

October 2018

## Preserving the Layers: Storage and Metadata for Geographic Information Systems

Jonathan D. Cartledge  
*Smith College, jcartled@smith.edu*

Follow this and additional works at: <https://online.vraweb.org/vrab>

---

### Recommended Citation

Cartledge, Jonathan D. (2018) "Preserving the Layers: Storage and Metadata for Geographic Information Systems," *VRA Bulletin*:Vol. 45: Iss. 1, Article 5.  
Available at: <https://online.vraweb.org/vrab/vol45/iss1/5>

This Feature Articles is brought to you for free and open access by VRA Online. It has been accepted for inclusion in VRA Bulletin by an authorized editor of VRA Online.

---

# Preserving the Layers: Storage and Metadata for Geographic Information Systems

## **Abstract**

This paper is an introduction to maps created with Geographic Information Systems, or, GIS. It describes the makeup and uses of these electronic maps and goes on to describe how this material is cataloged, utilizing CSDGM and ISO 19115 standards. The paper concludes by chronicling a couple of pilot projects conducted at Smith College to store, catalog and retrieve these files.

## **Keywords**

GIS, cataloging, standards

## **Author Bio & Acknowledgements**

Jon Cartledge is the Digital Imaging Specialist at Smith College. He has a B.A. in Anthropology, an MFA in Printmaking, and a masters in Library Science. He has presented at both regional and national VRA conferences and recently finished three terms as the secretary of the Visual Resource Association of New England. He is also a printmaker and a former board member of the Boston Printmakers. He has shown internationally and his work is in the collections of the University of Arkansas and the University of Cincinnati.

## Introduction

Geographic Information Systems, or, a GIS map product, is a digital map of geographic or spatial features, attached to a database of features and their attributes. Each row of the database corresponds to a feature on a map. Features are placed on a layer, much like a Photoshop layer, and a combination of layers allows you to build a visual representation of a geospatial phenomenon – for example, a correlation of soil types and vegetation across the country, or access to public transportation lines in low-income areas. These data features come from a wide variety of sources including satellite imagery, historical maps, or a researcher entering locations into a GPS unit. GIS software such as ArcGIS or QGIS allows you to import those features and assemble them into data layers. In this paper I will introduce the reader to maps created with GIS, describe how this material is cataloged, and a couple of pilot projects conducted at Smith College to store, catalog, and retrieve this information.

Imagine you have a set of transparencies. The first might be a standard roadmap – which will act as our base map. The next layer will be a set of coordinate points of all of the restaurants and bars I frequent – creating what is known as a vector layer (a layer made of only points or lines, no color gradations of any kind). The final layer will trace my route from Smith College to the pub in a line, creating another vector layer (figure 1).

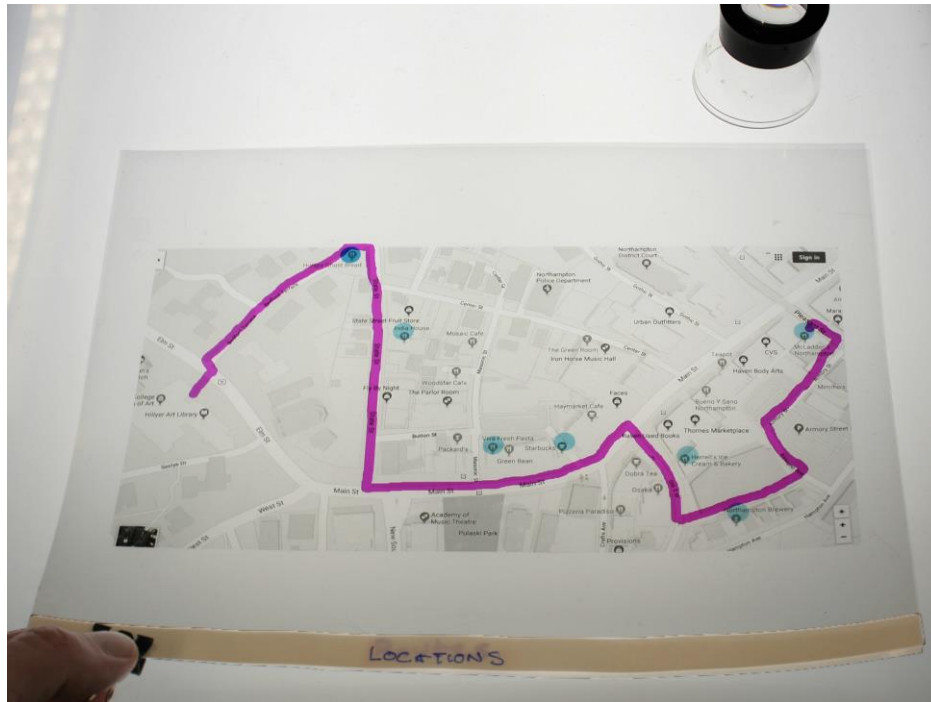


Figure 1

These layers are stored with a couple of important pieces of information. The first is something we're all familiar with from VRA Core & CCO – the coordinates – letting the software know where on the globe the map corresponds. The second is its projection. There are a number of ways you can turn a spherical (and not even perfectly spherical, at that) globe into a flat map. Each one has a set of algorithms that GIS software uses to create your map on the screen – but you need to select one, and that selection is stored with your map layers so the machine can properly read it (figure 2).

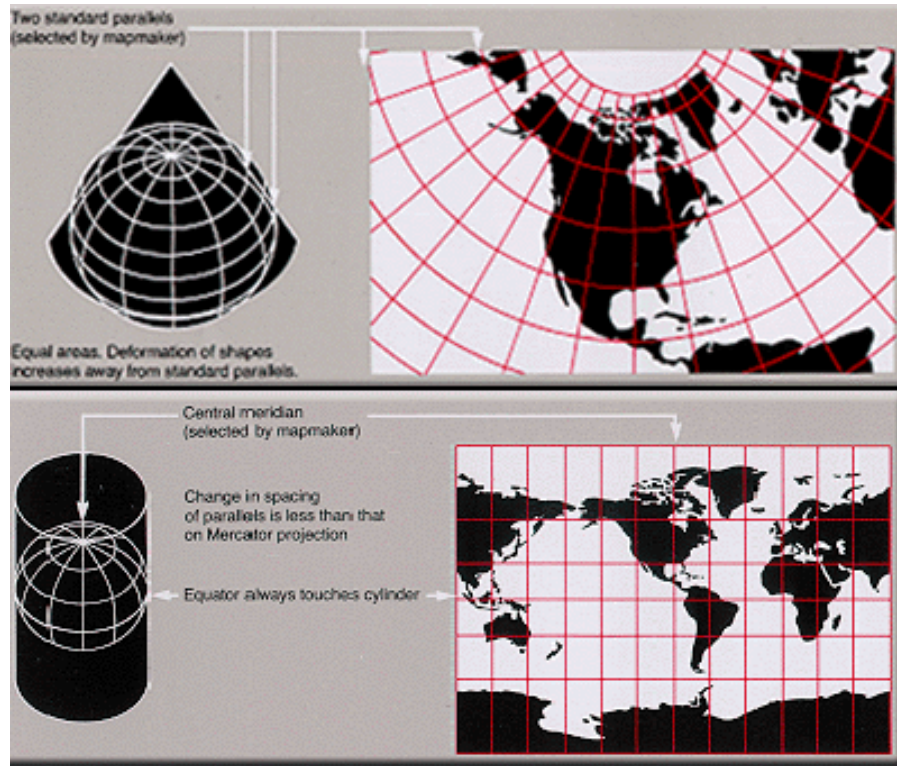


Figure 2

Each of these data layers can be re-used or altered for a new project in GIS software. The authorship is attached to each layer of features, but those layers can be reused in any number of combinations in GIS software. Unlike a scanned artwork, GIS maps are designed to be a changeable asset.

My interest in the cataloging and retrieval of GIS data is the result of our Art History faculty's ongoing interest in maps. The history of cities, medieval pilgrimages, Roman architecture, and Mesoamerican culture are just a few of our course topics that rely heavily on map imagery in the Smith classroom. For instance, we have fielded scanning requests for multiple details of Nolli's map of Rome. So, it makes sense that we'd be interested in the University of Oregon's GIS version of Nolli's map. Along with this topical interest, the field of art history has begun to utilize these tools for analysis, as can be seen in a project by Greg Prickman at the University of Iowa on the expansion of printing in Europe. GIS has begun to take hold in art history research. So, with an interest in giving our students the experiences they require to succeed in art history, our department has begun to explore its uses in the classroom, and I have begun to acquaint myself with the ways these files are created, cataloged, stored, and retrieved.

## Cataloging

In the United States, the most widely used standard for GIS data is the Federal Geospatial Data Committee's Content Standard for Digital Geospatial Metadata. The FDGC provides a great deal of documentation on the standard at <https://www.fgdc.gov/metadata/csdgm-standard>, including full technical

specifications, quick guides, and required element boilerplates. The standard has some similarities to visual resource cataloging standards, but, while the VRA Core and CCO focus on the question: what is the object being cataloged?, the CSDGM focuses more on who is responsible and how accurately the object was made.

There are seven general categories of elements within the CSDGM:

**Identification** – these elements are what the Visual Resource Association community would consider standard cataloging data: title, originator, date, descriptive keywords, and use constraints (copyright). Identification also features a couple of elements that reflect the changeable nature of the asset – Progress (with fixed values of Complete, In Work and Planned) and Currentness\_Reference (a free text description of when the data were gathered vs. when published).

**Data Quality** – Data quality elements (Positional Accuracy, Completeness) describe how correct the data is. Was there an independent field check of points, did the originator run a software to check for bad values, is there data that is missing, and what was the process used to create the asset? Often these are free-text elements that describe, for instance, the margin of error between the actual coordinate location and the map or the selection criteria for what is represented on the map.

**Spatial Data Organization** – The type of data used (vector, raster). Data is represented in two ways in GIS files. Vector data is point data that can act as individual points, or create lines or polygons – the kind of data you might see on an architectural floor plan. Raster data is pixel data which can describe changes in data with value and hue. A map displaying different elevations areas in different colors would use raster presentation.

**Spatial Reference**- Information about the coordinate and projection system used (see above) and how precise the resolution of the asset is, that is to say are your points of data accurate to the nearest couple of centimeters, meters or kilometers?

**Entity and Attributes** – A description of the asset database, with an explanation of the values you used, published codesets you might have used, and any non-standardized values that need explanation for the viewer.

**Distribution Information** – contact and liability information of the distributor.

**Metadata reference** – metadata standard used (in this case, CSDGM) and contact information for the standard (which would be the FDGC).

These data live in an XML document associated with each layer of a GIS map. Geospatial data are input by the device used to collect the data (typically a GPS unit) and descriptive data is input by the creator.

In the past few years, the geospatial community has been developing a new international data standard, ISO 19115. Created by Technical Committee 211, ISO 19115 is part of a suite of XML data standards for GIS objects. ISO 19115 NAP (North American Profile) and its XML format guide ISO 19139, are relational (rather than having CSDGM's flat structure), with a set of 12 categories to describe a digital map. Crosswalks have been provided by FDGC to move the data from the CSDGM standard, which show the great similarity between the two (figure 3).

16	FGDC name	FGDC XML Tag	ISO name	Comment
17	<b>I. Identification Information</b>		<a href="#">See Note 11</a>	
18	1.1 Citation		MD_Metadata.identificationInfo.MD_DataIdentification.citation	
19	1.2 Description		---	Compound element, unnecessary mapping
20	1.2.1 Abstract	abstract	MD_Metadata.identificationInfo.MD_DataIdentification.abstract	
21	1.2.2 Purpose	purpose	MD_Metadata.identificationInfo.MD_DataIdentification.purpose	
22	1.2.3 Supplemental Information		MD_Metadata.identificationInfo.MD_DataIdentification.supplementalInformation	
23	1.3 Time Period of content		---	Compound element, unnecessary mapping
24	<a href="#">1.3.1 Time Period Information</a>		MD_Metadata.identificationInfo.MD_DataIdentification.extentEX_Extent.temporalElement.EX_TemporalExtent.gml:TimeInstant.timeposition	
25	1.3.2 Currentness Reference		MD_Metadata.identificationInfo.MD_DataIdentification.extentEX_Extent.description	
26	1.4 Status		---	Compound element, unnecessary mapping
27	1.4.1 Progress	progress	MD_Metadata.identificationInfo.MD_DataIdentification.status	
28	1.4.2 Maintenance and Update Frequency		MD_Metadata.identificationInfo.MD_DataIdentification.resourceMaintenance.MD_Maintenance.maintenanceAndUpdateFrequency	
29	1.5 Spatial domain		---	Compound element, unnecessary mapping
30	1.5.1 Bounding coordinates		MD_Metadata.identificationInfo.MD_DataIdentification.extentEX_Extent.geographicElement.EX_GeographicBoundingBox (where EX_Extent.extentTypeCode = 1)	
	1.5.1.1 West Bounding Coordinate		MD_Metadata.identificationInfo.MD_DataIdentification.extentEX_Extent.geographicElement.EX_GeographicBoundingBox.westB	

Figure 3

For instance, the *abstract* element (1.2.1) in the CSDGM Identification section corresponds to the *MD\_DataIdentification.abstract* (idAbs) attribute in the ISO 19115 Identification information section. Designed for XML, the standard is extensible and, like Dublin Core, few of the elements are required- so you can go deep in some areas and skip others altogether, depending on your data. Under the section MD\_Identification, for example, there are two required classes – citation and abstract, along with other non-required classes like credit, status, pointOfContact and more. Nested within the first required class, citation, are the attributes: title, date, edition, citedResponsibleparty, presentationForm and others. Only title is required for the citation class. The other required class, abstract is a free text class. So, within MD\_identification, you could include a wide variety of information like who worked on the resource, their contact information, status of the research or you could simply put in a title and a short, free text abstract. The ISO also uses standard codelists for the higher level descriptive data and to describe each element type. The Topic Categories attribute uses a prescribed set of general terms to classify the file (terms include economy, health, society, and transportation). In addition, the ISO prescribes standardized descriptors for some attributes. For the Contact Section, there is a set of descriptors for roles, very similar to the recommended Creator Roles controlled list outlined in CCO. This format allows for more flexibility than the CSDGM, but it also creates some standardization via these codelists.

In the same way that Photoshop offers Bridge to create embedded metadata for digital images, ArcGIS software offers ArcCatalog, which will, among other things, walk the user through the cataloging process of new layers with a metadata wizard. The wizard begins by asking the user to select a cataloging standard – offering the older CSDGM, the 2003 version of ISO 19115, and ISO 19139 - which is the current, verifiable XML version of ISO 19115. The user enters the title, description, summary, and descriptive terms - the software automatically populates the technical terms, including coordinates and projection information. The result exports as an XML document which is stored with the other files that create your layer. When you open it, you can see your information embedded in the XML code (figure 4).










 Migration Countries.dbf	OpenOffice.org XML 1.0 S...	2 KB
 Migration Countries.prj	PRJ File	1 KB
 Migration Countries.sbn	SBN File	1 KB
 Migration Countries.sbx	Adobe Illustrator Tsume F...	1 KB
 Migration Countries.shp	SHP File	107 KB
 Migration Countries.shp.xml	XML Document	25 KB
 Migration Countries.shx	SHX File	1 KB

Figure 4

This embedded metadata acts primarily as an identifier to the asset itself. It is not searchable by a Windows browser and even ArcCatalog requires the user to select the file to be indexed *before* the metadata can be searched by the program. While this feature is useful when you have a new selection of layers from a particular project, it becomes problematic for locating an older free layer that may have never been indexed. Why, then, is it necessary? Unlike typical visual resource embedded metadata, which describes the work that the image depicts, GIS metadata describes itself and its authorship – and this function is crucial as layers are commonly reused in multiple projects by multiple users. It is entirely possible a user may obtain a layer for use, not from the author, but from a project several iterations from the original use. In this case, the only way for the user to know the subject, authorship, rights, and methods is to examine the metadata.

### Pilot Projects in Storage and Retrieval

Several years ago, I approached Smith College's Spatial Analysis Lab about a collaboration. I was interested to see if we could store GIS material in our former image database, LUNA. As the ArcGIS software does allow you to store and access your layers and map projects, either on a local storage tree or online, there was a question why storage in a digital asset database, with a metadata specialist overseeing production, is desirable. In our discussions, Jon Caris, the director of the SAL, outlined three reasons:

**Archival** – The primary function of ArcGIS is to create and repurpose map products, so ALL the layers that are accessible through ArcGIS are subject to re-use and continued development. This makes for a problematic archival storage environment. It would be similar to using Photoshop to store and retrieve finished TIFF images. By storing a finished package in another location, in this case LUNA, we make sure the original document is preserved – though students can download copies for use in ArcGIS.

**Cost** – ArcGIS' manufacturer, ESRI, does charge for storage space online – so long-term storage may not be the best option. ArcGIS does have a longer term storage software, ArcGIS Server, but this requires a database administrator to run it effectively, and at a cost.

**Metadata Quality** – The final reason is something we are all familiar with – the enforcement of data standards and richer, searchable metadata. As described in previous pages, ArcCatalog does feature a wizard to populate embedded metadata. But this data is not searchable within ArcGIS without user-initiated indexing, or even within a connected local storage tree. In addition, while ArcCatalog supports metadata standards, it is up to the user to enter complete and accurate data. Given the challenges of day-to-day research, this work often falls to the bottom of the to-do list. A separate storage solution offers more searchable metadata and a metadata specialist ensuring that standards are met.

For our first pilot, the SAL suggested a project to store some traditional map image scans, in particular, a set of historical maps of the Smith Campus, as well as a single GIS layer set. Our goal was the storage of map scans and finished GIS projects for students to refer to as they developed their own projects.

I worked with Jon Caris, SAL Director, and our LUNA administrator, Dan Bridgman, to come up with a usable data schema utilizing the CSDGM in LUNA to describe these objects in a relatively user-friendly and searchable way. Jon indicated that many of the CSDGM elements were required, so it would be a challenge to find a way to present all these data in a readable way. This was complicated by the fact that, the CSDGM schema reflects metadata for a single GIS layer – and a GIS map object is often made up of multiple layers. What we came up with is a schema that groups elements into discrete categories. The elements are repeatable, with an exclamation mark delimiter, to allow for multiple entries for multiple layers. If data were missing, or not relevant, the element heading would not appear. This allowed for paper maps, whose metadata falls only under the Identification category of CSDGM, to be included in the database.

The storage of the assets presented another challenge. Most digital access management systems cannot store or read GIS files; for LUNA, a ZIP file was required to package all the GIS layers. One reason for this incompatibility is that GIS software stores its object, not as a single file, but in parts used to create the map. It's as if Photoshop stored all the edit layers for an image as separate files. These parts include, among others, the shapefile (.shp) which stores the vector points, the shapefile's index (.shx), the database of features (.dbf), the projection information (.prj), and the metadata (.xml). (figure 5) The ZIP file is necessary to keep the components together and allow the material to be uploaded to LUNA. With this solution, the material could be archived and searched in LUNA, but it could not be viewed. What was needed was a storage platform that could put together the parts and display them while storing them in an archival and searchable manner.

One application that meets these criteria is an open-source web application called Geoserver. This storage and retrieval application runs in Java 8, serves out via Apache/Tomcat, and allows GIS objects to be stored and cataloged in an archival manner. The metadata entered is searchable, and the user can preview and download a copy of each GIS layer. Geoserver divides itself into three basic components: Workspaces, Stores, and Layers.

1. Workspaces are essentially a namespace or folder where you can save related stores and layers. So, if the cataloger had a number of stores that related to a particular topic, they can organize them in a single space.
2. Stores tell Geoserver where to locate a particular Layer file (in Smith's case a folder tree) and how to process it. Each Store connects to a single Layer.
3. Layers is where the actual GIS layer is stored.

Once you connect the layers to GeoServer, you can add an abstract and cataloging terms, as well as a reference that refers back to a thesaurus source for each cataloging term. The rudimentary versions of your layers can then be served up through Layers Display, allowing you to preview the file, adjust the resolution, and zoom. It will also download the file as a KML object – a GIS protocol that allows you to use the file in concert with Google Earth.



In the past year, the Spatial Analysis Lab installed GeoServer on a desktop machine as a test, so in the summer of 2017, the Imaging Center and the SAL put together another pilot. In this case, I'd be georeferencing and cataloging the archaeological maps of Liz Klarich, Smith's professor of archaeology. ISO 19115 (the XML version, 19139) would be used for the embedded metadata of a georeferenced file, and that information would be copied to the Geoserver cataloging interface. This revealed the first problem, which is that Geoserver, like many digital storage solutions, is unable to pull out the embedded metadata from a file, so it must be re-entered into the database. This issue occurs not only with identification information, but also with some technical metadata, like the projection. Failure to re-enter this data will, as I learned the hard way, create a number of image display problems. An issue that an end user will face is access to a preview image. Once the user finds the layer they may want, they are confronted with a number of viewing options that requires an experienced GIS user to know which type file is desirable for preview and what the local machine will allow. Finally, there was the issue of connecting a new file to the database, requiring the cataloger to know which drive and which folder in the folder tree to store the GIS files so they could be connected to the database. For this pilot, this was an issue easily remedied, but for an active program of cataloging and storage, it could quickly become a problem. By in large, the problems with storage and ease of accessibility can be remedied with two related software packages that are designed to work in concert with GeoServer: PostGIS for storage and GeoNetwork for searching, viewing and cataloging. PostGIS is an XML database that will store your GIS data. Rather than a local folder tree, the data is stored in an XML container that is easily connected to GeoServer. For access, GeoNetwork will create a web-accessible interface to GeoServer that will allow users to easily search and view stored layers in CSS-like styles. In addition, it offers a web-accessible cataloging tool for data entry. It remains to be seen (by this writer) how well GeoNetwork interoperates with the cataloging in GeoServer and whether it can access the embedded metadata, but we plan to test this functionality over the summer, with the hope of presenting a more complete picture of GIS cataloging at the VRA 2019 conference in Los Angeles.

## **Conclusion**

Since beginning my study of GIS and its cataloging, I was fortunate to partner with Jon Caris and Tracy Tien from Smith College's Spatial Analysis Lab to conduct a pilot of GIS for art history study in Professor Brigitte Buettner's Medieval Pilgrimages course. Students Xiyang Deng and Sophie Lei had the opportunity to create a map in ArcGIS, based on Theoderich's account of a Holy Land pilgrimage, and present it to the class in ArcGIS's StoryMaps online tool for their final project (figures 5 & 6).



Figure 5

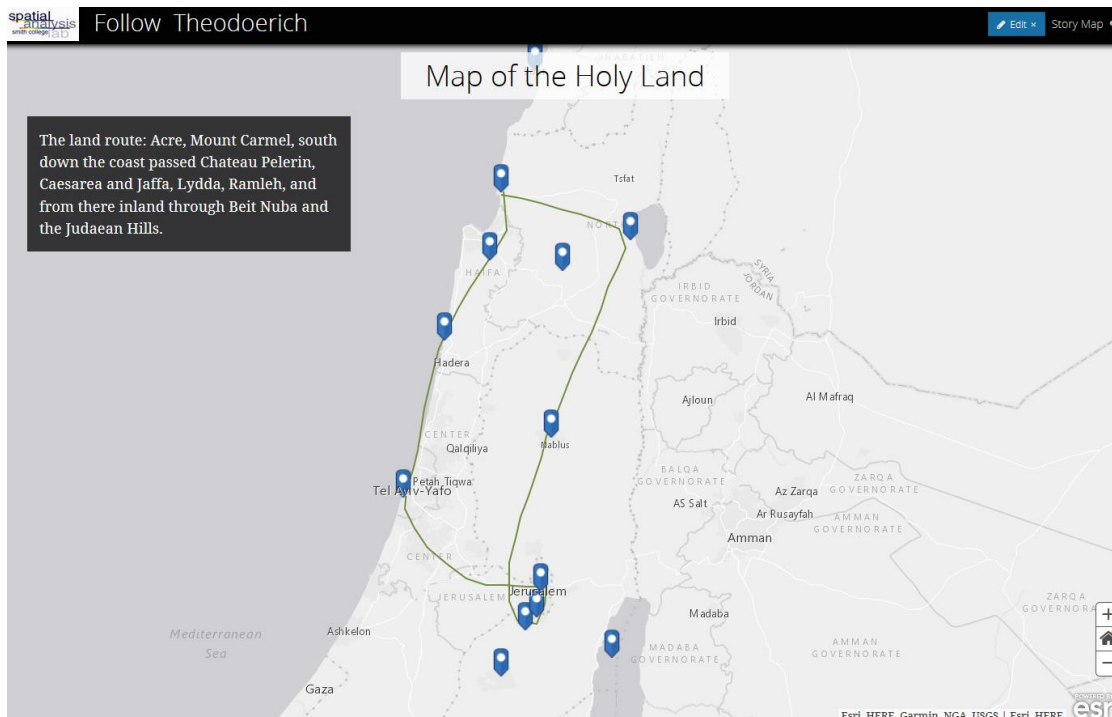


Figure 6

During the course of my preparation, I came across the Harvard Digital Atlas of Roman and Medieval Civilizations – which allows users to download multiple GIS layers of cited data on the roads, ports, kingdoms – even Viking raid routes for the Roman and Medieval periods. The data primarily

focused on the west, rather than the east of Mediterranean, so Ms. Deng and Ms. Lei's project could not incorporate it, but this resource showed me the wealth of information and possibilities available for future GIS projects, including the archiving of past faculty and student maps for future study and project use, and the need for the visual resource profession to acquaint themselves with the functionality, standards and storage & retrieval options for GIS mapping to help our institutions make the most of this growing technology.

## Bibliography

Caris, Jon. "ArcGIS Online Basic Workshop." Jan. 2015. 180 Introduction to Geographic Information Systems, Smith College. Microsoft PowerPoint presentation.

Caris, Jon. "ArcGIS Online Workshop: Basic." 180 Introduction to Geographic Information Systems, Smith College. Received January 2015. Course handout.

Federal Geographic Data Committee. "CSDGM Essential Metadata Elements." *FDGC.gov*, last modified May 2008, accessed May 2013. [https://www.fgdc.gov/metadata/documents/CSDGMEssentialMeta\\_20080514.pdf](https://www.fgdc.gov/metadata/documents/CSDGMEssentialMeta_20080514.pdf).

Federal Geographic Data Committee. "Content Standard for Digital Geospatial Metadata Workbook." *FDGC.gov*. last modified May 2000, accessed May 2018. [https://www.fgdc.gov/metadata/documents/workbook\\_0501\\_bmk.pdf](https://www.fgdc.gov/metadata/documents/workbook_0501_bmk.pdf).

Federal Geographic Data Committee. "Geospatial Metadata Fact Sheet." *FDGC.gov*. last modified July 2011, accessed May 2018. <https://www.fgdc.gov/resources/factsheets/documents/GeospatialMetadata-July2011.pdf>.

Federal Geographic Data Committee. "Geospatial Metadata Quick Guide." *FDGC.gov*. last modified April 2006, accessed May 2018. <https://www.fgdc.gov/metadata/documents/MetadataQuickGuide.pdf>.

Federal Geographic Data Committee. "Preparing for International Metadata." *FDGC.gov*. last modified October 2011, accessed June 2018. <https://www.fgdc.gov/metadata/documents/preparing-for-international-metadata-guidance.pdf>

Iacovella, Stefano. *GeoServer Beginner's Guide, 2<sup>nd</sup> ed.* Birmingham, Packt Publishing, 2017. Kindle.

International Standard Organization. *International Standard ISO 19115-1*. 1st ed., Geneva, International Standard Organization, 2014. PDF.

Law, Michael and Amy Collins. *Getting to Know ArcGIS*, 4<sup>th</sup> ed., Redlands, ESRI Press, 2015.

Loveless, Jack. "Module 3: Coordinate Systems, Datums, and Map Projections I." Modeling Our World: An Introduction to Geographic Information Systems GEO/ENV 150, Smith College. Received October 2015. Course handout.

Loveless, Jack. "Module 3: Coordinate Systems, Datums, and Map Projections II." Modeling Our World: An Introduction to Geographic Information Systems GEO/ENV 150, Smith College. Received October 2015. Course handout.

McCormick, Michael, Eurydice Georganteli, Leland Grigoli, Alexander More, Giovanni Zambotti and Lex Berman. "The Digital Atlas of Roman and Medieval Civilizations," Harvard University. last modified 2016, accessed December 2017. <https://darmc.harvard.edu/>

“ISO 19115 Geographic information Metadata Workbook Guide to Implementing ISO 19115:2003(E), the North American Profile (NAP), and ISO 19110 Feature Catalogue.” National Oceanic and Atmospheric Administration. last modified January 2012, accessed May 2018. [https://www.ncddc.noaa.gov/metadastandards/ftp://ftp.ncddc.noaa.gov/pub/Metadata/Online\\_ISO\\_Training/Intro\\_to\\_ISO/workbooks/MD\\_Metadata.pdf](https://www.ncddc.noaa.gov/metadastandards/ftp://ftp.ncddc.noaa.gov/pub/Metadata/Online_ISO_Training/Intro_to_ISO/workbooks/MD_Metadata.pdf).

Prickman, Greg. “The Atlas of Early Printing,” University of Iowa, last modified 2013, accessed March 2018, <http://atlas.lib.uiowa.edu/>

SDTS Task Force of the U.S. Geological Survey. “The Spatial Data Transfer Standard, a Guide for Technical Managers” *USGS.gov*, last modified 1995, accessed September 2018. <https://thor-f5.er.usgs.gov/sdts/articles/pdf/mgrs.pdf>

Tice, Jim and Erik Steiner. “The Nolli Map Website,” University of Oregon. last modified 2018, accessed June 2018. <http://nolli.uoregon.edu/>.

"GeoServer User Manual." *GeoServer*. last modified May 2013, accessed Sept. 2017. <http://docs.geoserver.org/latest/en/user/>.

“Shapefile File Extensions,” *ArcGIS Resource Center: Desktop 10*. last modified February 2011, accessed June 2018. [http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Shapefile\\_file\\_extensions/0056000000300000/](http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/Shapefile_file_extensions/0056000000300000/).