are not reproduced in great numbers are especially at risk. The vulnerabilities of the digital disk as storage medium are its volatility and the fast-changing technologies that employ it. In the digital disk's favor are its low cost and the fact that it can be duplicated perfectly and rapidly an indefinite number of times. When we

seek to preserve a digitized manuscript, especially when that digital information is complex and expensive to come by, as with spectral imaging and processing, we must work to overcome the vulnerabilities of digital storage and exploit its advantages.

The undertext of the Syriac Galen Palimpsest, like that of the more famous Archimedes Palimpsest, has been preserved for centuries on parchment leaves.[1] The modest goal of the project was to create a data set of spectral images that we hope will last for decades. To do this, the pattern established with Archimedes was followed. The Archimedes data set's design was informed by the Open Archival Information System (OAIS) reference model.[2] OAIS describes a model for the ingest, maintenance, and preservation of archival information. Key to the model are the collection and inclusion of data needed to preserve and access the information now and in the future. This includes technical and descriptive metadata—what file formats are present and what their content is—as well as information needed to verify the completeness and integrity of the data, and to migrate that data to new file formats and media as the old ones become obsolete. As such, the digital SGP is a self-documenting, verifiable set of archival-quality data and metadata.

The risk to that data and metadata is minimized by employing international standards and commonly used well-known file and metadata formats. Such practice serves two purposes. First, common, standard formats decrease the difficulty of working with the files in the present and migrating them to new formats in the future. Second, the likelihood of data use, and thus data duplication and proliferation, is increased by the use of common, popular file types. For core data, the digital SGP uses TIFF for archival images and UTF-8 encoded Unicode text documents for metadata.[3]

---

1   G. Kessel, "Membra disjecta sinaitica I: A Reconstitution of the Syriac Galen Palimpsest," *Manuscripta Graeca et Orientalia* 243 (2016): 469–96.

2   *Reference Model for an Open Archival Information System (OAIS),* Consultative Committee for Space Data Systems, 2012, https://public.ccsds.org/Pubs/650x0m2.pdf, accessed 1 June 2017.

3   TIFF (Tagged Image File Format) is typically recommended as an archival image file format for its broad use, the availability of tools, and the fact that—in contrast to a format like JPEG—a TIFF image does not compress image and color data to save disk space. TIFF is said

Metadata is also stored in each TIFF file's ImageDescription tag. Finally, open licensing is used for the image and metadata content. All content in the SGP archive may be used without prior permission of the archives that hold the manuscripts. This last point is of critical importance. The open licensing of the material, via the Creative Commons Attribution 3.0 Unported license (CC BY 3.0, https://creativecommons.org/licenses/by/3.0/),[4] is as important as the structure of the archive and the use of standards. Without open licensing, the free exploitation of the data that is vital to its longevity is impeded.

The Structure of the Digital Palimpsest

The structure of the data set, the names of the files, and the included metadata all serve their part in the digital SGP. While it is available on the web at http://digitalgalen.net, and currently hosted by the University of

---

to be "non-lossy": a TIFF image is a grid of pixels, and stores full grayscale or color information at each of those pixels. An 8-bit image—one that stores one byte, or eight bits per color sample per pixel—will typically store one or three bytes for each pixel in an image, one byte for a greyscale image, or three bytes for an RGB (red-green-blue) image. By contrast, JPEG images, which are commonly used for web pages because of their compact size, are "lossy." They save space by storing only partial pixel and color data and use interpolation to make relatively accurate guesses to reconstruct color information not stored in the file itself at the time of display.

For the TIFF specification, see *TIFF Revision 6.0* (Mountain View, Calif., 1992), no longer available online (accessed 30 July 2007). For information on TIFF as an archival format, see "Sustainability of Digital Formats: Planning for Library of Congress Collections: TIFF, Revision 6.0," Library of Congress, https://www.loc.gov/preservation/digital/formats/fdd/fdd000022.shtml, accessed 31 May 2017.

UTF-8 is a Unicode encoding scheme. Unicode provides a single numerical encoding for all the world's scripts, assigning a unique numerical value or "code point" to each character. It replaces older systems that relied on multiple, often conflicting, encoding schemes. See "What Is Unicode?," Unicode Consortium, http://www.unicode.org/standard/WhatIsUnicode.html, accessed 4 June 2017. For UTF-8, see "RFC 3629: UTF-8, a Transformation Format of ISO 10646," https://tools.ietf.org/html/rfc3629, accessed 4 June 2017.

4   "Attribution 3.0 Unported (CC BY 3.0)," https://creativecommons.org/licenses/by/3.0/, accessed 4 June 2017.

Pennsylvania Libraries, the digital SGP is not a website, nor is it a database. It is a single set of files arranged under a single directory structure. All its files are core data or in some way support the understanding and use of that data, by humans and computers. The top-level directory, the *root*, of the digital SGP is shown in figure 1. As you can see, the root directory has four files and five subdirectories. The files are an introductory Read Me file and a list of files in the archive. Both are in text and HTML formats. The directories are:

- Data—the core image data files and their supporting information
- Documents—documentation, project-specific documentation in an "Internal" subdirectory, and documentation of the standards and file types used in the project (such as TIFF, XMP, and MD5) in an "External" subdirectory
- ResearchContrib—important and useful image files that in one way or another do not conform to the standards for images in the core Data directory; for example, details or experimental images of the palimpsest leaves
- Supplemental—alternate presentations of source material used to generate text and other content included with the core data; currently empty[5]
- Support—functional files used by the archive or the data set; currently the Cascading Stylesheet (CSS) file used for the appearance and style of the archive's HTML documents

The focus of the archive is the Data directory. It contains the image files and the metadata that directly supports them. Figure 2 shows the first few subdirectories of the Data directory. Each of these directories contains capture and processed images from a single shot sequence of a bifolium of the disbound SGP. A portion of the contents of the directory for bifolium

---

5 For the Archimedes Palimpsest, the supplemental directory (http://archimedespalimp sest.net/Supplemental/) is used for treatise-length XML transcriptions and text files of the line-by-line coordinate mappings used to generate the line-mapped, per-folio transcriptions found in the core data directory.

Name	Last modified	Size
0_ReadMe.html	2010-07-21 21:31	18K
0_ReadMe.txt	2010-07-21 21:31	11K
1_FileList.html	2010-07-21 22:37	2.0M
1_FileList.txt	2010-07-21 22:34	683K
Data/	2016-04-29 16:26	-
Documents/	2010-07-09 11:43	-
ResearchContrib/	2010-07-19 14:46	-
Supplemental/	2010-07-08 13:51	-
Support/	2010-07-09 14:44	-

FIGURE 1. SGP home directory.

42v+43r is shown in figure 3. This directory contains twenty-seven archival TIFF images. Of this number, twenty-three are "pack8" files. These are 8-bit versions of the grayscale 16-bit capture images that have been computationally adjusted for visibility. The remaining four are processed images. These were computationally generated from select capture images to enhance the visibility of the undertext.

Each TIFF image file is accompanied by three other supporting files. These are a JPEG thumbnail image, an XMP "sidecar" metadata file, and an MD5 file containing the TIFF file's checksum digest. In more detail:

- The *JPEG thumbnail* file is one-tenth the size of the TIFF image of the same name. It is intended to provide a quick web preview of the TIFF's content. Example: 042v-043r_A_0365_pack8.jpg.[6]
- The *XMP file* contains the TIFF image's metadata following the Archimedes Palimpsest metadata standard.[7] It provides direct access to

---

6   http://digitalgalen.net/Data/042v-043r/042v-043r_A_0365_pack8.jpg.
7   XMP is an ISO standard for embedding metadata in and sharing metadata about digital data files, developed by Adobe Systems and adopted in 2012 as ISO standard 16684-1:2012. See "Adobe XMP Developer Center," http://www.adobe.com/devnet/xmp.html, accessed 31

	Name	Last modified	Size
	Parent Directory		-
	001r-004v/	2011-08-17 12:14	-
	001v-004r/	2011-08-17 12:14	-
	002r-003v/	2011-08-17 12:14	-
	002v-003r/	2011-08-17 12:14	-
	005r-006v/	2011-08-17 12:14	-
	005v-006r/	2011-08-17 12:14	-

FIGURE 2. Portion of the Data directory.

	Name	Last modified	Size
	Parent Directory		-
	042v-043r_A_0365_pack8.jpg	2010-07-21 19:16	124K
	042v-043r_A_0365_pack8.tif	2010-07-06 04:26	37M
	042v-043r_A_0365_pack8.tif.md5	2010-07-06 08:44	61
	042v-043r_A_0365_pack8.xmp	2011-08-17 12:04	12K
	042v-043r_A_0450_pack8.jpg	2010-07-21 19:17	128K
	042v-043r_A_0450_pack8.tif	2010-07-06 04:26	37M
	042v-043r_A_0450_pack8.tif.md5	2010-07-06 08:44	61
	042v-043r_A_0450_pack8.xmp	2011-08-17 12:04	12K

FIGURE 3. Image directory for SGP bifolium 42v+43r.

the metadata embedded in the TIFF's ImageDescription tag. Example: 042v-043r_A_0365_pack8.xmp.[8]

May 2017. XMP has come to be the preferred method for adding metadata to media files due to its extensibility and ability to handle Unicode character encodings.

8   http://digitalgalen.net/Data/042v-043r/042v-043r_A_0365_pack8.xmp.

- The *MD5 file* contains the MD5 checksum digest of the TIFF file[9] and is to be used to verify file integrity. Example: 042v-043r_A_0365_pack8.tif.md5.[10]

Digital Galen Metadata

The digital SGP uses the Archimedes Palimpsest Metadata Standard (APMS) for images.[11] The APMS defines six types of information:

1. Identification information—identification of the individual image and its lineage (field prefix: *ID_*)
2. Spatial data reference information—relation and orientation of the image to the physical object (field prefix: *SPTL_*)

---

9   The MD5 file can be used to verify the image file's integrity via checksum. A checksum, in this case, is a string of characters called a message digest (e.g., 5764db9519084649aee36 237b11b94f8). When a file's content is read by a checksum program, like md5sum, the digest it outputs will always be the same, provided the file's bit content has remained unchanged. A change of a single bit in a file, no matter how large the file, will cause the checksum digest to be significantly different. For the SGP, MD5 checksums were used. The MD5 message-digest algorithm is defined in RFC 1321, https://tools.ietf.org/html/rfc1321.

The following lines from a command line session show the content of the MD5 file "042v-043r_A_0365_pack8.tif.md5," the checksum of the TIFF file as output by the md5sum program, and how the md5sum program uses the MD5 file to verify the TIFF file's integrity.

```
$ cat 042v-043r_A_0365_pack8.tif.md5
35f8f7f17ee603a078d6818e1281013d *042v-043r_A_0365_pack8.tif

$ md5sum 042v-043r_A_0365_pack8.tif
35f8f7f17ee603a078d6818e1281013d  042v-043r_A_0365_pack8.tif

$ md5sum —check 042v-043r_A_0365_pack8.tif.md5
042v-043r_A_0365_pack8.tif: OK
```

10   http://digitalgalen.net/Data/042v-043r/042v-043r_A_0365_pack8.tif.md5.
11   *Archimedes Palimpsest Metadata Standard 1.0,* 2006, http://archimedespalimpsest.net/ Documents/Internal/Image_Metadata_Standard.pdf, accessed 31 March 2016.

3. Imaging and spectral data reference information—parameters of the image's capture (such as aperture and exposure) and spectral characteristics (such as illumination wavelength and filter) (field prefix: *IMG_*).

4. Data type information—file format information and technical provenance (such as processing techniques) (field prefix: *DAT_*).

5. Data content information—content descriptive information, such as keywords and foliation information (field prefix: *CONT_*).

6. Metadata reference information—identification of the metadata standard used and its version (field prefix: *MET_*).

Several of the descriptive elements in the APMS are taken from the Dublin Core Metadata Element Set. The Dublin Core Metadata Element Set provides a set of fifteen terms for the description of resources, typically resources found on the web.[12] Dublin Core terms are commonly used on the web to provide basic, machine-readable descriptive metadata, like Title, Description, Subject, and Date. In the APMS, Dublin Core terms have the expected names, like 'identifier', 'date', and 'subject'. All other elements have prefixes that indicate the element's information type. These six information types can be thought of as axes of information that intersect uniquely in each image. For example, all twenty-seven images of folio 42v+43r have the same data content information, but only one of those images has the spectral information associated with the 365-nanometer capture.

The APMS information types are expressed in over one hundred separate elements. Which elements are used depends on the image. Capture images have imaging and illumination information associated with them. Processed images, on the other hand, which are generated from multiple capture images, lack most imaging and illumination information, but are rich in data content information that describes the tools and techniques that were used to generate it. For an example, see the Appendix, which has the full text of an XMP sidecar file for an 8-bit version of a capture image and shows all the APMS metadata elements used.

---

12 They are Title, Creator, Subject, Description, Publisher, Contributor, Date, Format, Identifier, Source, Language, Relation, Coverage, and Rights. See *Dublin Core Metadata Element Set, Version 1.1,* 2012, http://dublincore.org/documents/dces/, accessed 4 June 2017.

```
identifier 21-000022846
date 2010-05-08T08:53:21-04:00
creator Bright, Allison
creator Christens-Barry, William A.
creator Coyle, Meghan
creator Easton, Roger
...
subject Medical Palimpsest Image
...
publisher Owner of the Syriac Palimpsest
description Image of folio 001r-004v of Syriac Liturgical Text with
    Medical Palimpsest, Private Collection (Hiersemann Katalog 500,
    1922, Nr.20)
ID_File_Name 001r-004v_C_0365_pack8.tif
...
SPTL_X_Resolution 23.98
SPTL_Y_Resolution 23.98
...
SPTL_Grid_Coordinate_System Over text UP, running horizontally
    starting at 0:0 at upper left corner of image
...
IMG_Spectral_Range 350-1100
IMG_Spectral_Range_Measurement_Unit nm
IMG_Illumination LED365
IMG_Illumination_Filters Acrylic prismatic diffusers common to all light sources
left and right
IMG_Acquisition_Filters Infrared filter removed from camera
...
```

FIGURE 4. ImageDescription metadata excerpt.

The full metadata for each image is also stored in the TIFF's Image Description tag.[13] Figure 4 shows a portion of the metadata extracted from 001r-004v_C_0365_pack8.tif. Note the Dublin Core elements 'identifier', 'date', 'creator', and others, as well as the APMS specific-fields with prefixes, like 'ID_File_Name' and 'SPTL_Grid_Coordinate_System'.

Collection of the Capture Data

Building on the experience of building the Archimedes Palimpsest data set, the SGP project was the first of a number of iterations in the development of an integrated system for the collection of spectral imaging data and metadata. It was known from the beginning exactly what metadata would

---

13 The TIFF standard specifies a number of text fields called *tags* in the image's header—that is, the non-image metadata section of the binary file (*TIFF 6.0*, 14–16, 117–18). As noted above, XMP has come to be the preferred format for embedding metadata, largely replacing the older system of TIFF tags.

be collected and what form the data set would take. The goal was a rapid and low-cost turnaround from the time of capture to publication. The team designed an integrated workflow that included both human and software processes for the collection of data and metadata from capture through processing. This system included software to manage the entire life cycle of the metadata, worksheets for collecting system geometry, and spreadsheets to collect structured information about image processing methods. The software was a Ruby on Rails application called the Multi-Spectral Imaging Manager or MSIMM. All metadata collected by the project was eventually entered into MSIMM.

The one-hundred-plus data elements in our implementation of the APMS represent multiple strands of data and come from multiple sources. Data for content information came from scholars and manuscript experts. Some information, like exposure time and camera settings, was collected by the camera software and embedded in each image's header. Detailed information about the illuminants and filters was provided by scientists on the project, as was detailed information about techniques used to generate processed images. Likewise, information about the camera system's geometry—the relative positions and angles of camera and lights to the imaged object—had to be collected by hand. The identities of project participants for acknowledging contribution to the project also had to be collected.

The first link in the chain of data collection came from MSIMM, which generated base file name components. The file name was the lynchpin of data management for the SGP project. Data managers, scientists, and software that worked with the data relied on the file name to identify image content and circumstances of image creation. The file name conveys each image's unique identity.

Before a bifolium was imaged, MSIMM was used to create the first segments of the filename. Take, for example, the filename 016v-021r_A_0735.dng. The initial two segments say that this is an image of bifolium 16v+21r, and it is from the first sequence of images captured of that bifolium, designated by the letter A. If a bifolium was imaged again, sequence letters B, C, D, and so on would be used. When a user entered the folio or bifolium to be imaged, MSIMM would generate a series of possible file name bases.

016v-021r_A
016v-021r_B
016v-021r_C
016v-021r_D
016v-021r_E
016v-021r_F

The camera operator would enter the first of these base names, '<EXT_CODE>016v-021r_A</EXT_CODE>', into the camera software, MegaVision PhotoShoot, which used it to create the file names of captured images. Any important information about a set of shots was logged in MSIMM for future use. This log was exported to Excel and distributed to project members. Figure 5 is a portion of the log that shows how the notes were useful later, especially for understanding why there may be multiple shot sets for a bifolium and which of them to use.

PhotoShoot provided the next link in the data management chain in the form of a shot designation code. PhotoShoot was configured to manage the spectral light system (built by project scientist William A. Christens-Barry), as well a filter wheel that would move color filters in front of the camera lens for certain exposures. Each exposure in the capture sequence was configured in PhotoShoot's *n-shot table* with its own exposure time, light sources, and optional acquisition filter. These configurations were represented by twenty-three four-character codes, shown in figure 6. The file base-name component and exposure code constitute the unique identity of each image. These codes and the corresponding configuration information were added to MSIMM's database, as were all other relevant metadata values, like spatial reference information and camera setup geometry. To construct a capture image's metadata, MSIMM would parse the image's file name, collect the relevant metadata based on bifolium, shot sequence letter, and capture code, and combine that with metadata extracted from the image's header. This metadata was output in two formats: XMP serialized as XML and text file of a list of name-value pairs for insertion into the image's Image-Description tag.

A similar method was used with processed images. Separate codes were created by project scientists for each of their processing methods.

	A	B	C	D	E	F
1	folio_id	folio_name	shot_set_id	set_name	shot_time	comment
50	48	047v-054r	236	047v-054r_A	2010-02-25 14:10:00 -0500	
51	49	048r-053v	241	048r-053v_A	2010-02-25 14:19:00 -0500	
52	50	048v-053r	246	048v-053r_A	2010-02-25 14:29:00 -0500	Undertext visible in UV.
53	51	049r-052v	251	049r-052v_A	2010-02-25 14:38:00 -0500	
54	52	049v-052r	256	049v-052r_A	2010-02-25 14:39:00 -0500	Undertext visible under UV.
55	53	050r-051v	261	050r-051v_A	2010-02-25 14:52:00 -0500	Extra shots of other leaf deleted (24-40).
56	54	050v-051r	266	050v-051r_A	2010-02-25 14:59:00 -0500	Undertext visible under UV.
57	55	055r-060v	271	055r-060v_A	2010-02-25 15:12:00 -0500	
58	56	055v-060r	276	055v-060r_A	2010-02-25 15:20:00 -0500	Malfunction. System rebooted. Trashed data and reshot. Drive was switched. Undertext visible in UV.
59	57	056r-059v	281	056r-059v_A	2010-02-25 15:39:00 -0500	Undertext visible in UV.
60	58	056v-059r	286	056v-059r_A	2010-02-25 15:50:00 -0500	
61	59	057r-058v	291	057r-058v_A	2010-02-25 15:57:00 -0500	
62	60	057v-058r	296	057v-058r_A	2010-02-25 16:06:00 -0500	Set #2 accidentally shot first. Reshot. Possibly misordered.
63	60	057v-058r	297	057v-058r_B	2010-02-25 16:20:00 -0500	Undertext lightly visible under UV.
64	61	061r-0000	301	061r-0000_A	2010-02-25 16:28:00 -0500	
65	62	061v-0000	306	061v-0000_A	2010-02-26 09:25:00 -0500	
66	63	062r-069v	311	062r-069v_A	2010-02-26 09:46:00 -0500	
67	64	062v-069r	316	062v-069r_A	2010-02-26 09:55:00 -0500	
68	65	063r-068v	321	063r-068v_A	2010-02-26 10:06:00 -0500	Program unexpectedly quit and appeared to not have saved set #2.
69	65	063r-068v	322	063r-068v_B	2010-02-26 10:11:00 -0500	

Syriac_Palimpsest_2010 +

FIGURE 5. MSIMM log data.

```
0365 - 365 nm LED illumination
0450 - 450 nm LED illumination
0470 - 470 nm LED illumination
0505 - 505 nm LED illumination
0535 - 535 nm LED illumination
0570 - 570 nm LED illumination
0615 - 615 nm LED illumination
0630 - 630 nm LED illumination
0700 - 700 nm LED illumination
0735 - 735 nm LED illumination
0780 - 780 nm LED illumination
0870 - 870 nm LED illumination
0940 - 940 nm LED illumination
RAIR - raking infrared (940 nm) illumination from the right
RABR - raking blue (470 nm) illumination from the right
RAIL - raking infrared (940 nm) illumination from the left
RABL - raking blue (470 nm) illumination the left
CFUR - ultraviolet (365 nm) illumination with red color filter
CFUG - ultraviolet (365 nm) illumination with green color filter
CFUB - ultraviolet (365 nm) illumination with blue color filter
CFBR - blue (450 nm) illumination with red color filter
CFBG - blue (450 nm) illumination with green color filter
CFBB - blue (450 nm) illumination with blue color filter
```

FIGURE 6. Capture codes.

PCABCB Single ▼		ACTION	Archimedes Sample
File Types	PCABCB_175v-178r_EXP_B-C1, etc.	N/A	
Short Name	pcabcb_single	N/A	
DAT_File_Processing	Eight-bit TIFF version of one plane, designated by the file name component _C<n>, of a three-plane selection from a principle component analysis of four images captured at 365 nm, 570 nm, and 735 nm illuminations, and 365 nm illumination with blue camera filter, interspersed in a 12-plane, 16-bit image cube, each image of which was cropped and normalized.	PLEASE VERIFY	A linear contrast stretch adjusted the 16-bit, minimum and maximum values of the input image to 8-bit, black and white values, respectively.
DAT_Joining_Same_Parts_of_Folio	Yes	PLEASE VERIFY	No
DAT_Joining_Different_Parts_of_Folio	No	PLEASE VERIFY	No
DAT_Samples_per_Pixel		NEED VALUE	3
DAT_Bits_per_Sample	8	PLEASE VERIFY	8
DAT_Type_of_Contrast_Adjustment		NEED VALUE	None
DAT_Type_of_Image_Processing		NEED VALUE	linear/stretch
DAT_Software_Version		NEED VALUE	1.1
DAT_Processing_Program	ImageJ, with BIJ plugin	PLEASE VERIFY	Archie
DAT_Processing_Comments		NEED VALUE	
IMG_White_Balance		NEED VALUE	None
PCABCB RGB		ACTION	Archimedes Sample
File Types	PCABCB_175v-178r_EXP_B-1-3-inv6_rgb, etc	N/A	
Short Name	pseudo-color	N/A	
DAT_File_Processing	Eight-bit color TIFF created from three-planes selected from a principle component analysis of four images captured at 365 nm, 570 nm,	PLEASE VERIFY	A linear contrast stretch adjusted the 16-bit, minimum and maximum values of the input image to 8-bit, black and white values,

FIGURE 7. Processing worksheet.

The details of these methods were provided in Excel spreadsheets filled in by the project scientists and loaded into MSIMM. An example is shown in figure 7.

From the data management perspective, an application like MSIMM is a source and hub for project data and metadata. The identifying file name components originated with MSIMM. Images with those file name components were captured, collected, and then duplicated multiple times for distribution to image scientists for processing. Once processed, new images, bearing the same identifying components, were returned to the author, the project's data manager, for assembly and publication. The list of all those files along with exported metadata from them was loaded into MSIMM. Once all required metadata from all sources was entered, MSIMM was used to create final metadata. Custom scripts were used to build the published data set from the processed images and MSIMM-generated metadata.

Syriac Galen Data Since 2010

As reported in the *New York Times,* Grigory Kessel has uncovered six more leaves of the SGP, at Harvard's Houghton Library, the Bibliothèque natio-

nale de France, the Vatican, and St. Catherine's Monastery in the Sinai.[14] Images of these leaves are online with the rest of the digital palimpsest.[15] The later images were captured by separate projects, with different equipment and methods. As such, their metadata is not as full as that from the 2010 session. They were, however, captured or processed by all or part of the same team that worked on the first session. The format of the file names and thus the appearance of the data is similar to that from 2010.

The later images were added quickly in 2015 and 2016 to ensure their availability and the completeness of the archive. As of this writing, the Read Me file, file list, and documentation have not been updated to reflect the new data. When those documents have been updated and corrected, the SGP data set will form a complete digital record of the known Syriac Galen Palimpsest.

Spectral Data Sets Since 2010

Since 2010, all or part of the team that worked on the SGP project has participated in other spectral imaging projects. The author has worked on two significant projects of this kind, the National Endowment for the Humanities–funded David Livingstone 1871 Field Diary project[16] and the Arcadia-funded Sinai Palimpsests Project.[17]

Unlike Archimedes and Galen, the Livingstone and Sinai projects required working with multiple documents. These more complicated projects presented new challenges for the collection, management, and presentation of data. It is important to understand that in all these spectral

---

14 Mark Schrope, "Medicine's Hidden Roots in an Ancient Manuscript," *New York Times,* 1 June 2015, https://www.nytimes.com/2015/06/02/science/medicines-hidden-roots-in-an -ancient-manuscript.html, accessed 4 June 2017.

15 These are Houghton Syriac MS 172, BnF syr. 382, Vat. sir. 623, Vat. sir. 647, and Sinai Syriac NF frag. 65. See http://digitalgalen.net/Data/, accessed 4 June 2017.

16 *Livingstone's 1871 Field Diary: A Multispectral Critical Edition*, http://livingstone.library. ucla.edu/1871diary/index.htm, accessed 4 June 2017.

17 *Sinai Palimpsests Project: About the Project*, http://sinaipalimpsests.org/about-project accessed 4 June 2017.

imaging projects, a data manager assembles hard drives of captured data and sends them to multiple imaging scientists for processing. The scientists send the resulting processed images to the data manager, who must collect and collate them, and prepare them for publication with complete APMS metadata for each file. This requires not only tracking hard drives and outgoing and incoming files, but ensuring that the images can be associated with the one-hundred-plus metadata elements that uniquely describe each file. On a relatively small project like the SGP, there are several thousand files. At the other extreme, the five-year Sinai project dealt with over 60 documents and fragments, over 6,500 folio sides, and approximately 600,000 image files.

A hard lesson learned on the Livingstone project was the risk of using content-identifying information—namely, shelf marks and page numbers—in file names. Hundreds of files from one of the Livingstone documents were misnamed when, during imaging, two pages were accidentally turned instead of one, throwing all page numbers off from that point forward. This complication affected both the files and the database that had associated capture metadata for those images. Not only did files need to be renamed, but a change in the naming format affected both the files and the database records that depended on those names and required the manual correction of the associated records in the version of MSIMM that was used for that project. Many hours were lost addressing this problem.

Also complicating the Galen and Livingstone projects was the need to collect required, detailed metadata about processing parameters from the imaging scientists. Very often the information was collected well after processing had been completed, making information-gathering time-consuming and inefficient. The association of processing metadata with the correct files relied on parsing the file names in order to link images with the correct metadata, a method susceptible to error.

The scale of the Sinai project required different approaches for file naming, quality control, and the handling and processing of metadata.

First, the reliance on content information for file names was done away with. KatIkon, a new cataloging and data management application created for the project, assigned arbitrary numbers to each sequence of images.

Most folios were shot as a single sequence, but if a folio was imaged twice or more, each time it would be done using a different *shot sequence number*. The relationship between this number and content metadata was maintained in KatIkon's database until an image was ready for publication. Then each file was renamed with the subject's shelf mark and folio number. Before that time, if an error was discovered in the association between the folio and the shot sequence number, it could be corrected in one place in the database *without the need to change file names*. Because arbitrary sequence numbers make for somewhat cryptic file names, project systems and practices were developed to help team members identify the content of images before publication. These included the distribution of contents lists and the addition, to each directory of images, of a text file named with the sequence's shelf mark and folio number.

Second, the practice of manually entering file names in the camera software was eliminated, and a system of post-capture independent verification and validation (IV&V) was used to ensure quality and accuracy. For file names, KatIkon was used to generate an ordered *shoot list* of folios to be imaged with their identifying shot sequence numbers. For the project, Megavision added a feature to the PhotoShoot camera software to manage imaging and name files based on the computer-generated shoot list. The operator would select the correct shot sequence from the shoot list, and PhotoShoot handled the rest, using the shot sequence number for the beginning of each file name. Immediately after a sequence was completed and all files were written to disk, the IV&V operator inspected the images to verify accuracy, rotation, and image quality.

To manage processed image metadata, a system was established for validating and packaging processed images called "spindle."[18] To be valid, processed images had to have required metadata in their headers and conform to the project's requirements for valid file names. Imaging scientists used a spindle script called "deliver" to validate and then package a directory of

---

18 The spindle project is hosted on Github at https://github.com/EarlyMssElectronic Libary/spindle.

images for delivery to the project system administrator. Upon receipt, the system administrator would validate a package of images by running the "receive" script. Other scripts in the spindle suite were used by the system administrator to add capture and processed images to the project's working repository and to extract metadata from them to be added to KatIkon's online database. This data, along with cataloging and other information, is used by KatIkon to construct zip archives of metadata and packaging instructions used by custom scripts to assemble packages of data for delivery to scholars and for archiving.

A final innovation of Sinai project data delivery was the inclusion of 16-bit monochrome capture TIFF images with the 8-bit monochrome and 24-bit color processed images. The SGP does not include the 16-bit capture images from the project. This decision was made following the Archimedes project. That data set is one terabyte in size, which was quite large in 2008 when the data was published. The overall size owes in part to the very large size of the individual files, each of which is a 256 MB, 24-bit color TIFF. For Archimedes, the 48-bit color capture images were excluded for two reasons. For general access, the 24-bit images had to be included because most commonly available image software supports only 8- and 24-bit files. Second, the full-depth 48-bit images were each 512 MB. Their inclusion would have tripled the size of the archive to an unmanageable three tera-bytes. The selection of the 8- and 24-bit images for the Galen project was a holdover from Archimedes, even though the 8-bit monochrome files were 37 MB, meaning that the capture images were approximately 75 MB each, and would likely add only another 500 GB to the 300 GB of the present archive. In the intervening seven years, storage costs have come down con-siderably. Given the archival goals of the data set, the decision to exclude these best-quality capture images is worth reconsidering.

Conclusion

The Syriac Galen Palimpsest project established a standard method for collecting spectral imaging data and assembling a data set. Archimedes

offered the template for the project's output, but Galen was the first such project that began with a clear idea of a final product, with a clear roadmap of how to proceed. The project was not without its complications and areas that were later improved upon. Nevertheless, later projects have been exactly that: improvements and refinements of the method established with Galen.

Appendix: XMP Sample

The following is the content of the XMP data for SGP image file 001r-004v_C_0365_pack8.tif.

```
<?xml version="1.0" encoding="UTF-8" ?>
<x:xmpmeta xmlns:x="adobe:ns:meta/" x:xmptk="Adobe XMP Core 4.1.1">
  <rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#">
    <rdf:Description rdf:about="" xmlns:dc="http://purl.org/dc/elements/1.1/">
      <dc:identifier>21-000022846</dc:identifier>
      <dc:date>2010-05-08 12:53:21</dc:date>
      <dc:creator>
       <rdf:Seq>
         <rdf:li>Bright, Allison</rdf:li>
         <rdf:li>Christens-Barry, William A.</rdf:li>
         <rdf:li>Coyle, Meghan</rdf:li>
         <rdf:li>Easton, Roger</rdf:li>
         <rdf:li>Hill, Meghan</rdf:li>
         <rdf:li>Knox, Keith</rdf:li>
         <rdf:li>Safford, Janet</rdf:li>
         <rdf:li>Ware, Mary</rdf:li>
       </rdf:Seq>
      </dc:creator>
      <dc:subject>
       <rdf:Bag>
         <rdf:li>Medical Palimpsest Image</rdf:li>
         <rdf:li>Troparia Image</rdf:li>
         <rdf:li>11th Century Manuscript Image</rdf:li>
         <rdf:li>8th-9th Century Manuscript Image</rdf:li>
         <rdf:li>Parchment Manuscripts Image</rdf:li>
         <rdf:li>Galen Manuscript Image</rdf:li>
         <rdf:li>Medical Manuscript Image</rdf:li>
       </rdf:Bag>
      </dc:subject>
      <dc:publisher>Owner of the Syriac Palimpsest </dc:publisher>
      <dc:contributor>
       <rdf:Seq>
         <rdf:li>Boydston, Ken</rdf:li>
         <rdf:li>Carney, Vincent</rdf:li>
         <rdf:li>Emery, Doug</rdf:li>
         <rdf:li>Noel, William</rdf:li>
         <rdf:li>Quandt, Abigail</rdf:li>
         <rdf:li>Toth, Michael B.</rdf:li>
```

```
    </rdf:Seq>
  </dc:contributor>
  <dc:type>Image</dc:type>
  <dc:source>
   <rdf:Bag>
    <rdf:li>Image 15523: 001r-004v_C_001.dng</rdf:li>
   </rdf:Bag>
  </dc:source>
  <dc:rights>
   <rdf:Bag>
    <rdf:li>Licensed for use under Creative Commons Attribution 3.0
      Unported Access Rights,
      http://creativecommons.org/licenses/by/3.0/legalcode.</rdf:li>
   </rdf:Bag>
  </dc:rights>
  <dc:format>image/tiff</dc:format>
 </rdf:Description>
 <rdf:Description rdf:about=""
  xmlns:xapRights="http://ns.adobe.com/xap/1.0/rights/">
  <xapRights:Marked>true</xapRights:Marked>
  <xapRights:WebStatement>
   http://creativecommons.org/licenses/by/3.0/legalcode
  </xapRights:WebStatement>
  <xapRights:UsageTerms>Licensed for use under Creative Commons Attribution
   3.0 Unported Access Rights,
   http://creativecommons.org/licenses/by/3.0/legalcode.
  </xapRights:UsageTerms>
 </rdf:Description>
 <rdf:Description rdf:about=""
  xmlns:ap="http://www.archimedespalimpsest.org/ns/metadata/1.0/">
  <ap:ID_File_Name>001r-004v_C_0365_pack8.tif</ap:ID_File_Name>
  <ap:SPTL_X_Resolution>23.98</ap:SPTL_X_Resolution>
  <ap:SPTL_X_Resolution_Unit_Of_Measure>pixels per
   mm</ap:SPTL_X_Resolution_Unit_Of_Measure>
  <ap:SPTL_Y_Resolution>23.98</ap:SPTL_Y_Resolution>
  <ap:SPTL_Y_Resolution_Unit_Of_Measure>
   pixels per mm
  </ap:SPTL_Y_Resolution_Unit_Of_Measure>
  <ap:SPTL_Upleft_X_Coordinate>0</ap:SPTL_Upleft_X_Coordinate>
  <ap:SPTL_Upleft_Y_Coordinate>0</ap:SPTL_Upleft_Y_Coordinate>
  <ap:SPTL_Loright_X_Coordinate>7216</ap:SPTL_Loright_X_Coordinate>
  <ap:SPTL_Loright_Y_Coordinate>5412</ap:SPTL_Loright_Y_Coordinate>
  <ap:SPTL_Upleft_X_Boundingcoordinate>
```

```
  0
</ap:SPTL_Upleft_X_Boundingcoordinate>
<ap:SPTL_Upleft_Y_Boundingcoordinate>
  0
</ap:SPTL_Upleft_Y_Boundingcoordinate>
<ap:SPTL_Loright_X_Boundingcoordinate>
  7216
</ap:SPTL_Loright_X_Boundingcoordinate>
<ap:SPTL_Loright_Y_Boundingcoordinate>
  5412
</ap:SPTL_Loright_Y_Boundingcoordinate>
<ap:SPTL_Grid_Coordinate_System>
  Over text UP, running horizontally starting at 0:0 at upper left
  corner of image
</ap:SPTL_Grid_Coordinate_System>
<ap:SPTL_X_Posn_Accuracy>1 pixel</ap:SPTL_X_Posn_Accuracy>
<ap:SPTL_Y_Posn_Accuracy>1 pixel</ap:SPTL_Y_Posn_Accuracy>
<ap:SPTL_Count_Horizontal_Regions_On_Folio>
  1
</ap:SPTL_Count_Horizontal_Regions_On_Folio>
<ap:SPTL_Count_Vertical_Regions_On_Folio>
  1
</ap:SPTL_Count_Vertical_Regions_On_Folio>
<ap:SPTL_Count_Total_Regions_On_Folio>
  1
</ap:SPTL_Count_Total_Regions_On_Folio>
<ap:SPTL_Position_Number_Folio_Region>
  0
</ap:SPTL_Position_Number_Folio_Region>
<ap:IMG_Spectral_Range>350-1100</ap:IMG_Spectral_Range>
<ap:IMG_Spectral_Range_Unit_Of_Measure>
  nm
</ap:IMG_Spectral_Range_Unit_Of_Measure>
<ap:IMG_Illumination_Filters>
  Acrylic prismatic diffusers common to all light sources left and right
</ap:IMG_Illumination_Filters>
<ap:IMG_Acquisition_Filters>
  Infrared filter removed from camera
</ap:IMG_Acquisition_Filters>
<ap:IMG_Imaging_System>
  E6 SN:31R094428200/Lens: APO-DIGITAR 5,6/120 M-16
</ap:IMG_Imaging_System>
<ap:IMG_Lens_Brand>Schneider</ap:IMG_Lens_Brand>
```

```
<ap:IMG_Lens_Focal_Length>120</ap:IMG_Lens_Focal_Length>
<ap:IMG_Lens_Focal_Length_Measurement_Unit>
  nm
</ap:IMG_Lens_Focal_Length_Measurement_Unit>
<ap:IMG_Sensor_Spectral_Range>
  350-1100
</ap:IMG_Sensor_Spectral_Range>
<ap:IMG_Sensor_Spectral_Range_Unit_Of_Measure>
  nm
</ap:IMG_Sensor_Spectral_Range_Unit_Of_Measure>
<ap:IMG_Camera_Incidence_Angle_Deg>
  0.0
</ap:IMG_Camera_Incidence_Angle_Deg>
<ap:IMG_Camera_Imaging_Depth_Bits>16</ap:IMG_Camera_Imaging_Depth_Bits>
<ap:IMG_Illumination_Wavelength>365</ap:IMG_Illumination_Wavelength>
<ap:IMG_Illumination_Wavelength_Unit_Of_Measure>
  nm
</ap:IMG_Illumination_Wavelength_Unit_Of_Measure>
<ap:IMG_Illumination_Source_Wattage>
  2 X 11.2
</ap:IMG_Illumination_Source_Wattage>
<ap:IMG_Illumination_Type>LED</ap:IMG_Illumination_Type>
<ap:IMG_Illumination_Spectral_Range>
  355-375
</ap:IMG_Illumination_Spectral_Range>
<ap:IMG_Illumination_Spectral_Range_Unit_Of_Measure>
  nm
</ap:IMG_Illumination_Spectral_Range_Unit_Of_Measure>
<ap:IMG_White_Balance>none</ap:IMG_White_Balance>
<ap:IMG_Illumination_Source_No>2</ap:IMG_Illumination_Source_No>
<ap:IMG_Illumination_Incidence_Angle_Az_0>
  43.6055030767404
</ap:IMG_Illumination_Incidence_Angle_Az_0>
<ap:IMG_Illumination_Incidence_Angle_Az_180>
  44.7527324496602
</ap:IMG_Illumination_Incidence_Angle_Az_180>
<ap:DAT_File_Type>TIFF</ap:DAT_File_Type>
<ap:DAT_Format_Version_Number>6.0</ap:DAT_Format_Version_Number>
<ap:DAT_Format_Version_Date>1992-06-03</ap:DAT_Format_Version_Date>
<ap:DAT_Decompression_Technique>None</ap:DAT_Decompression_Technique>
<ap:DAT_Compression_Technique>Uncompressed</ap:DAT_Compression_Technique>
<ap:DAT_File_Size>38138</ap:DAT_File_Size>
<ap:MET_Set_Id>1</ap:MET_Set_Id>
```

```
<ap:SPTL_Coordinate_Unit_Of_Measure>
  mm
</ap:SPTL_Coordinate_Unit_Of_Measure>
<ap:DAT_File_Processing>
 <rdf:Bag>
   <rdf:li>Processing Type 1: All images were flattened with a corresponding image
of white surface. The flats were smoothed by blurring with a gaussian-like filter,
normalized to have a maximum value of unity and divided into the corresponding
spectral image.</rdf:li>
    <rdf:li>Processing Type 2: A linear contrast stretch applied to the
      16-bit single-wavelength images. The black and white values were set 3 standard
deviations away from the average value. The values beyond 3 standard deviations were
clipped to black or white.
    </rdf:li>
  </rdf:Bag>
</ap:DAT_File_Processing>
<ap:DAT_Joining_Same_Parts_Of_Folio>
 <rdf:Bag>
   <rdf:li>Processing Type 1: No</rdf:li>
   <rdf:li>Processing Type 2: No</rdf:li>
 </rdf:Bag>
</ap:DAT_Joining_Same_Parts_Of_Folio>
<ap:DAT_Type_Of_Contrast_Adjustment>
 <rdf:Bag>
   <rdf:li>Processing Type 1: globally adjusted</rdf:li>
   <rdf:li>Processing Type 2: globally adjusted</rdf:li>
 </rdf:Bag>
</ap:DAT_Type_Of_Contrast_Adjustment>
<ap:DAT_Type_Of_Image_Processing>
 <rdf:Bag>
   <rdf:li>Processing Type 1: linear stretch</rdf:li>
   <rdf:li>Processing Type 2: linear stretch</rdf:li>
 </rdf:Bag>
</ap:DAT_Type_Of_Image_Processing>
<ap:DAT_Software_Version>
 <rdf:Bag>
   <rdf:li>Processing Type 1: 1.2</rdf:li>
   <rdf:li>Processing Type 2: 1.2</rdf:li>
 </rdf:Bag>
</ap:DAT_Software_Version>
<ap:DAT_Processing_Program>
 <rdf:Bag>
   <rdf:li>Processing Type 1: Archie 1.2, rect, div</rdf:li>
```

```
    <rdf:li>Processing Type 2: Archie 1.2, packimage</rdf:li>
   </rdf:Bag>
  </ap:DAT_Processing_Program>
  <ap:DAT_Processing_Comments>
   <rdf:Bag>
    <rdf:li>Processing Type 2: For viewing purposes only</rdf:li>
   </rdf:Bag>
  </ap:DAT_Processing_Comments>
  <ap:CONT_Content_Keyword>
   <rdf:Seq>
    <rdf:li>Syriac Palimpsest</rdf:li>
    <rdf:li>Palimpsest</rdf:li>
    <rdf:li>Syriac Manuscript</rdf:li>
    <rdf:li>Private Collection</rdf:li>
    <rdf:li>Medical Palimpsest</rdf:li>
    <rdf:li>Troparia Image</rdf:li>
    <rdf:li>11th Century Manuscript</rdf:li>
    <rdf:li>8th-9th Century Manuscript</rdf:li>
    <rdf:li>Parchment Manuscripts</rdf:li>
    <rdf:li>Galen Manuscript</rdf:li>
    <rdf:li>Medical Manuscript</rdf:li>
   </rdf:Seq>
  </ap:CONT_Content_Keyword>
  <ap:CONT_Source_Info>
   <rdf:Seq>
    <rdf:li>Palimpsest Leaf </rdf:li>
   </rdf:Seq>
  </ap:CONT_Source_Info>
  <ap:CONT_Foliation_Scheme>upper text</ap:CONT_Foliation_Scheme>
  <ap:CONT_Folio1_Number>1</ap:CONT_Folio1_Number>
  <ap:CONT_Folio1_R_V>r</ap:CONT_Folio1_R_V>
  <ap:CONT_Folio2_Number>4</ap:CONT_Folio2_Number>
  <ap:CONT_Folio2_R_V>v</ap:CONT_Folio2_R_V>
  <ap:CONT_Source_Citation>
   <rdf:Bag>
    <rdf:li>Syriac Liturgical Text with Medical Palimpsest, Private Collection
(Hiersemann Katalog 500, 1922, Nr.20)</rdf:li>
   </rdf:Bag>
  </ap:CONT_Source_Citation>
  <ap:CONT_Language>
   <rdf:Bag>
    <rdf:li>Syriac</rdf:li>
   </rdf:Bag>
```

```
     </ap:CONT_Language>
     <ap:MET_Metadata_Status>Valid</ap:MET_Metadata_Status>
     <ap:MET_Metadata_Date>2006-06-07</ap:MET_Metadata_Date>
     <ap:MET_Review_Date>2010-07-06</ap:MET_Review_Date>
     <ap:MET_Metadata_Contact>
       <rdf:Seq>
         <rdf:li>Emery, Doug</rdf:li>
         <rdf:li>Toth, Michael B.</rdf:li>
       </rdf:Seq>
     </ap:MET_Metadata_Contact>
     <ap:MET_Standard_Name>Archimedes Palimpsest Metadata
       Standard</ap:MET_Standard_Name>
     <ap:MET_Standard_Version>1.1</ap:MET_Standard_Version>
     <ap:MET_Extensions>
       <rdf:Bag>
         <rdf:li>none</rdf:li>
       </rdf:Bag>
     </ap:MET_Extensions>
     <ap:DAT_Joining_Different_Parts_Of_Folio>
       <rdf:Bag>
         <rdf:li>Processing Type 1: No</rdf:li>
         <rdf:li>Processing Type 2: No</rdf:li>
       </rdf:Bag>
     </ap:DAT_Joining_Different_Parts_Of_Folio>
   </rdf:Description>
   <rdf:Description rdf:about="" xmlns:tiff="http://ns.adobe.com/tiff/1.0/">
     <tiff:ImageWidth>7216</tiff:ImageWidth>
     <tiff:ImageLength>5412</tiff:ImageLength>
     <tiff:BitsPerSample>16</tiff:BitsPerSample>
     <tiff:Compression>1</tiff:Compression>
     <tiff:SamplesPerPixel>1</tiff:SamplesPerPixel>
     <tiff:ResolutionUnit>3</tiff:ResolutionUnit>
     <tiff:XResolution>239.76</tiff:XResolution>
     <tiff:YResolution>239.76</tiff:YResolution>
     <tiff:Make>MegaVision</tiff:Make>
     <tiff:Model>S/E6</tiff:Model>
   </rdf:Description>
   <rdf:Description rdf:about="" xmlns:xmp="http://ns.adobe.com/xap/1.0/">
     <xmp:ModifyDate>2010-05-08 12:53:21</xmp:ModifyDate>
     <xmp:CreatorTool>Processing Type 1: Archie 1.2, rect, div, Processing
       Type 2: Archie 1.2, packimage</xmp:CreatorTool>
     <xmp:CreateDate>2010-05-08 12:53:21</xmp:CreateDate>
     <xmp:MetadataDate>2006-06-07</xmp:MetadataDate>
```

```
    </rdf:Description>
    <rdf:Description rdf:about="" xmlns:exif="http://ns.adobe.com/exif/1.0/">
      <exif:PixelXDimension>7216</exif:PixelXDimension>
      <exif:PixelYDimension>5412</exif:PixelYDimension>
      <exif:DateTimeOriginal>2010-05-08 12:53:21</exif:DateTimeOriginal>
      <exif:DateTimeDigitized>2010-05-08 12:53:21</exif:DateTimeDigitized>
      <exif:ExposureTime>10</exif:ExposureTime>
      <exif:FNumber>11.0</exif:FNumber>
      <exif:FocalLength>120</exif:FocalLength>
      <exif:FocalPlaneXResolution>239.76</exif:FocalPlaneXResolution>
      <exif:FocalPlaneYResolution>239.76</exif:FocalPlaneYResolution>
      <exif:FocalPlaneResolutionUnit>3</exif:FocalPlaneResolutionUnit>
    </rdf:Description>
  </rdf:RDF>
</x:xmpmeta>
```