Accessing Hard Maps and Associated Digital Components: A Case Study of the U.S. Library of the Marine Corps

by

Samantha J. Tucker

A Thesis Presented to the
Faculty of the USC Graduate School
University of Southern California
In Partial Fulfillment of the
Requirements for the Degree
Master of Science
(Geographic Information Science and Technology)

May 2017

To Emma C. Tucker, my amazing wife and climbing partner

Table of Contents

List of Figures

List of Tables

Acknowledgements

Dr. Kemp thank you for helping me find the “nugget”. Without finding the nugget, I would not be seeing the light at the end of the tunnel for writing this thesis. Dr. Hastings thank you for connecting me to databases and the wonderful world present beneath the deck. Dr. Chiang thank you for the introduction to the connection between GIS and historical maps. Dr. Swift thank you for the positive encouragements, early morning calls, and more positive encouragements. Each professor I had at USC taught me about the different aspects of GIScience, and it is exciting to know that it is so much more than a system.

Thank you to my wife who encourages me to be the best I can. Thank you to my family, my military family, and friends for listening to my project and believing in me.

List of Abbreviations

AK	Actual key
ALA	American Library Association
BCE	Before Common Era
ebRIM	e-business XML Registration Information Model
ECS	NASA EOS Core System
eLibrary	Electronic library
ENVI	Environment for Visualizing Images
FGDC	Federal Geographic Data Committee
FK	Foreign key
GB	Gigabyte
GIF	Graphics interchange format
GIS	Geographic information system
GISci	Geographic information science
GRASS	Geographic Resources Analysis Support System
ISO	International Standards Organization
IT	Information technology
JPEG	Joint photographic experts group
m	Meter
MAGIRT	Map and Geospatial Information Round Table
MB	Megabyte
NASA	National Aeronautics and Space Administration
NYU	New York University

OCR	Optical character recognition
OCUL	Ontario Council of University Libraries
OGC	Open Geospatial Consortium
OWS	Open Geospatial Consortium Web Services
PK	Primary key
PNG	Portable network graphic
ppi	Pixels per inch
TIFF	Tagged image file format
UDDI	Universal Description, Discovery, and Integration
UNC	University of North Carolina at Chapel Hill
URL	Uniform resource locator
U.S.	United States
USC	University of Southern California
XML	Extensible Markup Language

Abstract

Maps provide information to researchers, as each map represents the environment from the perspective of the creator, at a set point in time. Libraries act as sources of information for researchers and can provide resources for users to locate, request, and receive maps. However, each library has differing capacities for users to access maps, and that capacity can even differ within the same organization. The United States (U.S.) Marine Corps is an example of an organization that has libraries with differing levels of accessibility for users accessing maps. At this time, remote users do not have access to hard, or paper maps, in the Library of the Marine Corps located in Quantico, Virginia. Therefore the aim of this thesis is to design an innovative process for remote users to access hard maps from the Library of the Marine Corps using digital components. To design a suitable process and spatial database, this project focuses on how users currently search and access hard maps and digital components from the Library of the Marine Corps, what the metadata standards for geospatial data and cartography currently consist of, common access capabilities of other libraries, and current practices for scanning, georeferencing, and extracting vector features from hard maps. A detailed entity-relationship model illustrates an efficient spatial database that can accompany the process, revealing the relationships between different spatial objects involved in users locating, requesting, and receiving digitized versions of hard maps and associated digital components from the Library of the Marine Corps. Finally, this project evaluates the spatial database and process for integrity and suitability for the Library of the Marine Corps. Other libraries housing important historical as well as current collections of hard maps could use the results of this research.

Chapter 1 Introduction

Maps are important sources of information for showing how a map designer perceives the environment at a specific point in time. This information provides researchers with information about the past, which assists in understanding the present or making predictions about the future. Maps are accessible from a variety of locations and libraries provide one of the easily accessible points for researchers. Each library has different capacities for researchers locating, requesting, and receiving a copy of a hard, or paper copy, map. Even libraries within the same organization can have different capacities for user access. The Library of the Marine Corps, located in Quantico, Virginia, is an organization with an extensive hard map collection and different levels of accessibility for users requiring access to hard maps.

This project has two primary objectives to assist the Library of the Marine Corps with allowing users to access versions of a hard map from a remote location. The first primary objective is to design a spatial database capable of meeting the requirements of different users and the library. To meet this objective, there are two secondary objectives. The first is to design a robust spatial database that captures metadata about a hard map and allows for collecting, storing, and downloading each hard map in various digital formats, such as an image or vector file. The second objective is to structure the format of the spatial database as an electronic library (eLibrary) allowing for terminology, spatial, and temporal searches of hard maps and digitized components of maps. The second primary objective is to propose a process for utilizing the spatial database. The process should explain how the spatial database interacts with the different stages of a request for a copy of a hard map in a digital format. The process explains various methodologies for scanning a hard map and how the spatial database archives obtained data about a map. Finally, the process should explain the extraction methodologies for spatial data and map features from a scanned map. Fulfilling these two primary objectives of a spatial database and outline of processes for utilizing the database can assist the Library of the Marine Corps in providing digital versions of hard maps to its remote users.

To meet these objectives and understand the importance of maps and the capacities of this library the rest of this thesis is structured as follows: this chapter discusses how a user currently accesses a hard map in the Library of the Marine Corps, why these maps are important, why the Library of the Marine Corps is a suitable library for this study, and why hard maps are important to geographic information science (GIScience). Chapter 2 contains a summary of literature relevant to this study in the fields of library capabilities, metadata, definitions, copyright, scanning, georegistering, and vector extraction. Chapter 3 proposes a streamlined process to request and receive a copy of a hard map in digital format from the Library of the Marine Corps, including the development of a spatial database model for archiving map features, generating metadata about each map, and facilitating sharing digital versions of the library's hard maps. Chapter 4 proposes detailed procedures for users to request and receive a digital map, map feature data, and metadata based on a hard map in the library's holdings. Lastly, Chapter 5 provides a reflection on continuing communications with the Library of the Marine Corps and recommendations for future implementations of the proposed process and spatial database and their applicability to other libraries with important hard map collections not readily accessible to remote users.

1.1 Current system: Hard copy map request

Users require access to libraries for research purposes and have various methods for contacting, requesting, and receiving information from a library. Generally a library has a variety of information available in different formats. The methods for contacting a library currently include interacting at the library with a librarian, calling a library and speaking with a librarian, emailing a library, and using a chat function available on a library's website. Users also have options to use a library's website to search a catalog of holdings either remotely or within a library's premises. After a user locates the required document, there is an option for a user to contact the library requesting the document. A user does not necessarily need to request the document from the library if the location of the information is readily accessible, for example, the call number in the library or a download option from the website. If a user is unable to physically access the library, a user can contact a library and request to loan a library holding if available. This option may not be feasible due to the age of a document or other lean restriction for some documents in a given library so that a remote user may receive the document in a different format compared to its original format. A hard map is an example of a document not suitable for loaning between libraries due to its potential uniqueness and fragile nature, regardless of age. This limitation requires the library to send a copy of the hard map to a remote user in a different format. This format could differ between requests from users, as each user's purpose for the map may differ. Therefore, this thesis focuses on hard maps as the document requested by users due to the varying formats required by users for research purposes.

1.2 Maps

Maps are a source of information displaying the relationship between the environment and a map creator's perspective at a set point in time. Brotton proposes that a map is "a graphic representation that presents a spatial understanding of things, concepts, or events in the human world" (2014, 7). This definition succinctly summarizes the intangible value of a map to a researcher. It also specifies that a map is a graphic representation, but does not specify the medium for that representation. Maps can occur in a variety of mediums including stone, animal skin, paper, and digital. For example, one of the oldest maps present today displays the interaction of a community and its environment in c.1500 BCE, carved onto stone, 2.3 by 4.6 m, and situated in Capo Di Ponte, Valcamonica, Italy (Brotton 2014). Maps also display a variety of subjects and perceptions of cardinality, the influence of religion, navigation, and politics (Brotton 2014). Regardless of the medium or subject of a map, for example, the stone map from Italy or a digital map created using ESRI's ArcGIS Story Map, each can provide information for research purposes (Esri 2016). This variety of mediums in which maps are available can affect accessibility by users for research purposes. For example, access to a digital map can be through a Web search or e-mail. Hard maps, unless already digitized, are restricted to the physical storage location of the map.

Today, researchers can have ready access to hard maps due to digitization technology. Unfortunately, many libraries have no standard requirements for handling the digitization of hard maps. Due to financial costs associated with digitizing a hard map, duplication of effort of having the same map both as a hard copy and a digital copy (Goodchild and Zhou 2003), time involvement, and varying digital formats requested by users, it is often unrealistic to digitize every hard map. Therefore, it is important to design a process and spatial database model that facilitates searches for a hard map, provides options for several digital formats of a digitized hard map, a method for digital data delivery to the requester, and a method that can validate hard map digitization efforts. In narrowing the focus of this project, it was necessary to select a single library for use as a case study since there are a variety of organizations that provide access to hard maps including libraries and museums.

1.3 Library of the Marine Corps

A library suitable for this thesis project required two components: a vast collection of historic as well as recent hard maps not currently digitized, and remote users not able to physically access hard maps from the library. The Library of the Marine Corps is a suitable library for this thesis as it has these two components. Currently, there are hard maps not stored in a digital format located at Quantico, Virginia, while the Marine Corps has bases stationed around the globe. Additionally, the Marine Corps requires Marines to continue their education throughout their careers. This means that Marines require access to a library regardless of their location, and potentially may require access to a hard map that is only located in Quantico, Virginia. These components make the Library of the Marine Corps a highly suitable library for designing a solution that facilitates Marines' access to the library's holdings.

U.S. Marines have the opportunity for continuing education in-residence (attend in person) and through distance learning. For example, a Marine may receive orders for one year to Command and Staff College at Quantico, Virginia and attend in-residence or to attend the Command and Staff College Distance Education Program while stationed elsewhere attending to other duties. When enrolled in a college or university, a Marine consults various source documents, such as maps, to learn about past battles. For instance, while in-residence at Command and Staff College, the students attend a staff ride (field trip) to the First Battle of Bull Run, also known as Battle of First Manassas, fought during the Civil War in Prince William County, Virginia. In preparation for this staff ride, students create a slide on a particular aspect of the battle by consulting maps of the movement of the Confederates and Unionists, first-person accounts by soldiers and generals, and interpretations by subject matter experts. The learning materials for this staff ride are available from the Library of the Marine Corps and through public

Internet searches. Marines attending the Command and Staff College Distance Education Program have a similar exercise except there is no staff ride component. Additionally, resources come from the Marine Corps Community Services Library and public Internet searches. Since a distance learning Marine is not located in Quantico, Virginia, the Marine does not have access to the actual battle site or the Library of the Marine Corps. While the physical access to the battle site is cost prohibitive for a distance learning Marine, access to the Library of the Marine Corps could be facilitated. By changing the capacity of the Library of the Marine Corps to extend services to distance learning Marines, the distance learning Marine is now only missing access to the physical battle site in comparison to an in-residence Marine.

The access to items at each library differs greatly. The Library of the Marine Corps has the greatest availability of library items, while the Marine Corps Community Services Libraries offers the lowest level of access. This is because the Library of the Marine Corps is designated as a research library, while the Marine Corps Community Services Library is a community library. As previously stated, the disparity between access to library items and the requirement for a Marine to continue education regardless of location is one of the primary criteria for the selection of the Library of the Marine Corps for this thesis.

The Library of the Marine Corps provides access to holdings to Marines, other Service Members, retirees, and family members. Its mission is to support “the Marine Corps University and Marine Corps by providing comprehensive storage, retrieval, analysis, and distribution of warfighting-related information” (Ramkey 2014, 2). This mission statement is important in the context of this thesis for two reasons. Firstly the mission statement provides support for implementing a system that improves retrieval and distribution of information, and secondly hard maps are a source of warfighting-related information. As previously discussed, maps provide an insight into a user's perspective at a point in time, and a Marine can review information from a map to make informed decisions about a future course of action.

The library has four different sections to help support the mission of the Library of the Marine Corps. Only two sections, Archives and Special Collections and the Virtual Library, pertain to this thesis as these sections contain hard maps. Both sections manage the storage, digitization, and preservation of hard maps. The Archives and Special Collections have three primary categories of information – personal papers, official Marine Corps records, and Marine Corps University materials. Only personal papers and official Marine Corps records contain hard maps. Currently, there is no access to these collections online as the Archives and Special Collections branch building is undergoing maintenance (Marine Corps University 2016).

An example document published by the Archives and Special Collections section does provide some guidance for accessing a collection. This document, about Desert Shield/Desert Storm, explains the library's referencing system, which is important when designing a suitable methodology for improving access to hard maps and how users currently bring a personal laptop and scanner to record information (Marine Corps Archives and Special Collections Branch 2010). As previously discussed, a Marine may be stationed outside Quantico, Virginia and not have access to the Archives and Special Collections branch, so this method of physically visiting the library and collecting materials is not suitable. The Virtual and Research Branches Collection Development Committee selects, digitizes, and distributes materials to Marine Corps Community Services Libraries (Ramkey 2014). However, this branch does not currently have any digitized maps accessible online (Maslowski 2016). The Virtual and Archives and Special Collections are key stakeholders to consider for implementing a process to serve users requesting access and copies of hard maps.

1.4 GIScience and maps

As discussed, maps are an invaluable source of information and GIScience includes a field of research dedicated to researching the extraction of information from hard and digital maps. Other research fields that use information from maps or are concerned with the process of digital information extraction include image processing, document analysis and recognition, machine learning, data integration, and geoinformatics/digital cartography (Chiang, Leyk and Knoblock 2014). Chiang et al. summarize the connection between geographic information system (GIS) and maps succinctly: "Converting geographic features (e.g. place names) in map images into a vector format is the first step for incorporating cartographic information into a geographic information system (GIS)" (2016, 21). Chiang envisages that in the future all hard maps and digitized components could be linked, which involves updating library databases, improving search capabilities, and automating the scanning, georeferencing, and vector extraction software (Chiang 2015). The aims of this project address the designing of a database to link the digitization as well as the data integration process by completing the spatial database design that can support both the cataloging of digital versions and archiving of information retrieved from hard maps at the Library of the Marine Corps.

Beneficial results of incorporating cartographic information into GIS can be the generation of gazetteers and the corresponding improvement of optical character recognition (OCR) software (Chiang and Knoblock 2015). In such a project new words are added to a gazetteer while processing multiple maps from different time periods creates additional validation references (Yu, Luo and Chiang 2016). The Archives and Special Collections branch of the Library of the Marine Corps contains historical maps that could contribute to historical gazetteers, improving OCR capabilities for data retrieval from historical documents from different eras.

Chapter 2 Literature Review

As part of this thesis work, review of two important areas of literature occurred to provide the Library of the Marine Corps with a process that can assist remote users in accessing digital versions of hard maps. First, a review happened on the literature and current practices for several libraries which provide online access to maps to understand how users access spatial data in these libraries, how digital versions of hard maps are stored, and how libraries provide requested spatial data to library users. Second, a literature review revealed the technical processes of scanning hard maps and extracting features from maps into data models. In the context of this thesis, a data model is a vector file, such as a point or shape file containing spatial information and descriptions about an object. By understanding the current procedures libraries use today for handling digital versions of hard maps and the technology and software available for processing hard maps, a conceptual model for the process and spatial database for the Library of the Marine Corps was developed. The following sections provide a breakdown of these related studies, including existing library databases, database design processes, and a database design evaluation framework.

2.1 Related studies

To successfully design a spatial database and process for the Library of the Marine Corps so users can request and download digitized hard maps and related components, a literature review was conducted on a variety of related topics. The following topics as discussed in this chapter, pertaining specifically to libraries, are library capabilities, library user access, metadata, copyright, definitions, scanning, georeferencing, extraction of vector features, and data types.

2.1.1. Library capabilities

It is possible to determine a modern library's capability for utilizing a spatial database and process for providing digitized maps and features by reviewing the software and hardware, staff expertise, and geospatial catalog design and frameworks required to support remote access. These elements encompass confidentiality, agreements and licenses, data management policies, archiving methods, and storage methods. As a part of this project, the knowledge of experts including consortiums (Scaramozzino, et al. 2014, Forward, Leahey and Trimble 2015), committees (Kollen, et al. 2013), and journals such as *Journal of Map & Geography Libraries: Advances in Geospatial Information*, *Collections & Archive*, *Code4Lib*, and *New Review of Academic Librarianship* were reviewed.

The first consortium evaluated currently consists of five university libraries, the University of North Carolina at Chapel Hill (UNC), Texas A&M University, New York University (NYU), North Carolina University, and California Polytechnic State University. Scaramozzino et al. (2014) reviewed this consortium to learn how each library is transforming supplying data from a single access point in a limited range of media formats to providing and receiving data from multiple access points in a variety of formats. This review regarding software, hardware, and staffing for each library provides insight for this project regarding options for the Library of the Marine Corps to implement a system capable of handling geospatial data and providing remote access to it. Table 1 summarizes the software and staffing components from each library involved in that study. The discussion on hardware was limited to locations of where users could access software on computers or what software users could download onto personal computers, so is not included in Table 1 as part of the summary for Scaramozzino et al. (2014) review.

Table 1 Summary of software and staffing from five library universities

University	GIS Software	Search and access Software	Staffing
University of North Carolina at Chapel Hill	Esri, Environment for Visualizing Images (ENVI), GME, Google Earth, Geographic Resources Analysis Support System (GRASS), QGIS, GeoDa, R, statistical software	GIS Data Finder, Andrew File System, Endeca Information Access Platform	Three full-time librarians, two student assistants with skills in GIS and data service, assistance from information technology services
Texas A&M University	ArcGIS	Java and Flash-based Web-mapping services	Two professional, one assistant, a GIS coordinator, four to six student assistants
New York University	Esri ArcMap, ArcPad, ArcIMS, ArcGIS Desktop, GeoLytics, Google Earth, Google Earth Pro, Simply Map, Policy Map, Social Explorer	Libguide, OpenGeoportal.org	Staff from NYU Libraries and information technology services
North Carolina University	GRASS, other open source packages, ArcGIS	GIS Lookup database designed by NC State	One data services librarian
California Polytechnic State University	ArcGIS 10.2, Google Earth, SketchUp, AutoCAD 2014, MATLAB, SPSS, Adobe Creative Suite	SLO Datafinder Web site	Two full-time, one half-time staff, three assistants

Source: Scaramozzino, et al. 2014

Forward et al. (2015) provided information about another consortium, the Ontario Council of University Libraries (OCUL), composed of 21 libraries. The authors' reviewed the consortium's approach to handling metadata standards and utilizing a single online portal for users to access geospatial information. OCUL implemented a self-designed Scholars Portal to allow users access to spatial data from a Web site. This is a successful example of a group of libraries successfully collaborating to creating their own software application, in this case, a spatial data portal so that users can search for and access spatial data from a remote location using the Internet. It is assumed that the collaboration to produce a single application was economically preferable to each library attempting to create its own data portal. An added benefit is the use of common metadata standards across 21 libraries which will streamline continuing usage and maintenance of the data.

The Spatial Committee, formed by the American Library Association (ALA) Map and Geospatial Information Round Table (MAGIRT) Geographic Technologies Committee, interviewed a number of libraries to provide guidance to academic libraries using and implementing spatial data catalogs (Kollen, et al. 2013). Kollen et al. (2013) summarized these interviews to discuss common technologies libraries use for assisting users in searching for an item. The common technologies for searching for an object within a system included databases using PHP and MySQL, content management software, Geonetwork, ArcGIS Server, Dspace, and Solr/Lucene. This variety of technologies used by different libraries reviewed in this study indicates that there is no universal spatial data collection technology currently in use by research libraries. Therefore, the spatial database or process could adapt to any of these technologies depending on the software and staffing resources at the Library of the Marine Corps.

Durante and Hardy (2015) provide a case study using the Stanford libraries and the implementation of a full-service repository for geospatial data. The Stanford libraries spatial data infrastructure utilizes GeoHydra, Extensible Markup Language (XML), EarthWorks, and GeoBlacklight technology. Hydra is a combination of Ruby language, Fedora repository, Blacklight discovery platform, and Apache Solr search index. The GeoHydra, an open source software, and EarthWorks are specific to the Stanford libraries. The Stanford libraries gather information from a variety of sources, not just hard maps. While the objective of this project involves only scanning hard maps for the Library of the Marine Corps, the processes used at the Stanford libraries could be adapted for the Library of the Marine Corps.

Each of the libraries reviewed provide an online search tool using different software to provide users with the ability to search and access geospatial data. UNC experienced difficulty with users requesting access to geospatial data from external locations and changed from GIS Data Finder and Andrew File System to the Endeca Information Access Platform for Libraries (Scaramozzino, et al. 2014). The staffing at libraries seemed dependent on course requirements and funding, and normally employed one GIS expert and one database expert (Scaramozzino, et al. 2014). The reviews showed no universal method for libraries to provide search and access to users regardless of software, hardware, or staffing (Durante and Hardy 2015, Forward, Leahey and Trimble 2015, Scaramozzino, et al. 2014). Since there are a variety of software options for libraries to construct a website for users to search and access geospatial data (Durante and Hardy 2015, Forward, Leahey and Trimble 2015, Scaramozzino, et al. 2014), it is important to understand how existing users access libraries.

2.1.2. Library user access

For the purposes of this project, a user is a stakeholder that requires access to digitized hard maps and associated digital components from the Library of the Marine Corps. This relationship between the library and user is similar to the relationship between other libraries and their users. This relationship involves a user identifying the collection, searching within the collection for the required information (Goodchild and Zhou 2003), requesting the information, and receiving the information. It is important to identify this approach as users have a variety of methods available for searching data, including asking a librarian for assistance or using search capabilities on a library's website, which does not change the relationship between the library and the user. It is important to understand each step for a user, as while the relationship remains constant, the information shared within and between each stage can be different.

To determine how users access and search within libraries it is important to review previous studies and surveys. Kollen et al. (2013) summarized a survey sent to academic libraries by ALA MAGIRT and determined that users access information using search tools that are alphabetical, topical, include basic and advanced text searches, spatial browsing, and have spatial searching capabilities. Users typically access these search tools through Open Geospatial Consortium (OGC) Web Services (OWS) using one of five standard interfaces: a web coverage service, a web map service, a web feature service, a sensory planning service, and a catalog service (Chen, et al. 2010). GIS users typically frame search requests around spatial, thematic, and temporal themes rather than text-based searches (Ahonen-Rainio 2006). This means users typically prefer a visual interface for locating and reviewing findings, rather than reviewing the text of metadata (Ahonen-Rainio 2006). The design of the spatial database for this project, therefore, should have the ability to interact with one of more of the standard search interfaces used by the OGC OWS and provide a variety of search methods to users.

2.1.3. Metadata

Metadata standards provide uniformity for information about objects which translates into attributes associated with an object in a database. The definition for metadata is "a set of data that describes or gives information about other data" (Ahonen-Rainio 2006, 38). Metadata can be external to an object as an accompanying file, or internal to the object, and is readable by both human users and computer applications (Ahonen-Rainio 2006). In the context of this thesis, an attribute is information that explains a certain part of an object, such as the attribute 'author' contains the name of the author that designed the hard map. It is important to maintain metadata for each object separate from a user’s application, so an object is searchable within a catalog with or without a web application (Durante and Hardy 2015), such as an online data portal. As the objectives of this project focus on the U.S. Library of the Marine Corps and relevant spatial components of the library's hard maps and proposed data models, the review considered American metadata standards, standards related to library map cataloging standards, and geospatial information standards.

The Federal Geographic Data Committee’s (FGDC’s) Content Standard for Digital Geospatial Metadata (1998) and the International Standards Organization (ISO) are common metadata standards for libraries (Kollen, et al. 2013, Chen, et al. 2010, Durante and Hardy 2015). Table 2 is a compilation of common standards that could be considered in the construction of the spatial database design. Due to the variety of standards listed in Table 2 and the varying purposes of objects by users and organizations, there is no standard metadata template for American libraries.

Table 2: Common metadata standards

Standard	Purpose
19110	Geographic information – methodology for feature cataloging
19115-2	Geographic information – metadata – gridded imagery – part 2
19115	Geographic information - metadata
19119	Geographic information – services
19139	Geographic information – metadata- xml schema implementation
19157	Geographic information – data quality

Source: Durante and Hardy 2015

Current map cataloging standards for libraries utilize the *Resource and Description Access* and *Functional Requirements for Bibliographic Records* (Traill 2014). These two documents assist in creating frameworks, such as the Library of Congress Bibliographic Framework Initiative, which uses an XML for users to search for an object within a system

(Traill 2014). The generation and maintenance of metadata are both manual and automated. Owners that generate the data or receivers that store the data can manually update metadata information associated with an object, or computer programs can automatically retrieve the information necessary for the metadata template of an object (Ahonen-Rainio 2006).

Additionally, text-based cataloging systems used by many libraries are different from common metadata standards (Ahonen-Rainio 2006), meaning there is no simple integration between the two standards, resulting in additional differences between organizations. While metadata creates the frameworks for entering data, there is minimal guidance in the standards on standard terminology for the entry fields of objects. Since there is no accepted standard, there will be minimal metadata fields used in the designing of the spatial database. Future consultation with the Library of the Marine Corps can of course add additional fields as required.

2.1.4. Geospatial catalog and frameworks

The structure of a geospatial catalog is the key to how a library can share information with users. A geospatial catalog facilitates "sharing, discovery, retrieval, management of, and access to large volumes of distributed geospatial resources" (Chen, et al. 2010, 411). In 2010, the Spatial Data Committee established by the ALA and MAGIRT Geographic Technologies Committee provided guidance to academic libraries on spatial data catalogs (Kollen, et al. 2013). The OGC uses two primary models for cataloging digital resources, e-business XML Registration Information Model (ebRIM) and Universal Description, Discovery, and Integration (UDDI) (Chen, et al. 2010). The ebRIM uses an object-orientated approach which allows for flexibility when integrating OWS, while the UDDI is a methodology intended for an organization wanting to use OWS (Chen, et al. 2010). The design of the spatial database and process for this project does not intend on using OWS for the Library of the Marine Corps to provide access to digitized hard maps. Nevertheless, the ebRIM with its object-orientated design is important in describing how an item is stored within the system and can be accessible to users. A framework, such as the metadata lifecycle, supports the successful implementation and maintenance of a geospatial catalog.

The metadata lifecycle explained by Kollen et al. (2013) describes a library's management of an item: collection, creation, storage, publication, discovery, retrieval, and access and update. For this thesis, the hard map already exists within the library, so a review of the collection step is unnecessary, whereas the rest of the steps in the metadata lifecycle require review. This lifecycle allows users to locate an item within a library and is adaptable to different libraries in terms of order and methodologies.

Durante and Hardy (2015) provide a framework focused more on geospatial data compared to the metadata lifecycle with the aim of designing a "durable digital library assets, delivering vector, raster, and georeferenced maps via a spatial data infrastructure, discovering geospatial layers with a variety of applications and contexts, and curating geospatial data" (124). The focus of this project is on access to hard maps, digitized hard maps, and associated vector formats. According to Durante and Hardy (2015), observation is important to design a structure capable of accommodating a variety of geospatial data formats.

In addition to a solid framework, it is important to include collection metadata during the discovery stage (Goodchild and Zhou 2003). This component adds an upper level for users to query before searching in detail for an item. Collection metadata is available to libraries in two formats: inherent, meaning auto analysis, and contextual, meaning supplied by the creator. Regardless of the selected collection metadata format, each item within a collection still requires metadata so users are not sifting through large collections (Durante and Hardy 2015).

2.1.5. Definitions

To meet the objective of this project to design a robust spatial database that can locate and request an object and remove duplicates, consistency is a key factor (Ahonen-Rainio 2006). In 2008 the Library of Congress announced its Policy and Standards Divisions would develop a genre/form thesaurus for cartography (Traill 2014). In 2010 the division published the *Library of Congress Genre/Form Terms for Library and Archival Materials* that implemented multiple headings for an object, rather than a single heading including multiple terms (Traill 2014). There was no reference to hard maps in this publication. However, this change to headings for an object is important to this study, as removing multiple terms from an attribute improves the integrity of the database.

Hard maps also can be classified as rare materials and occur either as a map sheet or a book. This means it is necessary to have definitions capable of incorporating both formats. In 2007 the *Descriptive Cataloging of Rare Materials (Books)* and *Descriptive Cataloging of Rare Materials (Cartographic)* provided standards for these formats (Traill 2014). This project's spatial database design must take into consideration these dual source formats for hard maps.

2.1.6. Copyright

Libraries and source items within a library are subject to copyright terms and conditions. These terms and conditions may restrict which users can access an item, whether an item can leave the library premises, what the item can be used for or in, or how the item is to be referenced. Additionally, the Library of the Marine Corps has different classifications that mean only users with a certain level of clearance can access items. A spatial database for a library should consider these terms and conditions during the design phase in consultation with the library.

2.1.7. Data types

Maps are a compilation of data showing the designer's perspective of the environment at a single point in time. Data represents the environment either in a discrete or continuous format. There is no correct format for data representing the environment, and the data format depends on the user's intended use or subject. Discrete data is either a point, line, or area, while continuous data is pixel based, forming a raster. The data format and software program used to create and store the data determines the file format of the data. For example discrete data can use a shapefile format (.shp) while continuous data can use an image file format, such as a portable network graphic (PNG) transparent file (.png), joint photographic experts group (JPEG) raster format (.jpg, .jpeg), tagged image file format (TIFF) (.tiff, .tif), and graphics interchange format (GIF) transparent file (.gif).

The creation of a hard map is from discrete and continuous data. Often the background of a hard map is continuous representing a surface, while areas of interest are discrete. For example, a street map can user either a satellite image or elevation surface as background, while the roads, houses, and county boundary areas are placed on top of the surface as a combination of points, lines, and areas. Since the hard map is an image, it is continuous data and saved in one of the file formats listed above for rasters. If required, a user can also digitize items from a hard map and create discrete data.

2.1.8. Scanning of hard maps

The scanning of a hard map is important to this project for two reasons. The first reason is that a user can receive a scanned map compared to a hard map. It is not always possible or realistic to send a hard map to a user, as it may by priceless, have large dimensions, or not be a suitable medium to mail. The second reason is that there is software available that can extract features from scanned maps, such as points, lines, polygons, text and numbers, and symbols (Chiang, Leyk and Knoblock 2014). By extracting digital features from hard maps, it is possible to perform analyses of the features digitized from the hard map and combine these with additional features from other hard maps, in order to perform new analyses (Chiang 2015). This project's literature review does not provide a detailed process on how to scan maps. Nevertheless, the general process is important to understand, so that the process must be described within the spatial database for scanning and capturing features and other information form the map.

Stanford University Library's Digital Production Group digitizes maps (Stanford University Libraries 2016). This group uses a variety of equipment to scan a map, including a robotic book-scanner, manual book-scanner, PhaseOne P65 and reproduction system, WideTek wide format color scanner, flatbed scanner, and sheet feed scanner. Each piece of equipment has different specifications regarding the level of resolution, with the robotic scanner having the lowest resolution of 200 to 600 ppi, and the flatbed scanner having the highest maximum resolution of 6400 ppi. One of the members of the group selects the suitable piece of scanning equipment during the digitization process.

There are a number of steps required to digitize a map. The Stanford Library created a video showing a typical method utilizing a camera for imaging a map. The author interpreted the steps as follow from the video (Stanford University Libraries' Digitization Labs 2012):

1. Obtain resolution for image capture (based on the finest detail)
2. Adjust height of camera to capture necessary pixels
3. Take a picture of target used for quality insurance and software
4. For a map requiring multiple images, make sure the tiles overlap

5. Imaging software crops each tile
6. Upload tiles onto server for quality control
7. Stitch and photoshop tiles together with enhancement or embellishments

If the digitizer chose another piece of equipment to scan the hard map, then step 2 would be adjusting the piece of equipment to ensure the correct resolution is captured for the image. Other items that might change the process or require noting in the metadata for the scanned image include the compression process and any aging artifacts if it is a historical map (Chiang, Leyk and Knoblock 2014).

It is unknown what equipment the Library of the Marine Corps has for scanning maps. Currently, the Library of the Marine Corps invites users to bring personal computers and scanners to the Archives and Special Collections branch to scan documents (Marine Corps Archives and Special Collections Branch 2010). This method was unsuitable for this project since the author does not reside near the library location. Additional consultation is required with the Library of the Marine Corps to determine current practices used by the library staff and possibility of expansion of hardware to scan maps within the library.

2.1.9. Georeferencing

After a librarian or technician digitizes a hard map, the newly created image file needs to be georeferenced so, the digital map works in a GIS. This means adding the appropriate projection and coordinates to the raster. The degree of difficulty for georeferencing an image depends on the clarity of features within the image, the scale of the image, the coverage of the image (Campbell and Wayne 2011), and the published date of the original hard map (Yu, Luo and Chiang 2016). If features are distinguishable within the image, the image occurs within one projection, the area is large enough in the image to distinguish the area from other areas and was recently published, then a GIS technician using a standard GIS software package could georeference the image with a lesser degree of difficulty. For an image with any differences related to these four criteria, the degree of difficulty increases.

This project's objectives focus on the assumption that hard maps from the Archives and Special Collections branch of the Library of the Marine Corps can be digitized. The hard maps are likely to be hand drawn and from a different era. This means any digitized hard maps are likely to be older maps; the features may be difficult to identify on the map, and the physical scale of the image could change depending on the experience of the cartographer. Therefore, it is likely to be a challenge to georeference any of the digitized images of older maps. To simplify the goals proposed in this project, it is proposed that each map image be georeferenced with a point for the general area of the map, rather than georeferencing the image itself. The spatial database can also contain this simplified map location, datum, and projection as attributes (Chiang, Leyk and Knoblock 2014). However, for the extraction of features from the digitized image, the image must also be georeferenced. Therefore, additional research is required in the processing of the digitized images to ensure each image is georeferenced with an acceptable degree of error.

2.1.10. Extraction of vector features

The extraction of vector features from a georeferenced image is proposed as the final stage in the process for this project. There are four options for processing a scanned map for collecting vector features: (i) raster-to-vector conversion tool with minimum automation, (ii) semi-automatic systems, (iii) fully automatic systems for specific map types, and (iv) fully or semi-automatic systems for specific types of map features (Chiang, Leyk and Nazari, et al. 2016). Regardless of the processing system selected there are some challenges to extracting vector features, in particular, labels, from historical maps. These challenges include artifacts (noise, intersection of other text labels, and non-textual artifacts), unpredictable rotation and orientation of text, and resolution of the image (Yu, Luo and Chiang 2016). Chiang, Leyk, and Knoblock (2014) reviewed text recognition programs and associated challenges in detail. Some of the information from this review is incorporated into the design of the spatial database for this project, such as the resolution of the image, sampling spot size (blur), the bit depth of resulting data, and brightness/contrast (Chiang, Leyk and Knoblock 2014).

2.2 Existing library databases

A review of other libraries cataloging standards and literature reviews of entity-relationship models assisted in designing a spatial database capable of allowing users to locate, request, and receive digitized versions of hard maps from the Library of the Marine Corps. The author sent emails to libraries and Web sites that host digital maps to understand cataloging standards and databases. Chen et al. (2010), Goodchild and Zhou (2003), and Durante and Hardy (2015) provide models for cataloging standards. The Web sites of each organization, the responses to the email, and the example models provide insight into how to design a spatial database for this project.

2.2.1. Library cataloging examples

The organizations contacted to inform this project were either a research library or had a Web site that hosted digital maps. Table 3 provides a list of the observed and contacted organizations. From the six contacted organizations, the Library of Congress, Stanford Digital Repository, History of War, and the Naval History and Heritage Command provided responses. The responses are included in Table 3 along with observations from each organization's Web site.

Table 3 Observations of organizations with online map collections

Name of System; Organization	Function	Model	Data Collection practice	Data dissemination functionality
Perry-Castaneda Library Map Collection; University of Texas Libraries	Current and historical maps online. Contains public maps from the LoMC Archives and History Division and other military services. There are 70,000 maps online and 250,000 in print.	Unknown	From non-copyright sources. Individual users cannot request digitization of a map. A user can email a request of a suggestion for a map requiring digitization. Scanning uses two different scanners with 400 to 800 dpi Older maps are 150 to 200 dpi.	Maps are downloaded via a hyperlink. Can search via broad categories. There is no preview panel before download. Within the web browser, there is an option to zoom. The only format for saving is JPEG.
Military History Encyclopedia on the Web; History of War	Maps of operations in which Marines have participated	Access database for the articles, pictures, and maps	Unknown	Able to search data via two methods – indexes (categories) and a search box (categorized via subject terms). If accessing via the gallery, then there are thumbnails available with hyperlinked names.
Naval History and Heritage Command	Maps of Marines participation in operations	OCLC database or Library of Congress system of cataloging and classification	Unknown	There is no direct ability to search for online map media. Maps are associated using topics.

Name of System; Organization	Function	Model	Data Collection practice	Data dissemination functionality
Library of Congress	17,400 map collection online with some related to Marine Corps operations.	Library of Congress Integrated Library System (LC ILS) with an upgrade to Voyager 2000.1.3. Database) resides on a Sun E10000 server	A variety of sources including maps, atlases, reference works, globes, relief models, CDs, and DVDs.	Accessed via online map collections website.
David Rumsey Map Collection; Cartography Associates	71,000 maps online from 16th through 21st century with free access to high-resolution images	LUNA Browser	A variety of optical equipment and digital scanners. Images are compressed and viewable via the internet.	It is possible to view maps via a variety of applications including LUNA Browser, Atlases, Georeferencer, MapRank Search, Google Earth, Google Maps, Second Life, 3D GIS, Insight Java Client, and Ticker. LUNA browser users can search by keyword, advanced search, or category.
Stanford Digital Repository	Provide digital collections searchable via Search Works	Search Works catalog	Digitized by staff or gathered collections from third parties.	Search Works via the internet hosted by Stanford University Libraries. Unable to view online maps, unless they are available for public access.

As the Archives and Special Collections branch was not available for comment in the time frame of this study, it is unknown what model and data collection practices the branch currently uses. The Naval History and Heritage Command responded that it uses the Library of Congress cataloging and classification system, which means the Library of the Marine Corps, may conform to the same standards. The responses from each organization enforce the earlier reviews from Section 2.1, that there is currently no standard methodology for designing a spatial database, and that it depends on existing infrastructure, each organizations' requirements, and needs of the users.

2.2.2. Entity-relationship model examples

The literature review concerning different entity-relationship models applicable to spatial databases that can serve academic or research libraries includes objects and attributes that may be suitable for designing a spatial database. While each component is important in an entity-relationship model, only some components apply to this project. Figure 1 is a simplified entity-relationship model displaying objects and relationships. Objects relevant to this project's spatial database include 'registry object', 'classification,' 'classification scheme', and 'registry entry'. These four components allow for identification of a hard map and for a user to search for the hard map. The 'extrinsic object' is the search component of this model. This allows a user to request an item. Another important item is the use of a unique identifier. This unique identifier allows objects to be incorporated into the collection-level metadata (Goodchild and Zhou 2003), allowing each object to belong to multiple classifications (Durante and Hardy 2015). By allowing an object to occur within multiple classifications, this removes duplication and allows a user greater search flexibility.

Another set of authors designed an overview for extracting data and metadata from an object (Durante and Hardy 2015) . This methodology included processing data so users could access information via an online data portal, which is different than the proposed methodology for this project. The outline presented in Durante and Hardy's (2015) paper does present an alternate idea for the step of registering using two parallel processes as displayed in Figure 2.

Chen et al. (2010) provided a detailed perspective of the attributes associated with registering an object integrating three metadata standards ISO19115, FGDC, and an additional standard NASA EOS Core System (ECS). The NASA ECS is not under consideration for this project as it was designed for NASAs remote sensing data. Figure 3 displays the interactions between three different objects utilizing the different metadata standards; the core dataset, responsible organization, and reference system.

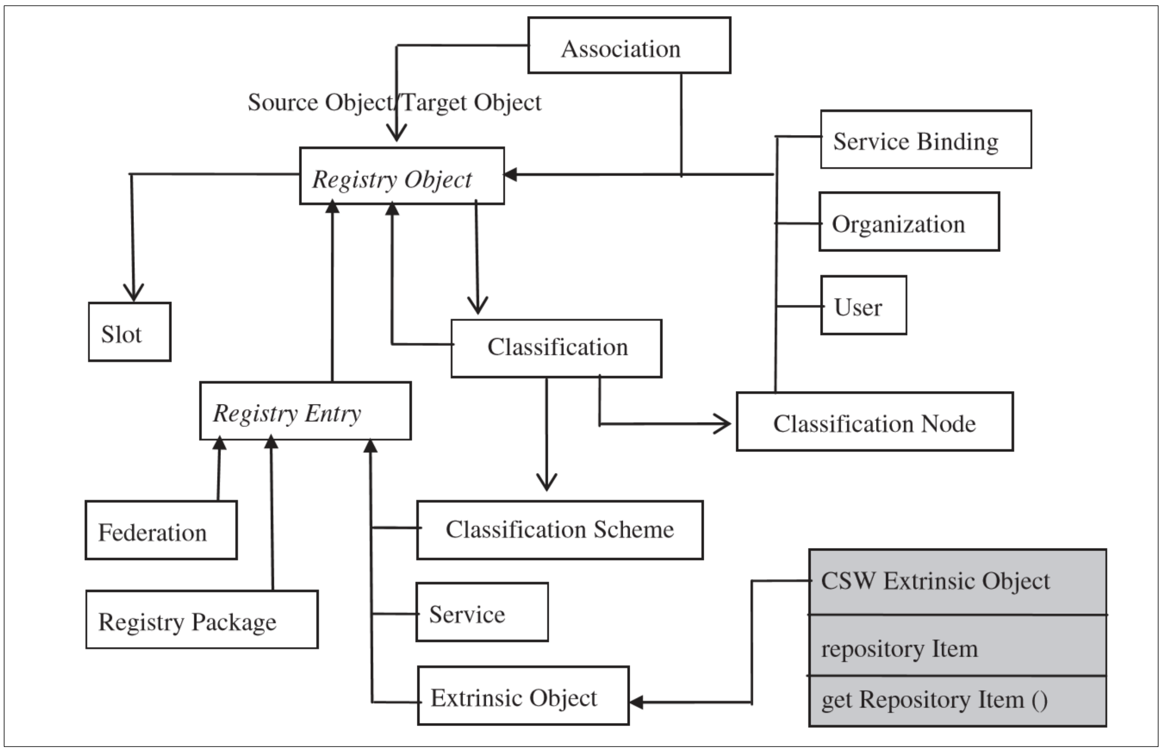

Figure 1 Overview of ebRIM model with catalog standard extensions
Source: Chen et al. 2010, 414

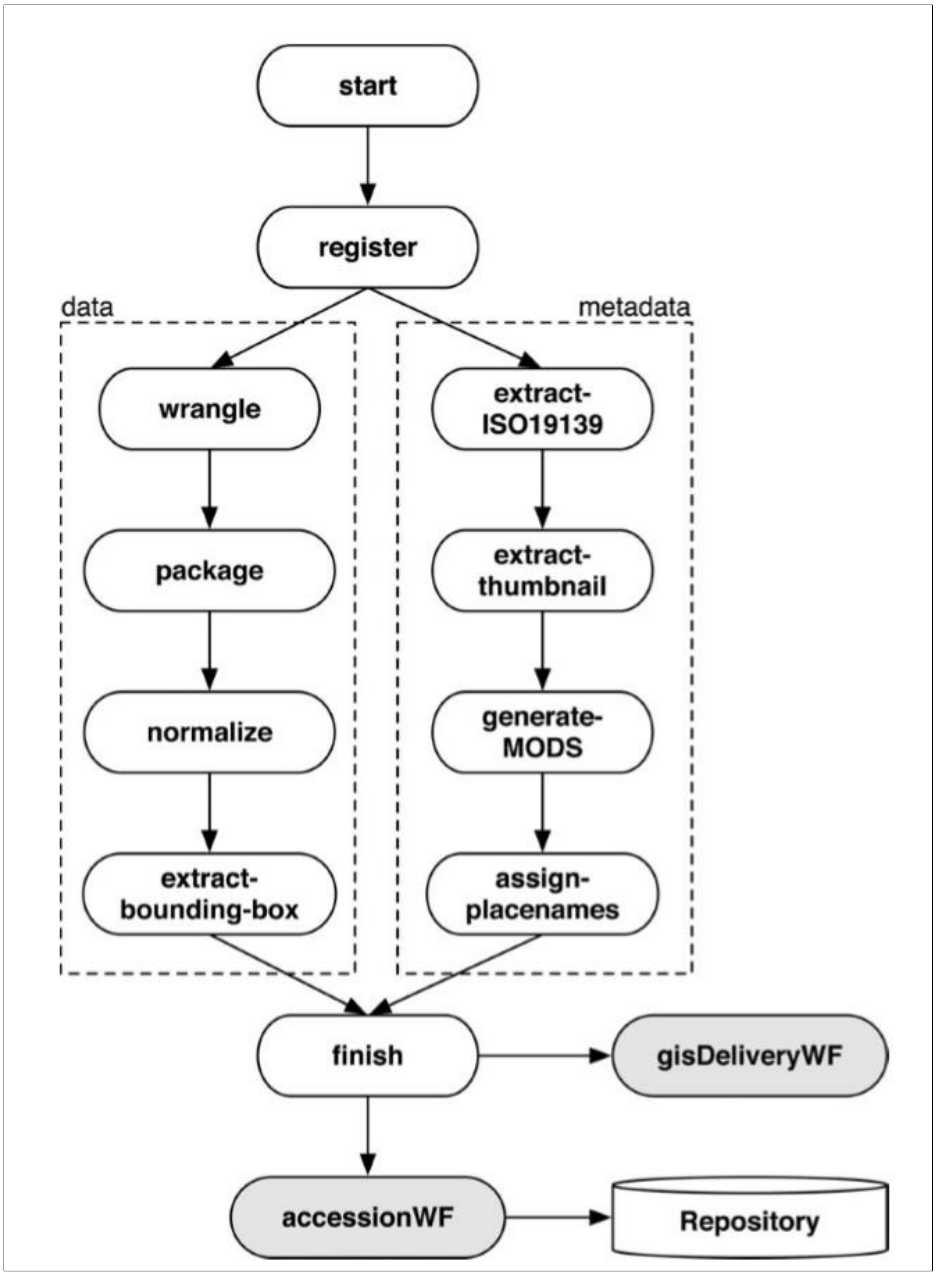

Figure 2 Outline of parallel processes registering an object
Source: Durante and Hardy 2015, 133

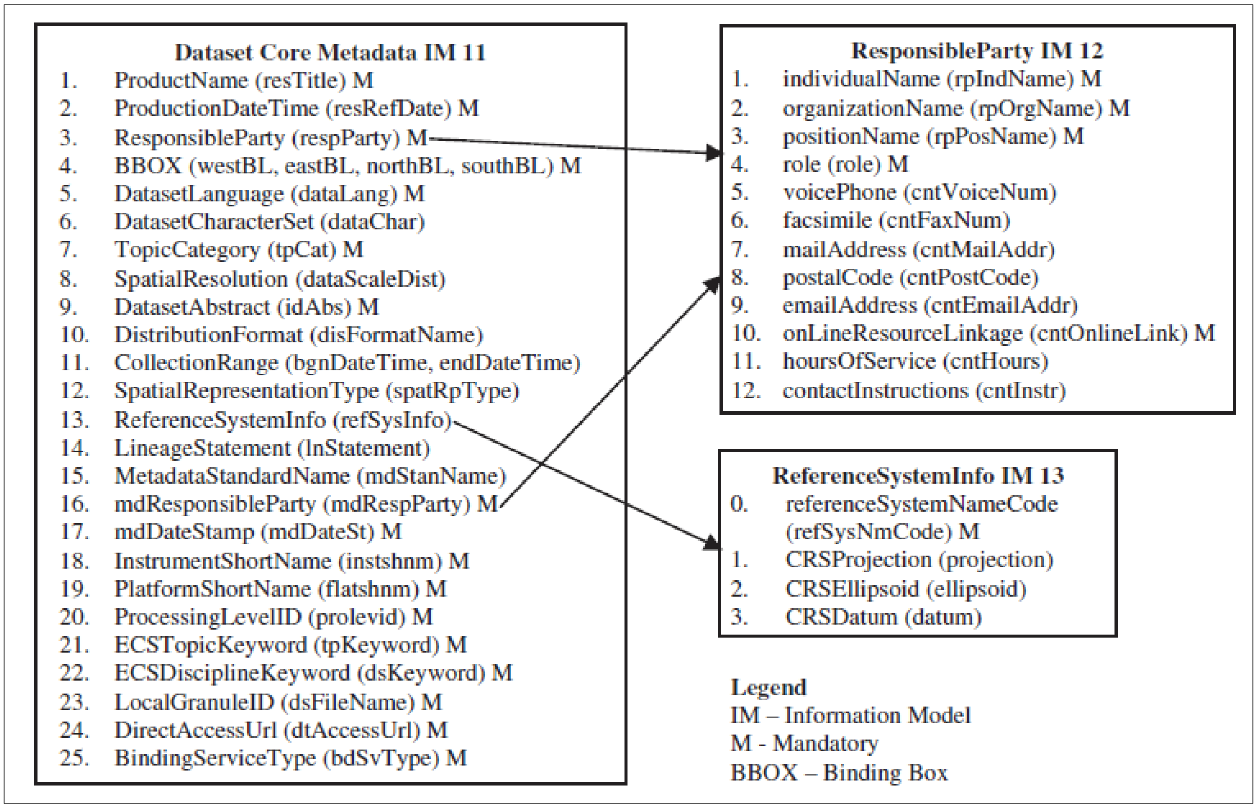

Figure 3 Example of a dataset incorporating ISO19115, NASA ECS, and FGDC metadata
Source: Chen, et al. 2010, 414

2.3 Database design process and evaluation framework

To ensure the relationships between the different objects in the entity-relationship diagram, there are three constraints to consider: (i) domain constraint, (ii) entity constraint, and (iii) referential constraint (Yeung and Hall 2007). A domain constraint establishes permissible value limits for each attribute. The domain is the smallest representative unit within each attribute and cannot contain composite or multivalued attributes. The entity constraint is the use of a primary key (PK) within each object and ensuring there are no null values. The referential constraint involves the entry of new items within an object that has a foreign key (FK). The item must first have a PK before a corresponding entry is created in a table where it has a FK, otherwise the integrity of the database is compromised. The three constraints were considered during the design phase of the spatial database for this project, and the resulting spatial database was subsequently evaluated against the three constraints, described in detail in Chapter 4.

Evaluation of the process objective utilized the nine-stage GIS planning methodology listed in Table 4 (Tomlinson 2013). By reviewing how this project's process meets these nine-stages it is possible to determine how the Library of the Marine Corps could receive the process and spatial database and implement these items into the existing service system.

Table 4 Nine-stage GIS planning methodology

Stage	Definition
1: Consider the strategic purpose	Consider an organizations goals, objectives, mandates, and management accountability for measuring success of this project.
2: Build the foundation	Define planning and implementation stages to the organization and receive funding for the planning stage. Organize a GIS team to plan the design of the project.
3: Conduct a technology seminar	Identify what the organization needs from GIS by hosting technology seminars about the benefits of GIS and the planning process.
4: Describe the information products	Identify tasks for stakeholders and design suitable information products involving the method of construction of the product, frequency of product outputs, input data, tolerance of error, and benefits of a new product.
5: Consider the data design	Define spatial data characteristics, data sources, data and technology standards, and conversion and interoperability guidelines.
6: Choose a logical database model	Select a database model that best represents the input data, relationships, attributes, and behaviors.
7: Determine the system requirements	Identify GIS workstations, software, communications bandwidth, core capacity, and location of storage. Present findings to organization.
8: Consider benefit-cost, migration, and risk analysis	Determine method for implementation and prepare a presentation for organization using a benefit-cost analysis and risk factors.
9: Plan the implementation	Write report addressing staffing and training, legacy hardware and software, implementation schedule, managing changes, and benefit-cost analysis.

Source: Tomlinson 2013

Chapter 3 Data and Methods

A spatial database and process were designed as part of this thesis, aimed at improving a library's searching, requesting, and sending functions for digital versions of a hard map. By designing a process for accessing a record of a hard map from an electronic library catalog in a remote location, and for requesting scanned hard map and vector features from the selected hard map, it is possible to design a spatial database to support this process. Figure 4 displays the methodology for designing the spatial database and process for this project. Chapter 2 addressed the first section 'Research Design'. The rest of this chapter focuses on proposed 'Data requirements and sources' and 'Procedures and analysis' by discussing the different user communities, the design principles, the spatial database design, and proposes collection of example data.

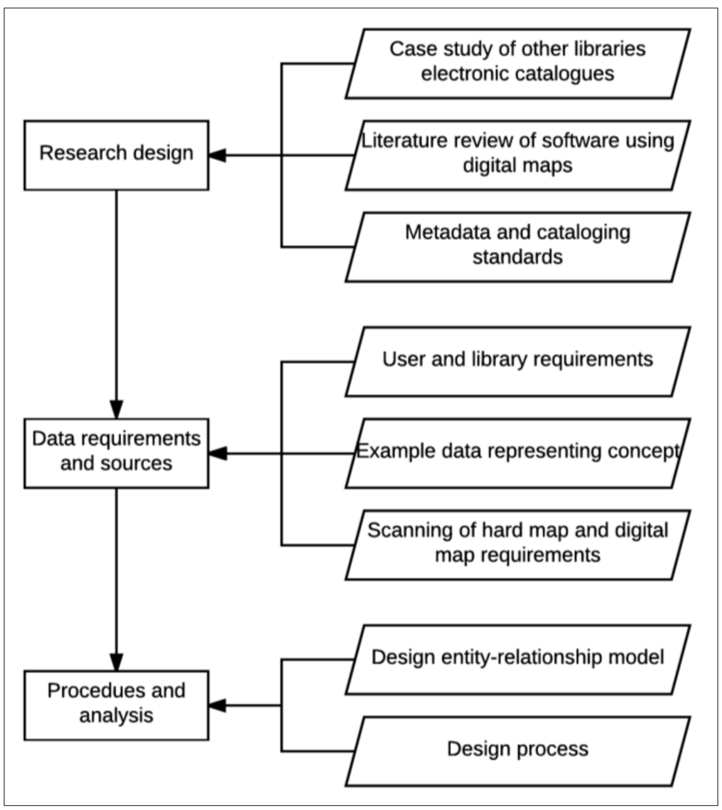

Figure 4 Overview of proposed methodology

3.1 User communities

It was important to consider the key stakeholders' involvement in the process of accessing and providing access to digital versions of hard maps. By understanding the requirements of each stakeholder a process and spatial database for accessing both digital versions and catalog record of hard maps was designed. There were four primary stakeholders identified for the interaction with the hard maps and the Library of the Marine Corps. Unfortunately, as already discussed in Chapter 2, the Archives and Special Collections branch was not available for consultation. Therefore, the four stakeholders represent those parties that are assumed by the author to be present within the Archives and Special Collections or another branch of the Library of the Marine Corps. Figure 5 provides an example of the four stakeholders, how the stakeholders relate to the process, and their associated tasks. The next sections in the chapter discuss each stakeholder in detail.

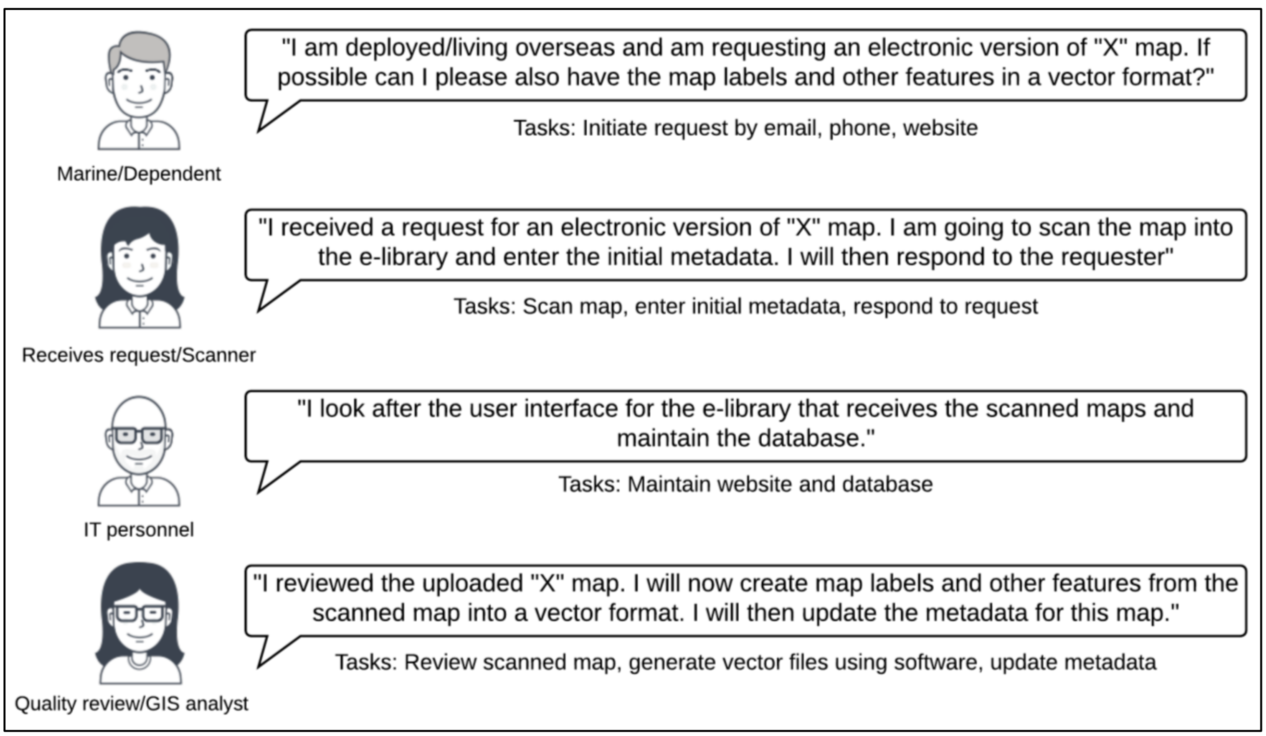

Figure 5 Key stakeholders associated with process

3.1.1. Marine/Dependent

The Marine or Marine's dependent are the primary stakeholders, as these users are the initiators and start the process of digitizing a hard map. Without a Marine or dependent requesting a copy of a hard map or map features in a vector format, there is no initiator for selecting a hard map for processing. As discussed in Chapter 2, it is probably not feasible to process every hard map in a collection into a digital version due to money and time constraints of a library and its staff. Therefore, the Marine automatically provides prioritization for the library for processing hard maps in a collection into a raster and vector format.

It is necessary to capture information about each user, including best method of contact, details about the requester, and level of clearance. As Marines are stationed or deployed in various countries it is important to capture the best method of contact in order to enable communication between the Marine and the library and to ensure delivery of the product(s). By gathering other details about the requester, such as the Marines unit or level of education, this information can be used for future analysis by the Library of the Marine Corps.

Also as discussed in Chapter 2, it is necessary to consider copyrights and classifications of hard maps. By collecting the level of clearance of the requester, this information acts as a mechanism for ensuring requesters can only access appropriate materials.

3.1.2. Librarian (receives request and scanner)

The librarian that receives the request from the initiator is the next key stakeholder in the process. The librarian is a staff member, and for this project performs two tasks – responding to the request and scanning and entering the initial metadata about the selected hard map.

Response to the request by the librarian validates the request of a user for a digital copy of a hard map or map features in vector format. The librarian identifies the version of the hard map in the collection and responds to the user. After the hard map is transformed into a raster or vector format, the librarian responds to the user by providing or transmitting the final product(s) to the user.

There are various hardware and software available for scanning hard maps as previously discussed in Chapter 2. The skill level of the librarian for using the necessary hardware and software depends on the resources available at the library. It is unknown what resources the Library of the Marine Corps currently possesses for scanning a hard map, therefore, it is not possible to identify the skill level. The librarian does require data entry skills to enter initial metadata about the hard map into the spatial database.

3.1.3. IT personnel

The database and the user interface for the process of entering and managing the hard map require information technology (IT) personnel. In the initial phase of creating a process for converting hard maps to scanned and vector formats, the IT personal construct the spatial database and the user interface. After the spatial database and user interface are operational, the IT personnel maintains the software and hardware necessary for converting hard maps into digital versions and the user interface for access.

3.1.4. GIS analyst

After the librarian scans in a copy of a hard map and creates a unique entry in the spatial database, the GIS analyst uses appropriate software to contribute additional information from the map. This additional information is in the format of vector features and can include labels, roads, buildings, or other manmade features. As the GIS analyst creates each vector feature a connection remains in the spatial database to the original hard map. This reduces duplication of effort if a different user requests the same hard map and provides the ability in the future for users to search for vector features via the user interface since the spatial database is now populated with information from hard maps. The GIS analysist also accesses the user interface to update the metadata with the necessary information for each digital version of the hard map.

3.2 Design principles

The spatial database for supporting the information and products gathered from scanning a hard map involved four main processes: (i) initiation of the task, (ii) scanning of the hard map, (iii) metadata entry, and (iv) creation of vector features. Taking into consideration the previous section about key stakeholders this section continues the discussion about the principles behind the design of the four processes.

3.2.1. Task initiation

Task initiation is a pivotal process that enables a central point for the process of scanning a hard map. The task initiation incorporates factors such as details about the user requesting information from the hard map, which librarian processes and scans the map, and details from the existing library catalog about the hard map. Task initiation also allows the system to record a start date, which allows management and administrators to track information regarding the average time to process a request or the amount of time used by a staff member on a single request.

3.2.2. Scanning of hard map

The scanning of a hard map depends on the technology and software available to an organization. An organization such as the Stanford Library has multiple technologies for scanning hard maps of various sizes and mediums. This means the collected data regarding the scanning and processing of a hard map could be different, but there are some common principles as listed in Table 5.

Table 5: Common principles from scanning a hard map

Principle	Purpose
Resolution	Resolution is determined by the hardware or the personnel and provides information about what the scanned map could be used for in relation to visual displays and creation of vector features.
Hardware	Recording the actual hardware used to scan the map provides information to other users about the product. This could provide insight into accuracy of the scanned map.
Error	This provides information to the user of a scanned map any errors that may have occurred during the scanning of the map.
Personnel	Knowing who scanned the map provides assistant in quality reviews of items. This can also provide assistance in the level of accuracy for a scanned map.

3.2.3. Metadata entry

There are two types of products that require data entry during the processing of a hard map, firstly the hard map itself and secondly any raster or vector data formats. Depending on the location of a library or the preferences of a library there is no worldwide industry standard for the metadata for a hard map or a raster or vector data format. This means library data catalogs could differ and flexibility is required in designing a spatial database to capture all necessary information about the hard map and its additional products.

3.2.4. Vector feature creation

A GIS analyst can create a vector feature using suitable software. Discussed in Chapter 2, there are different types of software for extracting vector features from a scanned hard map. It is important to capture generic information about the feature extracted from the scanned image and capture this information in the metadata. The data entry must also remain simple to ensure consistency between different features and users. Similar to capturing information about the scanned map is it important to capture the identification of the staff that generated the vector data, the software used to create the data, the file format of the features, any errors encountered during the processing, and other comments about the extracted information from the scanned image. Capturing this information in the spatial database allows other users to locate and use the vector features of a hard map.

3.3 Spatial database design

A spatial database supports the structure of a system that contains all the necessary information for each object and the relationships between each object (Yeung and Hall 2007). This type of database is recommended for this project because it is able to contain both spatial information about an object and non-spatial descriptors. During the design of this spatial database, it was important to document the stakeholders assumed to be involved in designing the database structure, the guiding principles of the process, and the proposed functionality of the user interface. This chapter already discussed the stakeholders and principles behind the process. This section introduces a user interface for searching for a hard map and how the spatial database interacts within that system. Then the basic principles behind constructing a spatial database are proposed.

3.3.1. Overview of process

The process of a user requesting a scanned image of a hard map with features from the map in a vector format is assumed to require a number of steps and interactions of different components. Figure 6 provides a proposed overview of the different components involved in the process. There are two main differences in the components as indicated by the two dashed containers.

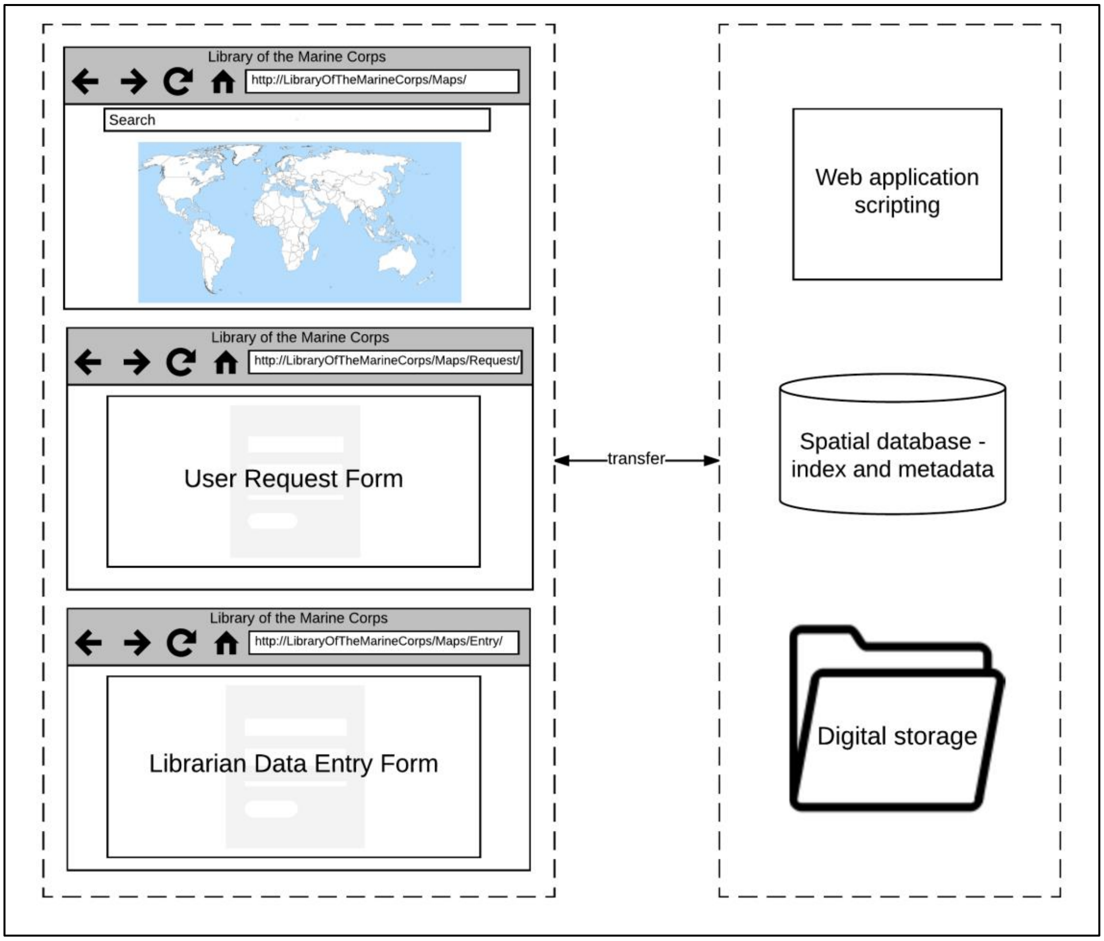

Figure 6 Overview of components and interactions

The container on the left is the user interface and is composed of three different parts. The first provides the ability for a user to search for a hard map using keywords, or spatial or temporal parameters. Once the user selects a hard map, the program initiates a request form and sends it to the Library of the Marine Corps. Once the librarian receives the request and the scans the map, the librarian enters the metadata for the scanned image using the librarian data entry form. The GIS analyst also uses this form to enter in new vector data files or update the scanned metadata.

The container on the right houses all the software and hardware that supports the user interface. The web application scripting creates and maintains the website. The spatial database contains the index and metadata information for the hard maps and is the primary focus of this thesis. The digital storage is where the scanned maps and the vector features obtained from the scanned image are stored. The design has the raster and vector files stored separately from the spatial database as the file size of a scanned map can range from 200 MB to 2 GB (Rumsey, 2016). The greater the storage requirements of the database, the greater the requirement for the amount of memory to index and search for items within the system. Therefore, to allow for growth in the system and not grow the database such that it becomes inefficient, this project's design proposes the raster and vector files are stored separately.

3.3.2. Spatial database principles

The proposed design of the spatial database used an entity-relationship data model. The entity-relationship model provided a workspace for visualizing the types of objects and relationships between objects associated with a process. The workspace can be accessed and understood by various stakeholders including the client, the database designer, the database developer, and the end users (Yeung and Hall 2007). In this case, the proposed client is the Library of the Marine Corps, the database designer is the author, the database developer is the IT personnel, and the end user is the Marine or dependent. As the client is further engaged, the entity-relationship model can later be adapted to meet the needs of the Library of the Marine Corps and the users.

The actual entity-relationship model includes tables representing each object. An object in the spatial database can represent a collection of items, such as people or books. Each object has a list of attributes. Depending on the objects within the entity-relationship model there could be geospatial feature classes, non-spatial tables, and spatial tables. The relationship between each object is represented by line symbols with different markings as shown in Figure 7. It is possible to have one-to-one, one-to-many, and many-to-many relationships with the additional rule of mandatory or optional between tables within the spatial database.

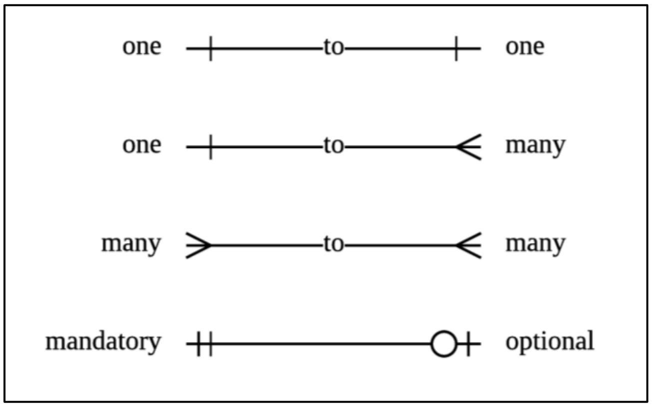

Figure 7 Types of relationships possible within an entity-relationship model

3.4 Collection of data

This author consulted with the Library of the Marine Corps and reviewed necessary documents from the Library of the Marine Corps website to understand the requirements of the library and its users. The author requested example data from the Archives and Special Collections branch of the Library of the Marine Corps. Due to the physical moving of the branch at the time this research was conducted, the Archives and Special Collections was not available for consultation.

As previously mentioned there are other libraries that service the Marine Corps which the author consulted, including the Marine Corps Community Services Libraries at Camp Foster, Okinawa, Japan to determine how remote users normally access hard maps from the Library of the Marine Corps. The Camp Foster library currently provides remote users with the contact details for the Library of the Marine Corps. The Camp Foster library only had three hard maps within the collection, as the library has a community service focus rather than a research focus.

Therefore, the three maps, two of which were present in another Marine Corps library, were not suitable as example data for this thesis.

Chapter 4 Results

To meet the objectives of this project, this chapter presents a spatial database design, evaluates the spatial database design, and evaluates the process for making hard maps available to users, Marines and dependents, proposed in Chapter 3. The first section of this chapter reveals the proposed spatial database structure and provides a discussion about each object for interest. The second section evaluates the spatial database design. The final section evaluates the process as presented in Chapter 3.

4.1 Spatial database

The spatial database presented in Figure 8 is a combination of objects and relationships between objects. Each object is comprised of attributes that contain additional information about an object. This section discusses, in particular, the construction of the entity-relationship model and the selected attributes for each object.

4.1.1. Spatial database structure

The objectives of this project informed the resulting structure for the spatial database. Firstly, the spatial database needed to contain information that would allow a user to search for a map using either a term, a category such as ‘civil war,' spatial information such as place name or spatial coordinates, or temporal information. Secondly, the spatial database must facilitate access to stored digital formats of a hard map, including the ability of a user to download the map image. These two objectives appear straightforward and one object, such as a map record, could be the primary object. However, a single object does not include information about who has clearance to access maps, features on the map, or information about the staff responsible for looking after the task of scanning and digitizing the hard map Figure 8 displays the objects deemed to be required to meet the objectives of this project.

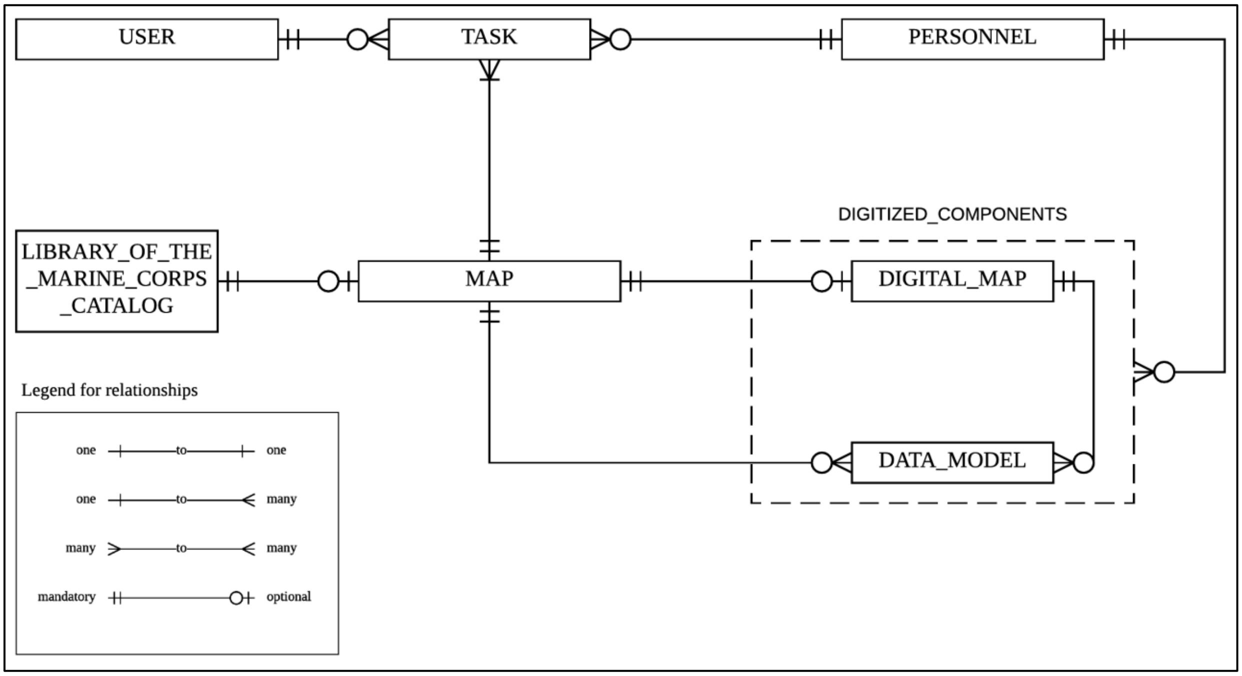

Figure 8 Entity-relationship model displaying just objects for the spatial database

The spatial database is comprised of seven different objects and nine relationships. These objects are MAP, LIBRARY_OF_THE_MARINE_CORPS_CATALOG, TASK, USER, PERSONNEL, and DIGITIZED COMPONENTS which contains DIGITAL_MAP and DATA_MODEL. There are a variety of one-to-one and one-to-many relationships, both optional and mandatory. The following paragraphs discuss the different objects and relationships present in Figure 8.

The object MAP is pivotal in the database as this object contains information about the hard map, links back to the Library of the Marine Corps catalog, is used by stakeholders, and connects to digitized components such as vector features gleaned from the map. MAP has a one-to-one relationship with the LIBRARY_OF_THE_MARINE_CORPS_CATALOG since each hard map is unique in the catalog and the catalog only has one record of that hard map. The relationship is mandatory between the hard map to the catalog since only some of the hard maps in the Archives and Special Collections are recorded in the catalog. Each time a hard map is involved in a task a unique record is created that links back to the catalog for the Library of the Marine Corps.

The object TASK is the active component of the database that links a user and personnel that work at the library to a hard map. A task must have only one user, but a user can initiate multiple tasks. A staff member from the library receives the task and may simultaneously work on multiple tasks. One hard map represents a single record in the object TASK. If the same hard map requires additional digitized components, the same map can occur again within the object TASK as a new record.

The object DIGITIZED COMPONENTS is a container for DIGITAL_MAP and DATA_MODEL since these two components have the same one-to-many relationship with the library staff. The container would not be present in the actual construction of the spatial database. The use of the container simplifies the diagram to display two components that are similar, reducing the number of lines required within the diagram. For example, the library staff member has the same relationship with both digital components, represented by one line. The relationship between the two objects means the personnel can work on one or multiple digitized components, while a digitized component can only have one personnel assigned. The MAP has two different relationships with the digitized components, which is why the relationship lines continue inside the container to the corresponding objects. MAP to DIGITAL_MAP is a one-to-one relationship because a user creates a task regarding a hard map and a library staff member scans the hard map. Once a library staff member scans a map, other personnel can create multiple data models. There is a one-to-many relationship between DIGITAL_MAP and DATA_MODEL since is it important to record which scanned map the data models were made from. As personnel can create multiple data models from a digitized version of a map, the hard map needs a one-to-many relationship to capture all the different models. As stated in Chapter 1, in the context of this thesis a data model is a vector file, such as a point or shape file containing spatial information and descriptions about an object

The object PERSONNEL represents library staff members, such as the librarian, GIS analyst, and IT personnel. As discussed in Chapter 3, each personnel have different capabilities and tasks involved in the process of completing a request for a user for a digitized version of a hard map. All personnel, regardless of ability, represent a single object as all personnel support MAP, TASK, and DIGITIZED COMPONENTS.

The object USER is the last object and is different from PERSONNEL, as the user is the object that initiates the task of processing a hard map into digitized components. The user only has one relationship in the diagram, which is to tasks as previously discussed. A library staff member is responsible for sending the digitized components to the user. It is not necessary to directly link USER to MAP or DIGITIZED COMPONENTS, since this is not vital information for either of these components. Instead, TASK captures this information if required by administration.

4.1.2. Attributes

Each object contains additional information describing the individual item and revealing the structure of the spatial database using different keys. Figure 9 expands on Figure 8 by providing the details of the attributes for each object. Each object contains a PK which is a unique identifier for different records. Since a PK is unique, the reference is not readily recognizable by users and staff members, so an actual key (AK) is used instead. An AK is normally a common name and recognizable by a user or staff member. In addition, some objects contain one or more FKs which represent a relationship with another object. The name of the FK in one object and the connecting PK in another object can differ as long as the content within the cells is the same, since this is what connects two records in different objects. The remainder of this section explains the different attributes within each object.

The object USER captures information regarding the person starting the request and information necessary for creating a profile to access the system. The object has one PK, UserID, and an AK UserName. The system automatically generates a unique PK. The user can create the AK. The user interface requires the Password attribute so the user can access the user interface for initiating tasks and searching for hard maps and digitized components. The Classification attribute acts as a level of access ensuring a user only accesses items with a suitable level of classification or clearance. The rest of the attributes consist of information contributed by the user which can be used for future analyses of the users, such as level of schooling of users collected by the 'Education' attribute.

The object TASK's main purpose is to combine information from other objects, so there are few attributes local to TASK. There is one PK, TaskID, a unique reference automatically generated by the system. There are three FKs, UserID, PersonnelID, and MapID. These are also unique references that link back to USER, PERSONNEL, and MAP respectively. The three unique attributes to TASK are DateStart, DateEnd, and Request. The DateStart automatically generates after a user submits a request. A library staff member enters the DateEnd in the same format in which the user has received the requested product(s). The Request is a copy of the comment from the user requesting digitization of a hard map and any other digitized components such as vector features. This is a text field and created for tracking and archiving purposes.

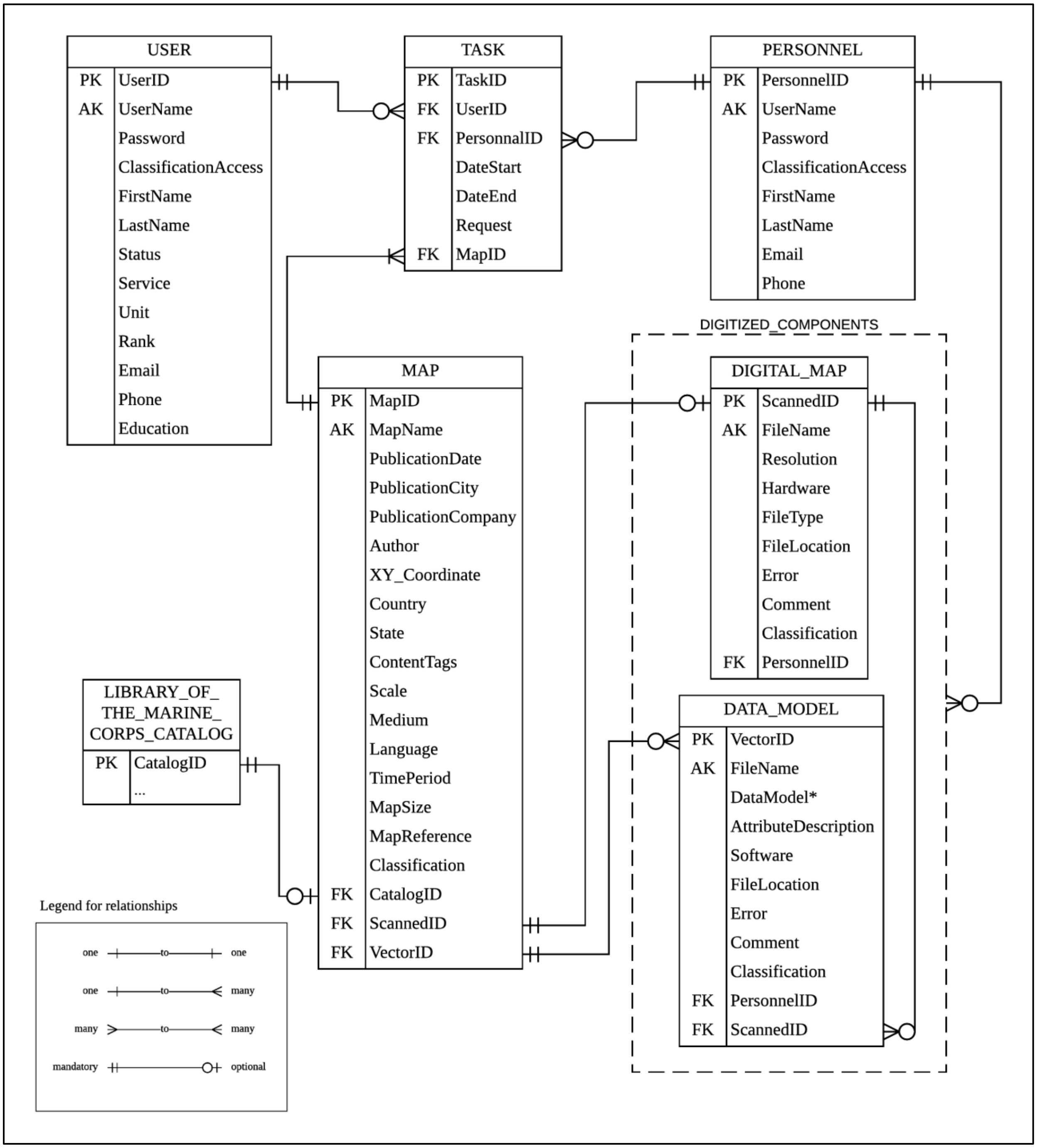

Figure 9 Entity-relationship model expanded to include attributes for each object

The object PERSONNEL is similar to USER in that it has one PK, PersonnelID, and one AK, UserName. The PK is a unique reference automatically generated by the system, and the AK is a unique reference created by the staff member. The staff member also has a level of classification which allows access to different classification documents. The other attributes consist of basic information collected about the staff members.

The object MAP contains the largest number of attributes compared to the other objects in the design. This is because MAP contains both metadata about the hard map and links to digitized components of the hard map. MAP also meets one of the objectives of this project by collecting information about search terms, geographic coordinates, and temporal information. A user can later use these attributes in query searches. In regards to structure, MAP has one PK, MapID, which is a unique reference automatically generated by the system. It has an AK, MapName, that is recognizable by users and library staff members. This attribute does not have to be unique, as there is the potential for different maps to have the same name. There are three FKs, CatalogID, ScannedID, and VectorID that connect MAP to LIBRARY_OF_THE_MARINE_CORPS_CATALOG, DIGITAL_MAP, and DATA_MODEL, respectively. The ScannedID and VectorID are the same in both objects since the system automatically generates these references. The format is not known at this time for the CatalogID. However the two must be the same for the relationship to work between the objects.

The object LIBRARY_OF_THE_MARINE_CORPS_CATALOG represents the Library of the Marine Corps catalog system. It is unknown what attributes the system presently maintains or the structure of the existing database. When this information is obtained in the future, the PK attribute name can be corrected and updated in MAP.

The object DIGITAL_MAP contains all the necessary information for when the hard map is scanned including the location of the scanned map on the server. DIGITAL_MAP has one PK, Scanned ID, that is a unique reference automatically generated by the system. It has one AK, FileName, that is unique and created by the library staff member when the file is generated by scanning the hard map. There is one FK, PersonnelID, linking PERSONNEL to DIGITAL_MAP. This allows a record to be kept of the staff member that scanned the hard map and entered in the information about the image file. If there are any discrepancies, the staff member can be contacted by the library. Additionally, by using the PersonnelID, rather than the UserName from PERSONNEL, the personal contact details of the staff member are protected. The rest of the attributes contain information about the scanner used to scan the hard map, the resolution of the image, the location of the file on the server, and any comments by the staff member during the digitization process. This information helps mitigate errors in the future use of this image.

The object DATA_MODEL is the next stage in processing the hard map after it is scanned. DATA_MODEL contains one PK, VectorID, which is a unique reference automatically generated by the system. Similar to DIGITAL_MAP it has an AK, FileName, that is a unique identifier generated by a library staff member once the data model is created. DATA_MODEL also has two FKs, PersonnelID, and ScannedID to create the relationships between PERSONNEL and DIGITAL_MAP respectively. PersonnelID is acting in the same manner as in DIGITAL_MAP providing a connection between the staff member that creates the various data models and the staff member's personnel information. The ScannedID links the digital map that is used during the process of extraction of various data models. This information provides a history for the data model and completes part of the metadata for the object. The rest of the attributes pertain to the type of data model created, what feature was extracted from the scanned map, what software was used to extract the feature, any errors encountered during the creation of the data model, and where the item is located on the server.

4.2 Evaluation of spatial database design

When designing a spatial database, there are three basic principles that a designer should adhere too: integrity constraints, business rules, and normal forms (Yeung and Hall 2007). These constraints ensure a simple and non-duplicating design. These principles were mostly adhered to when designing the spatial database to meet the objectives of this project. This result means the entity-relationship model presented in Figure 9 can be adapted to the Library of the Marine Corps software and system in the future. In addition, other organizations can adapt this entity-relationship model using different software, since the model follows the basic principles. An organization's software does not have to contain spatial capabilities. As this database stores the digitized components separately using a text or uniform resource locator (URL) field as a reference, there is no requirement for software to project the scanned image or vector file.

4.2.1. Integrity constraints

This design of the spatial database addressed some of the integrity constraints since the entity-relationship design only encompassed a portion of the constraints. In regards to the domain constraint, the majority of attributes were designed for single-valued attributes except for those listed in Table 6. After additional consultation with the stakeholder and software selected for construction of the spatial database and user interface, the composite and multivalued entries can be split into additional tables or ignored for simplicity purposes.

Table 6 Attributes that ignored the domain constraint

Object	Attribute	Composite/Multivalued description
MAP	XY_Coordinate	Contains a multivalued entry with both the X and Y coordinate. To reduce the number of attributes in MAP these values were kept together, but can be split into separate attributes during construction of database.
MAP	ContentTags	A map could have more than one searchable tag associated with it. Without access to example data it is unknown what the tags could involve. This attribute could use a gazetteer for tags or allow the librarian to enter tags. The first would remove duplication, while the second method would require additional rules when users search via content tags. This decision would be decided in consultation with the stakeholder.
MAP	Author	Contains a composite in two forms. The first contains both the first and last name of the author. The second is the possibility of more than one author. The decision to create a separate table with authors depends on the existing library catalog and stakeholder preferences.
MAP	MapSize	Contains a multivalued entry with both the width and the height of the hard map.
TASK	Request	Contains a composite entry as this is a copy of the text from the user requesting digitization of a hard map. This attribute could reference the hard map and the digitized components requested by the user.
DIGITAL_ MAP	Resolution	Contains the potential for a multivalued description if the height and width are different for the pixels selected for scanning.
DIGITAL_ MAP	Error	Contains a composite entry if text has references to multiple errors.
DIGITAL_ MAP	Comment	Contains the potential for a composite entry if the personnel enter a comment.
DIGITAL_ MAP	Hardware	Contains the potential for a composite entry if multiple hardware are used during digitization.
DATA_ MODEL	Error	Contains a composite entry if text has references to multiple errors.
DATA_ MODEL	Comment	Contains the potential for a composite entry if the personnel enter a comment.
DATA_ MODEL	Software	Contains the potential for a composite entry if multiple software are used during digitization.

As discussed in section 4.1.2, each object has one PK that is a unique reference automatically generated by the system. The automatic generation of a PK ensures there are no null values and stops a library staff member from accidentally creating a record without a PK.

This ensures the entity constraint and removes the additional rule during construction of the database that the PK is not null.

The referential constraint is unable to be enforced during the entity-relationship design stage. Each object discussed in section 4.1 has a PK and an FK with the same attribute name to help maintain the relationships between each object. The only relationship FK attribute that is unclear is the relationship between MAP and LIBRARY_OF_THE_MARINE_CORPS_ CATALOG since it is unknown at this time how that attribute is currently referenced in the library's catalog.

4.2.2. Business rules

Business rules continue to enforce the relationships between objects. The detail about the rule for each relationship is not explicit in Figure 9, however, it is possible to determine which relationships are mandatory and which are optional from the symbols in Figure 7. Table 7 explains in detail the business rules between the objects in Figure 9.

4.2.3. Normal forms

Normal forms are the next level of rules to ensure the integrity of the database. Normal forms are concerned with the relationships between attributes within an object. There are three rules, as presented in Table 8 that an object must meet to be in normal form. Consideration during the design of the spatial database included maintaining normal form. The final spatial database design unsuccessfully maintained normal form, as listed in Table 9. The breaking of normal form within the spatial design was a decision by the author and additional consultation with the client is required to fix any deviations from normal form.

Table 7 Business rules for objects presented in Figure 9

Objects	Business rule
USER and TASK	A task can only be initiated when a user makes a request
	A user can make multiple requests
PERSONNEL and TASK	A staff member must be assigned to a task at all times
	A staff member can be assigned to multiple tasks
TASK and MAP	A task must have a hard map
	The same map can be assigned to multiple tasks
MAP and LIBRARY_ OF_THE_ MARINE_CORPS_ CATALOG	Once a hard map is located it must have an entry in the Library of the Marine Corps catalog
	The hard map may not be present within the Library of the Marine Corps catalog prior to location by a staff member
PERSONNEL and DIGITIZED COMPONENTS	A digitized component must have a staff member assigned
	A staff member can be assigned to multiple digitized components
MAP and DIGITIZED COMPONENTS	A hard map may have one digital map or multiple data models
	All digitized components must have a corresponding hard map
DIGITAL_MAP and DATA_MODEL	A digital map can be used to create data models
	Data models must be created from a scanned map

Table 8 First, second and third rules for normal forms

Normal Form	Rule
First (1NF)	There are no repeating attributes in the table (that is, no two columns are allowed to store identical attributes, for example, the land use status of a parcel at different points in time).
Second (2NF)	The table is in 1NF, and all non-key attributes are functionally dependent on the primary key.
Third (3NF)	The table is in 1NF and 2NF, and there is no transitive dependency of attributes on the primary key ("transitive" in this context means indirect)

Source: Yeung and Hall 2007

Table 9 Objects that break the normal forms

Object	Rule	Reason
USER	1NF	There is the potential for Email and UserName to be the same if the process decides that a user name is to be the user's email address to help reduce duplication of accounts for the same individual. If this is the case, then one of the attributes needs removal to ensure consistency between the user name and the email.
PERSONNEL	1NF	There is the potential for Email and UserName to be the same if the process decides that a user name is to be the staff member's email. To improve security of the system this method is not suggested.
DIGITAL_ MAP	1NF	FileName is a partial duplication of FileLocation, excluding the address location of where the digital map is stored on the server. Currently it is acceptable to break this rule, so staff members can easily identify a digital map by its file name and it is clear in the design that there is a separate location for the digital map other than the database. To bring the object into 1NF the FileName can be changed to another recognizable format for the staff if decided by the organization.
DATA_ MODEL	1NF	Same problem and justification as DIGITAL_MAP.
MAP	1NF	MapReference is a combination of other attributes within MAP. It is currently an attribute to provide a bibliography or reference for the hard map in the entity-relationship model. With scripting this attribute can be removed and created in various bibliographic formats as requested by the user.

4.2.4. Test data

The next logical step is implementation and loading of data into the spatial database. After that evaluation of the spatial database can be continued by testing its usage. However, it was not possible to receive example data from the Library of the Marine Corps at the time this research was conducted, as previously mentioned. The author was unable to visit the library in person and request test data such as historical maps to scan and digitally archive. Therefore, an alternate method of evaluation of the success of this project was chosen, as described in the next section.

4.3 Evaluation of process

Chapter 3 introduced the outline of a process for a user requesting a digitized version of a hard map from the Library of the Marine Corps. The aim was to develop a process for using the spatial database. There were three secondary objectives associated with this aim. The first, as discussed in Chapter 3, was to explain the different interactions for requesting a digitized hard map. The other two secondary objectives, as discussed in Chapter 2, were scanning a hard map and extracting features from a digitized map. This chapter now evaluates the process introduced in Chapter 3 using the nine-stage GIS planning methodology listed in Table 4 (Tomlinson 2013).

The project scope changed to a conceptual study rather than a specific process and spatial database designed for the Archives and Special Collections branch, since the branch was not contactable at the time this project was conducted. The Library of the Marine Corps remained the representative organization for this project as Chapter 1 identified a need for the digitization of hard maps for users. This means that the nine-stages were only partly completed. Once successful consultation occurs with the Archives and Special Collections branch, the nine-stages can resume. The benefit of designing a non-specific process and spatial database means other organizations can adapt this process to make digitized hard maps available to users. Table 10 discusses the success of each stage for this project's process design.

The process design addressed six of the nine stages by conducting literature reviews, emails with other organizations, and reviewing available online documents from the Library of the Marine Corps. Therefore, the process presented in Figure 6 is a suitable product, for example for a GIS team to use when requesting funding for the planning stage of implementation of this project from the Library of the Marine Corps. It is anticipated that additional input from the

Library of the Marine Corps will change the rest of the products and documents in the other stages if there is a direct collaboration with the organization in the future.

Table 10 Comments on how the designed process met the nine-stages

Stages	Chapter	Comment
1	1	Chapter 1 identified the Library of the Marine Corps aims for "comprehensive storage, retrieval, analysis, and distribution of warfighting-related information" (Ramkey 2014, 2). The Archives and Special Collection is suitable for accountability of project's scope since the majority of hard maps are stored there. The Virtual Library branch may also measure success for the online user interface.
2	-	The author sent an email to the Greys Research Library and the response identified the Archives and Special Collections branch for consultation, even though it was in the process of moving and most likely would not respond. The author sent two additional emails to the branch for comment and there was still no response. The author did not present the different stages to the branch or receive funding. As this was a part of a thesis, the GIS team consisted of this author, two thesis advisors, and committee members.
3	1, 2	The author did not hold a technology seminar with the Archives and Special Collections branch. Instead the needs of the organization were identified by contacting other libraries and organizations that had hard maps or electronic libraries.
4	1, 2, 3	Figure 5 identified the tasks for four different stakeholders including input data and outputs. Chapter 1 and 2 identified the benefits of new products in the form of digitized hard maps and data models. The process or spatial database did not specifically discuss the tolerance of error. Each digital component did contain an error attribute, personnel attribute, and technology attribute in the metadata.
5	2, 3, 4	The majority of the data design was completed in Chapter 2 and Chapters 3 and 4 display the information in figures and tables.
6	3, 4	Chapter 3 introduces the spatial database and Chapter 4 evaluates the database design.
7	2	The author performed a literature review and contacted other libraries and organizations that had an electronic library to determine a scope for system requirements.
8	-	Not conducted.
9	-	Not written.

Chapter 5 Conclusion

The main aim of this project was to propose a design a process and spatial database for the Library of the Marine Corps, so remote users had access to digital copies of hard maps and associated digital components. This meant there were two objectives for this project. The first objective was designing a spatial database with objects, associated attributes, and relationships that supported a user requesting a digitized copy of a hard map, library staff scanning a hard map, a GIS analyst creating vector features such as text tables from a scanned map, and a user searching for a digital component. The second objective was designing a process showing the relationships between the different components of the user interface, web application scripting, spatial database, and separate digital storage. This final chapter summarizes how the components of this project met these objectives, and future work needed to completely fulfil the original goals.

5.1 Objectives

As a part of this project, the two primary objectives, to design a spatial database and a process to facilitate access to hard maps and digitized components from the Library of the Marine Corps, were met. The first objective was met by designing a spatial database based on a combination of background literature reviews and case studies relevant to this topic. However, the spatial database design was limited by lack of knowledge of the Library of the Marine Corps catalog or the current processes. Even without consultation, the design of the spatial database provided appropriate objects with the necessary attributes to capture metadata, digitized components, and relationships between hard maps and digitized components. The spatial database also includes attributes that facilitate searches in the object MAP based on terms and spatial and temporal information about the maps. The user interface and query tools could be developed to utilize these attributes to assist users in searching for the necessary hard map or digitized components. The second objective was met by designing a process for using the spatial database, as presented in Chapters 2 and 3. Chapter 2 covered the different methodologies for scanning hard maps and extracting spatial data and map features from hard maps. Chapter 3 showed the interactions of the spatial database and the user interface. Chapter 4 discussed the limitations of the process and that the majority of these limitations could be fixed through consultation with the Library of the Marine Corps. Overall it is assumed that the design of this project's spatial database and process could meet the needs of the Library of the Marine Corps to support remote user access to the hard map archive. The next step in this project would be to collaborate with the Library of the Marine Corps to build the spatial database and user interface and improve access to warfighting-related information to Marines and dependents located around the globe.

The overall success of the primary objectives of this project was considered positive. Nevertheless since the author was unable to construct a spatial database using data from the Library of the Marine Corps, the lack of example data highlighted two important facts.

Firstly, the Marine Corps does not have a process for supplying digital versions of maps to its stakeholders. The Marine Corps only station a portion of Marines at Quantico, Virginia, and the rest rely on community libraries or online access. If Marines could request the digitization of a hard map, this creates a priority for digitization of hard maps for the Library of the Marine Corps and improves the availability of warfighting-related information to Marines stationed around the globe.

Secondly, Chiang (2015) introduced a vision of having a connection between hard maps, digitized maps, and data models. As discussed in Chapter 2, the majority of information occurs in separate locations. The spatial database proposed in this project could provide a relationship bridge between the different objects on maps thus creating new information connections. While this spatial database proposed to use internal storage, the file paths to scanned images could be URL's that connect hard maps to data models completed by other organizations.

5.2 Advancing GIScience

The information and results in this thesis have implications for advancing GIScience. Currently, the science behind connecting a hard map, a digitized version of a hard map, and vector formats of features is not specifically summarized in the literature except in Chiang's (2015) vision paper. A spatial database and user interface provide the necessary components to connect these items together. The spatial database designed as part of this project provides a guide for how to connect the different components described in this effort and a user interface to access the different components. The obvious advantage of using a spatial database is to provide easier access to multiple digital components associated with a single hard map.

Also, other advantages of having a spatial database that links digital components to original hard maps include ease of access for researchers, greater access to historical information, and interest in software for extracting features from digitized maps. This literature review and reviews of other libraries' Web sites highlighted the lack of access to digitized maps and connection to vector objects digitized from maps. When a digitized map was available for download, often there were limited choices for file type, and there was no spatial component, except for maybe a textual tag. Researchers that work with GIS data spend a significant amount of time locating suitable data with the aim of answering a spatial question. Improving the ease of access for GIScience researchers to historical hard map archives could improve existing knowledge bases and vastly improve the time required to access these. For example, OCR software improves its performance in recognizing text labels correctly when analyzing multiple source maps from different time periods (Yu, Luo and Chiang 2016). Additionally, digitized historical information can contribute to temporal studies and improve gazetteers for given time periods. Finally, as additional digitized maps and historical data from hard maps become available and archives grow, the part of GIScience involved with advancing recognizing features on maps can further progress. Overall, improving access to digital maps and associated digitized components can assist the communities that work in GIScience or GIS by providing greater and timelier access to hard map data.

5.3 Future Work

The first priority for future work is receiving a response from the Archives and Special Collections branch. If this is not achievable, then the priority will be to establish a new library or consortium to work with. The Navy Department Library, Naval History and Heritage Command, could be a suitable organization, as this library is currently in the process of making maps available to the public. Since this is a project already underway, Tomlinson's (2013) nine-stages would involve a thorough investigation beginning at Stage 1 to understand the current situation and how GIS could assist the process. Other organizations that have similar parameters to the Library of the Marine Corps regarding limited access to digital data, use of war-related information by service members, and not being stationed near the central library, include the Naval War College, the U.S. Army Fort Leavenworth, and the Army War College. The spatial database and processes designed for the project could also be adapted and applied to these organizations. If applicable, then a GIS team could update the spatial database and process for creating digitized components.

If the GIS team establishes communication with the Archives and Special Collections branch, then Stage 2 is the starting point, and then the rest of the stages accordingly. It is vital to understand the existing library catalog and the specific hard map access processes currently utilized by the branch. The processes used by the branch could change the business rules governing the relationships between objects and attributes for objects.

After an organization is involved with the project, it is important to understand the needs, existing processes, and software and hardware of the selected organization. The GIS team can apply this information to the spatial database design, and then construct, implement, and load the spatial database with data. Next, testing and evaluation can occur and the database updated as required. Additionally, the IT personnel can create a user interface and storage location for the database and its associated files.

A preliminary user interface, displayed in Figure 6, had three proposed Web pages. The first contained the functionality for a user to search for a hard map or digitized components of a hard map using keywords or a spatial selection on a map. The second page was a user request form collecting information about the user and the request for a hard map and its digitized components. The last page was for staff members at the library to enter metadata information about the hard map and any digitized components. Consultation between the GIS team, the IT personnel, and the library determine the need for additional user interface pages and the detail on each page.

Once the GIS team, IT personnel, and organization complete the user interface and spatial database, the organization and users require instructions and tutorials on how to use the system. This involves creating additional products such as written documents with step-by-step instructions, video instructions, or classroom tutorials. These products allow users to understand how the system works and suggest methods for how students could utilize digitized components at school. The GIS analyst could also create instructions and tutorials on how digitized components work within GIS software to create maps and analyze patterns in data.

Additionally, the software and technology for extracting features from hard maps is a growing rapidly. Any future processes or designing of spatial databases should reference current literature and update accordingly.

Each of these future work tasks can assist a user's access to digital data, increase the amount of war-related information available for analysis, and advance the GIScience by encouraging the use of OCR software and reducing the amount of time required by a researcher to locate digital components after location of a suitable hard map. The information gathered in this project, the proposed process for a library to use a spatial database, and the design of the spatial database are ideas that can be used to improve access to digital hard map data that can be used to advance GIScience.

REFERENCES

Ahonen-Rainio, Paula. "Metadata for Geographic Information." *Journal of Map & Geography Libraries*, no. 2 (2006): 37-66.

Brotton, Jerry. *Great Maps.* New York: DK Publishing, 2014.

Campbell, James B., and Randolph H. Wayne. *Introduction to Remote Sensing.* 5th. New York: Guilford Press, 2011.

Chen, Aijun, Liping Di, Yuqi Bai, Yaxing Wei, and Yang Liu. "Grid computing enhances standards - compatible geospatial catalogue service." *Computers & Geosciences*, no. 36 (2010): 411-421.

Chiang, Yao-Yi. "Querying Historical Maps as a Unified, Structured, and Linked Spatiotemporal Source (Vision Paper)." *SIGSPATIAL*, 2015.

Chiang, Yao-Yi, and Craig A. Knoblock. "Recognizing Text in Raster Maps." *Geoinformatica* 19 (2015): 1-27.

Chiang, Yao-Yi, Stefan Leyk, and Craig A. Knoblock. "A Survey of Digital Map Processing Techniques." *ACM Computing Surveys* 47 (2014): 1-44.

Chiang, Yao-Yi, Stefan Leyk, Narges Honarvar Nazari, Sima Moghaddam, and Tian Xiang Tan. "Assessing the impact of graphical quality on automatic text recognition in digital maps." *Computers & Geosciences* 93 (2016): 21-35.

Durante, Kim, and Darren Hardy. "Discovery, Management, And Preservation of Geospatial Data Using Hydra." *Journal of Map & Geography Libraries: Advances in Geospatial Informaiton, Collections & Archives*, no. 11 (2015): 123-154.

Esri. *Story Maps.* 2016. https://storymaps.arcgis.com (accessed December 18, 2016).

Forward, Erin, Amber Leahey, and Leanne Trimble. "Shared Geospatial Metadata Repository for Ontario University Libraries: Collaborative Approaches." *New Review of Academic Librarianship* 21 (2015): 170-184.

Goodchild, Michael F., and Jianyu Zhou. "Finding Geographic Information: Collection-Level Metadata." *GeoInformatica* 7, no. 2 (2003): 95-112.

Hardy, Darren, and Kim Durante. "A Metadata Schema for Geospatial Resource Discovery Use Cases." *Code4Lib Journal*, no. 25 (2014).

Kollen, Christine, Cynthia Dietz, Joy Suh, and Angela Lee. "Geospatial Data Catalogs: Approaches by Academic Libraries." *Journal of Map & Geography Libraries: Advances in Geospatial Information, Collections & Archives* 9 (2013): 276-295.

Marine Corps Archives and Special Collections Branch. *Desert Shield/Desert Storm: A Register of Collections in the Marine Corps Archives and Special Collections.* Guide, Quantico: Library of the Marine Corps. Gray Research Center, 2010.

Marine Corps University. *Library of the Marine Corps Research Portal - Library of the Marine Corps.* January 14, 2016. http://guides.grc.usmcu.edu/lomc (accessed June 28, 2016).

Maslowski, Monika. "Personnal communication." *Systems Librarian/Webmaster and Interim Head, Virtual Library.* Quantico: Library of the Marine Corps, September 7, 2016.

Ramkey, Carol E. *Collection Development Plan for the Library of the Marine Corps consisting of the Marine Corps Research Library, the Archives and Special Collections, the Virtual Library, the Quantico Base Library.* Plan, Quantico: Marine Corps University, 2014.

Scaramozzino, Jeanine, et al. "Map Room to Data and GIS Services: Five University Libraries Evolving to Meet Campus Needs and Changing Technologies." *Journal of Map &*

Geography Libraries: Advances in Geospatial Informaiton, Collections & Archives 10 (2014): 6-47.

Stanford University Libraries. *Digital Production Group.* 2016. https://library.stanford.edu/research/digitization-services/labs/digital-production-group (accessed September 20, 2016).

Stanford University Libraries' Digitization Labs. Directed by StanfordDigitalLib. Performed by Stanford University Libraries' Digitization Labs. 2012.

Tomlinson, Roger. *Thinking About GIS: Geographic Informaiton System Planning for Managers.* 5th. Redlands, California: Esri Press, 2013.

Traill, Stacie. "Exploring the Terra Incognita of Access and Discovery: The Evolution of Cartographic Cataloging in the Twenty-First Century." *Journal of Map & Geography Libraries: Advances in Geospatial Information, Collections & Archives*, no. 10 (2014): 48-61.

Yeung, Albert K. W., and G. Brent Hall. *Spatial Database Systems: Design, Implementation and Project Management.* Dordrecht: Springer, 2007.

Yu, Ronald, Zexuan Luo, and Yao-Yi Chiang. "Recognizing Text in Historical Maps Using Maps From Multiple Time Periods." *Proceedings of the 23rd International Conference on Pattern Recognition.* 2016.