Journal of Map & Geography Libraries
Advances in Geospatial Information, Collections & Archives

ISSN: 1542-0353 (Print) 1542-0361 (Online) Journal homepage: http://www.tandfonline.com/loi/wmgl20

Functional Requirements Specification for Archival Asset Management: Identification and Integration of Essential Properties of Services-Oriented Architecture Products

Jonathan Wheeler & Karl Benedict

To cite this article: Jonathan Wheeler & Karl Benedict (2015) Functional Requirements Specification for Archival Asset Management: Identification and Integration of Essential Properties of Services-Oriented Architecture Products, Journal of Map & Geography Libraries, 11:2, 155-179, DOI: 10.1080/15420353.2015.1035474

To link to this article: https://doi.org/10.1080/15420353.2015.1035474

Published online: 04 Sep 2015.

Journal of Map & Geography Libraries, 11:155–179, 2015
Published with license by Taylor & Francis
ISSN: 1542-0353 print / 1542-0361 online
DOI: 10.1080/15420353.2015.1035474

Functional Requirements Specification for Archival Asset Management: Identification and Integration of Essential Properties of Services-Oriented Architecture Products

JONATHAN WHEELER and KARL BENEDICT
University of New Mexico, Albuquerque, New Mexico, USA

The complexity and size of geospatial data can constrain the capabilities of service providers and create risks to the long-term preservation and archiving of valuable information assets. While services-oriented architectures such as the Earth Data Analysis Center's Geographic Storage, Transformation and Retrieval Engine (GSToRE[1]) facilitate increased use and impact of geospatial data by mitigating these complexities by development of dynamic applications and interfaces, such services can often be primarily focused on the maintenance and delivery of only the most current versions of geospatial data that may nonetheless possess significant historical, cultural, or scientific value. Actions and documentation required to assure long-term preservation may not be supported by existing business models or may be otherwise compromised. However, general purpose archives offer a preservation capability that is complementary to the value created by dynamic service providers. We present an overview of the features of GSToRE, and the DSpace[2] repository platform and describe the requirements of a methodology for the harvest, quality assurance, and ingest of geospatial data into an institutional repository as a complement to the dynamic data access and visualization services provided by GSToRE and systems like it.

KEYWORDS *digital preservation, geoarchiving, institutional repositories, services oriented architectures, content transfer, open standards*

Address correspondence to Jonathan Wheeler, MSLIS, Data Curation Librarian, University of New Mexico, Centennial Science and Engineering Library, L259 A, Albuquerque, NM 87131. E-mail: jwheel01@unm.edu

INTRODUCTION

Geospatial data may possess historical, cultural, or legal value sufficient to justify their long-term preservation. Interest in these data and evolving government expectations within the United States and abroad toward the publication, sharing, and archiving of increasingly large and complex data assets have the practical impact of a twofold mandate: First, that data will be shared and readily accessible by the broadest possible user base at the time of their highest utility and currency. Second, that data which have been superseded or are otherwise of less immediate interest will be archived in support of continued, if limited, access as well as monitoring for required preservation actions including, for example, hardware changes and format migration. Whereas for many data types and formats these dual roles may be fulfilled by a single repository, the complexity and size of geospatial data can constrain the preservation capabilities of near-term service providers focused specifically on providing increasingly dynamic access to the most current data. Correspondingly, archives, and in particular general purpose archives such as the institutional repositories maintained by many academic libraries, possess the capabilities to preserve digital data but may lack sufficient domain expertise and technical architecture required to support dynamic, high-throughput geospatial data services and interfaces.

Because of the complementary nature of their respective functions, the potential for collaboration between data or service providers and institutional repositories is intuitive and a topic of ongoing interest (see, for example, Carly Strasser's January and February 2014 blog posts making the case for the complementarity of institutional and domain repositories[3]). In the most basic terms, the source provider facilitates data sharing and analyses while the archive offers potentially less interactive, if longer term, storage and quality assurance. However, while the preservation and maintenance of data integrity may be an essential function of an archive, in the case of data that have been dynamically enhanced or generated through their management within services-oriented architectures (SOAs), the question arises as to whether quality assurance is enough. Because the context and utility of such data are strongly dependent on the host environment, archival repositories may reasonably be expected to additionally support preservation actions sufficient to assess and maintain the fitness of the data for their potential future transfer to and exposure within an alternative SOA. At a minimum, this requires digital archivists to extend or enhance the content and metadata models supported by their repository application to incorporate access information or otherwise represent the services and capabilities of the source data provider.

In support of these objectives, we propose a model for the batch ingest and archiving of geospatial data into an institutional repository through which the requirements of preserving the fitness of the underlying data for specific services—as opposed to preserving the services themselves—are mapped to the capabilities and features of DSpace, an open source and widely adopted repository platform. Specifically, we identify methods by which the default DSpace metadata features can be customized and extended to capture complex geospatial metadata and data packaging specifications, with the additional inclusion of the source provider's Open Geospatial Consortium's (OGC's[4]) GetCapabilities requests as actionable URLs indexed and published with the corresponding asset metadata. This addition effectively creates an application layer that enables communication between the repository and the source SOA, provides repository managers with a means to monitor the status of the source provider and undertakes necessary preservation actions if and when the source provider is no longer able to support the originally specified capabilities. Further, it establishes the repository's ability to provide links to its users for access to the value-added services provided by the source provider's system.

Background

In an effort to develop a data management, discovery, access, and use platform that supports a wide variety of data types and discovery and use scenarios, the Earth Data Analysis Center (EDAC—with support from the National Science Foundation, NASA, and the New Mexico Legislature) has developed the GSToRE platform. GSToRE is a tiered data management system that supports both geospatial and nongeospatial data objects, a rich metadata model that can accommodate the range of documentation requirements of diverse data types and target representations (e.g., the Federal Geographic Data Committee's Content Standard for Digital Geospatial Metadata [CSDGM], ISO 19115 and related standards, Dublin Core, and JavaScript Object Notation[JSON]). It is also a Web-service-based access model that exposes data discovery, visualization, access, and administrative capabilities to client applications, using platform-specific representational state transfer (REST) service models and OGC Web Map (de la Beaujardiere 2006), Web Feature (Vretanos 2005), and Web Coverage (Whiteside and Evans 2006) services.

While GSToRE as a production architecture for rich interaction with diverse data types offers many advantages, it is designed primarily to provide a variety of value-added services built upon the data managed within the system, and was not developed as a long-term data archival solution. Recognizing the importance of the role of long-term archival access to data objects as a complement to dynamic data services, recent developers working on the GSToRE API have focused on facilitating the migration of data assets (with associated metadata) to archival systems such as the DSpace-based LoboVault system, the institutional repository system hosted and maintained by the University Libraries of the University of New Mexico.[5] This is a crucial step in the preservation planning and long-term sustainability of the data managed as part of the GSToRE collections, because in addition to offering an alternative discovery interface, institutional repositories provide complementary services including link resolution and file integrity validation not supported within GSToRE.

Because any attempt to preserve geospatial assets with reference to their actionable status within service-oriented architectures requires archivists to define the essential characteristics and preservation requirements of the tangible artifacts as well as the specification and documentation of value additions performed by the host service, a review of the complexities inherent in preserving and archiving geospatial data in the broader context is worthwhile. Naturally, these same issues attach to geospatial data maintained within and produced by services-oriented architectures and applications, but more importantly observations about the role and scope of activity for producers, consumers, and archives may provide insight into the potential capabilities of a successful preservation program. As discussed by Janée, Sweetkind-Singer, and Moore (2009), the principal actions of different stakeholders are influenced by the length of time over which each party or entity is expected to maintain authority over a given resource (p. 3). In this regard, whereas the emphasis within the SOA may be on providing the most up-to-to date version of a data set together with tools for dynamic analysis and visualization, the focus of the archive may be on publishing and documenting the relationship between multiple versions of a single, historically relevant data set over time. That is, the scope of dynamic services provided by the SOA may be broader than that of the archive, while the temporal scope of the archive will be broader than that of the SOA. Similarly, per further discussion by Janée et al. (2009) of preservation as a series of relay actions (p. 4), the functions and features of the archive will vary substantially from those of the SOA to the extent that the latter is designed to serve content to various clients for processing, while the former may primarily serve the content to other archives and repositories. Nonetheless, in consideration of the enhancements and value additions created by an SOA in advance of any transfer to an archive, there is a growing need for archives to preserve assets as functionally complete. That is, without preserving the features and capabilities of the SOA, archives must assure the fitness of preserved data for future incorporation within, and exposure by, comparable alternative systems.

Regarding geospatial data preservation issues in general and as described in the reports of the National Geospatial Digital Archive[6] and similar projects sponsored by the Library of Congress's National Digital Information Infrastructure and Preservation Program (NDIIPP),[7] significant risk factors for spatial data assets include the scale and size of the data, variations between image and tabular data formats, substantial metadata requirements for the preservation of sufficient context to promote comprehension and reuse over the long-term, and version control issues arising from the temporal quality of spatial data. In the context of the specific projects funded by the NDIIPP, see NDSA (2013) for a discussion of related findings of the North Carolina Geospatial Data Archiving Project (pp. 8–9); Janée, Mathena, and Frew (2008) describe the findings of the NDGA-sponsored research at the University of California, Santa Barbara (pp. 134–135).

The scale and size of geospatial data presents challenges on two fronts. First, the size of individual files can vary greatly depending on the data type, format, and resolution. Notably, the storage and preservation of high resolution raster files is problematic, as the size of full resolution images becomes a barrier to access across networks, while flattened or "dessicated" (Janée et al. 2008, p. 138) versions may be of limited quality due to data compression. Second, in addition to file size and, as reported by the National Research Council (2006) at the federal level, the short-term volume of data generated by remote sensing and satellite surveys can exceed multiple terabytes of space, sufficient to quickly exhaust the capabilities of many repositories and archives.

As noted by multiple authors, including McGavra, Morris, and Janée (2009) and more recently Locher and Termens (2012), spatial data are captured and represented using any number of a variety of proprietary and application-specific formats, most notably ESRI's shapefile (.shp). While there is some argument to be made for adopting the dominant proprietary format as a de facto preservation format, particularly as existing open formats have yet to achieve widespread adoption, such a strategy is only practical in the short term. Longer term archiving needs, which require that data and documents remain usable and accessible across decades—and, by implication, further require that the useful lifespan of selected assets exceed that of their creators and host institutions—are easily compromised by the closed nature of proprietary formats. In particular, due to the size and technical complexity of spatial data and as noted by Janée et al. (2008), there may be periods in the preservation of spatial data in which the designated stewards are required to take a minimal, "fallback" or "do-nothing" approach (p. 135). If changes in proprietary formats are not documented or monitored during such times, the required resources and capabilities for future archives and users to access and use the data are compromised.

Similar concerns extend to geospatial metadata as well. While some current standards, including Geography Markup Language (GML)[8] and the ISO 19115 family of related standards,[9] offer significant potential, complexity and other factors affect widespread and consistent adoption. The example of the Spatial Data Transfer Standard[10] provides a case in point: developed for and broadly adopted by federal and state agencies in the United States, the standard nonetheless has seen little uptake by commercial and international stakeholders (McGavra et al. 2009, p. 11). More recently, the capability of GML to act as a comprehensive data and metadata package inclusive of attributes and descriptors specified for the OGC's Web Mapping, Feature, and Context Services, offers spatial data service providers a single metadata standard promoting interoperability and comprehensive data exchange. As has been noted, however, the scope of the GML specification, including the ability to embed binary data, requires the development of application-specific profiles in advance of widespread adoption and implementation (Morris 2006, p. 299).

Adding further complication to the problem of metadata for geospatial data preservation is the scope of information required to support long-term comprehension and reuse. Spatial data are composed of a complex array of attributes, the loss of any of which may compromise the integrity or utility of the data. This complexity is both broad and deep: referring to information and metadata categories specified by the Open Archival Information System (OAIS) recommendation, spatial data routinely feature large and potentially fine-grained attribute sets detailing descriptive, administrative, technical, provenance, and rights information. As data are refined and augmented, they acquire further annotations, processing information, and so on. Consequently, some degree of expertise is typically required to minimally comprehend and make modest use of the data.

From the archival standpoint, the challenges presented by the breadth and depth of spatial metadata are twofold. First, the lack of a broadly adopted and consistently implemented standard complicates the identification of essential or required metadata and prevents interoperability. With regard to the concept of the designated community, defined within the OAIS Reference Model (2012) as "An identified group of potential Consumers who should be able to understand a particular set of information" (pp. 1–11), this lack of specification makes it difficult for archivists, who may be nonexperts, to capture and retain sufficient information to support the long-term utility of the data for a future community of experts. Second, because actions and annotations applied to spatial data are frequently specific to a particular use or outcome, much contextual information is on the one hand too narrow for archival purposes while on the other requires extensive documentation of provenance, source, and lineage. The question arises about whether and how to preserve and manage spatial data separately from, but with clearly expressed relations to, derived data products, including maps, reports, and data sets.

A third characteristic of spatial data that complicates preservation is the temporal nature of the data and, consequently, the frequent updates to which it is often subjected. Spatial data represent geographic features that are subject to change and recurring observation. Granular and high-interest features such as street center lines and cadastral plots are particularly subject to revision, yet as observed by Morris, Tuttle, and Essic (2009), the small, local organizations that produce and manage these data are often the most constrained with regard to preservation planning and archiving

(p. 527). As a result, depending on the varying policies of data producers, data may be frequently overwritten, resulting in a loss of important supporting and contextual information for decisions and processes based upon the superseded—often deleted—data.

Nonetheless, there is a strong rationale and requirement for the preservation of spatial data precisely because of these revision processes. Considered broadly, legacy spatial data possess historic and cultural value and document the recorded status of natural resources, political features and boundaries, and commercial or public infrastructure (Bethune, Lazorchak, and Nagy, 2009; Morris et al. 2009; NDSA 2013). By extension, historical spatial data may be necessary to resolve territorial disputes and for additional legal and legislative reasons.

The Geographic Storage, Transformation and Retrieval Engine

The Geographic Storage, Transformation and Retrieval Engine (GSToRE) has been developed as a platform designed to provide data discovery and access services as part of the U.S. National Spatial Data Infrastructure (NSDI [Clinton 1993]), as part of a broader Global Spatial Data Infrastructure, and more generally as a general-purpose platform for supporting geospatial (and nongeospatial) data discovery and access for multiple current and future applications. These broad objectives have driven the definition of the functional requirements of the platform and influenced the actual development of the system as the current foundation for the New Mexico Resource Geographic Information System (NM RGIS[11]), the New Mexico NSF EPSCoR Program's Data Portal,[12] and as the platform upon which the WC-WAVE NSF EPSCoR Tri-State Consortium[13] is building a "virtual watershed" data management platform for rapid data delivery to, and assimilation of, results for watershed modeling systems and data visualization systems. This section describes the functional requirements that have informed the development and evolution of the GSToRE platform and the specific capabilities that have been developed in response to these requirements; it also highlights the aspects of the system that are complemented by the capabilities of an archival platform such as LoboVault (DSpace).

The definition of the functional requirements for GSToRE as development was initiated in 2008 was driven by several factors:

1. Required support for the next version of the New Mexico Resource Geographic Information System—New Mexico's Geospatial Data Clearinghouse—replacing the NSDI Clearinghouse Node based on the Z39.50 application profile for Geospatial Metadata[14] established in the late 1990s and an initial online data search and access system that was developed and deployed in 2001

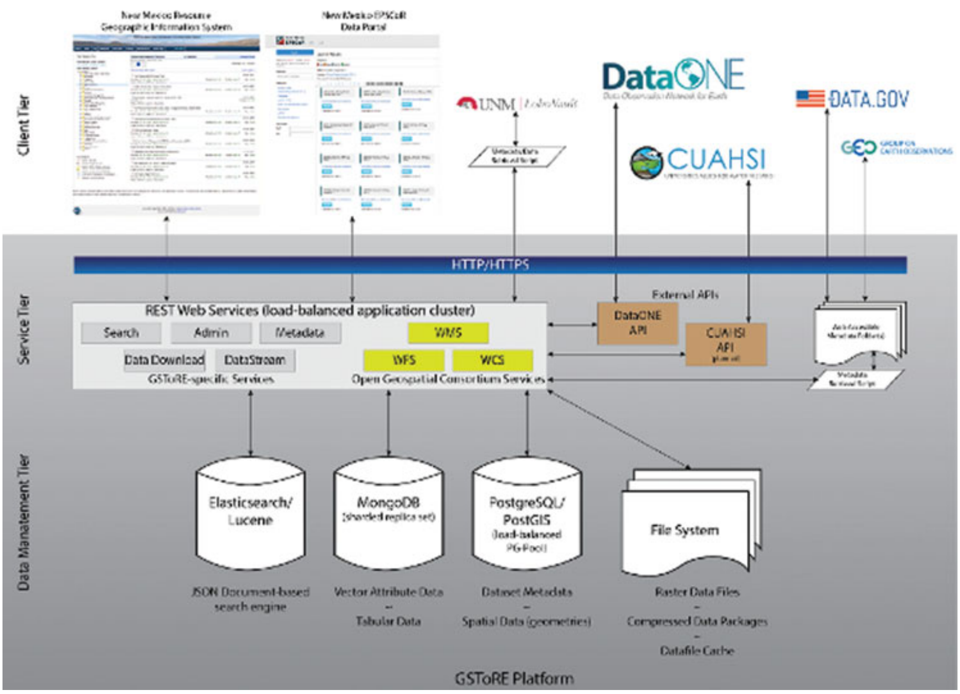

FIGURE 1 The GSToRE Tiered Architectural Model. At the bottom of the diagram is the *Data Management Tier* which contains the multiple databases and file systems components of the system. The middle *Service Tier* contains the code and Application Programming Interface (API) components that support client interaction with the managed data. The top *Client Tier* represents the various external applications that interact with the services published by the *Service Tier*.

2. Support for an expanded and diverse collection of research data products that were going to be generated by and acquired in support of two NSF-funded EPSCoR projects[15] and a NASA ACCESS Program[16]–funded data service and provenance capability development project
3. An overall strategic objective of developing a system based in a tiered SOA that provides well-defined Web standards–based interaction models for clients interacting with the data discovery and access capabilities of the platform, while also providing a degree of abstraction from the implementation details of the underlying data management systems with which the published services interact (Figure 1).

NSDI Participation

The first driver is directly related to the need to continue supporting the contribution that NM RGIS plays in the US NSDI—with the definition provided in Executive Order 12906 providing a foundational definition of how the US NSDI should be designed, and the implementation goals of the NM RGIS clearinghouse: "National Spatial Data Infrastructure (NSDI) means the technology, policies, standards, and human resources necessary to acquire, process, store, distribute, and improve utilization of geospatial data (Clinton 1993:1)."

The contribution of NM RGIS to the US NSDI includes the following milestones:

- initial development, in the late 1990s, of a Z39.50 (GEO Application Profile [Nebert 1999]) clearinghouse node (i.e., registered metadata search service) as part of the early US NSDI developed under the supervision of the Federal Geographic Data Committee (FGDC)
- development of a Web-based data discovery and download interface in 2001
- subsequent development[17] of Open Geospatial Consortium (OGC) Web Map Services (WMS, de la Beaujardiere 2006) for a limited set of data sets available through the RGIS Web site
- early contribution to the *Geospatial One Stop* data portal starting in 2002.

Building upon this history of development, the current NM RGIS NSDI capabilities include

- Registration of RGIS' Geospatial One Stop collection of over 2,500 FGDC metadata (FGDC 1998, 2002) to the current Data.gov platform's[18] metadata harvesting model as ISO 19139[19] metadata records that are written by an external metadata processing script into a Web Accessible Folder (WAF) that is registered with Data.gov as the source location for metadata used to populate the RGIS data collection
- Open Geospatial Consortium Web Map, Web Feature, and Web Coverage Services (WMS, WFS (Vretanos 2005), and WCS (Whiteside and Evans 2006), respectively) for geospatial data sets
- One or more available metadata representations (FGDC CSDGM, ISO-19139, Dublin Core,[20] Elastic Search JSON[21]).

While the completeness and accuracy of descriptive metadata for data sets managed within GSToRE depend on the information available from the originator of a given data set, additional metadata attributes are automatically generated for and inserted into each metadata record for the available data access services and metadata representations provided by GSToRE for each data set. Depending on the specific characteristics of each data set, links for service metadata for available OGC services (GetCapabilities requests in the service standard models for WMS, WFS and WCS), available data download formats, available metadata representations, and the source data upon which the data set within GSToRE is based are included within the metadata that are published through Data.gov, and available for download through the metadata links published by GSToRE. These embedded online linkage elements within the metadata are used by Data.gov to populate the user interface for discovered data sets with the links to the corresponding data and metadata items.

Specific Project Requirements

Building upon the core NSDI requirements described above, the GSToRE platform has been designed to be extensible, first, through expansion of the data storage and format options (additional source and output file formats); second, through the development of additional metadata content and representation options; and third, through expansion of the published Web services to provide additional machine accessible interaction models. Examples of project-specific GSToRE platform capabilities that have been developed or are under development include[22]

- development of an internal metadata schema and associated database content to support a wide range of both geospatial and nongeospatial data products
- implementation of support for external data discovery and access networks
  - The DataONE Earth Observation network[23]
  - The Consortium of Universities for the Advancement of Hydrologic Science (CUAHSI) hydrologic information network[24]
- implementation of experimental support for the World Wide Web Consortium's Provenance Ontology[25] as an additional metadata representation for selected data sets
- Web-service accessible data and metadata ingest capabilities to support linkages with watershed modeling systems for automated integration of model outputs into a separately running GSToRE-based platform
- repository support services designed to enable the identification of data sets that are designated for integration into specified repositories and networks and retrieval of information for those data sets to support access to related data and metadata for integration into those repositories
- implementation of the Network Common Data Form[26] data format support for expanded support of array-oriented scientific data structures.

General Architectural Strategy

Though the phrase, "services-oriented architecture" (SOA), had not yet been adopted as a descriptor for a Web-based architectural model that was based upon the publication and use of Web-accessible services for geospatial data discovery and access in 1993 (Erickson and Siau 2008), the definition and implementation of the US NSDI (provided above) and subsequent development of the NSDI Clearinghouse Node model and OGC service standards in the late 1990s and early 2000s as part of the NSDI set the stage for Earth Day Analysis Center's (EDAC's) focus on standards-based Web services as a core capability for the planned evolution of its data infrastructure. This focus began with EDAC's implementation of Z39.50 data search capabilities in the late 1990s, limited implementation of OGC WMS in 2001–2002 as part of the NM RGIS, and broader implementation of OGC WMS in conjunction with Simple Object Access Protocol[27] Web services behind the scenes, automatically generating those services and generating data/metadata packages.

Based on these experiences and a recognition by EDAC's development team of the streamlined implementation and adoption paths afforded by the Representational State Transfer (REST) Web services architectural model (Fielding 2000), GSToRE was designed as a tiered RESTful SOA that provides clear separation between the implementation of its data storage system and the service interfaces through which client applications access those data and their related services (see Figure 1). The RESTful service model is based on the core capabilities of HTTP (hypertext transfer protocol; Fielding et al. 1999; Fielding 2000), the core protocol that enables the request-response model for the Web, and, based on this foundation, provides a straightforward implementation path for both simple (e.g., read-only) and complex (e.g., read-update-create-delete) Application Programming Interfaces (APIs). The APIs published by GSToRE[28] are designed as publicly available RESTful services used by EDAC in developing its applications, but are also supported for use by other applications developed by project partners and others.

Resulting GSToRE Capabilities and Functional Emphases

As described in the preceding sections, the GSToRE platform has its roots in participation in the US NSDI and is based upon a SOA that emphasizes robust and flexible services for data discovery and access. This is well-aligned with the definition and focus of NSDI developments globally—an emphasis on data discovery and access through established standards and protocols. Beginning with the definition provided in Executive Order 12906 (Clinton 1993) and reflected in other definitions (e.g., Fu and Sun 2011; FGDC 2000; Nedovic-Budic, Knaap, Budhathoki, and Cavric 2009; de Andrade, Gomes, Baptista, and Davis 2014; Friis-Christensen, Ostländer, Lutz, and Bernard 2007; Crompvoets, Bregt, Rajabifard, and Williamson 2004; Kiehle, Greve, and Heier 2007; Rajabifard, Freney, and Williamson 2002), long-term archival storage is not an explicit focus of participants in the system. This lack of emphasis is reflected in the capabilities of the GSToRE system. While GSToRE is built and maintained following best practices for data storage, security, and

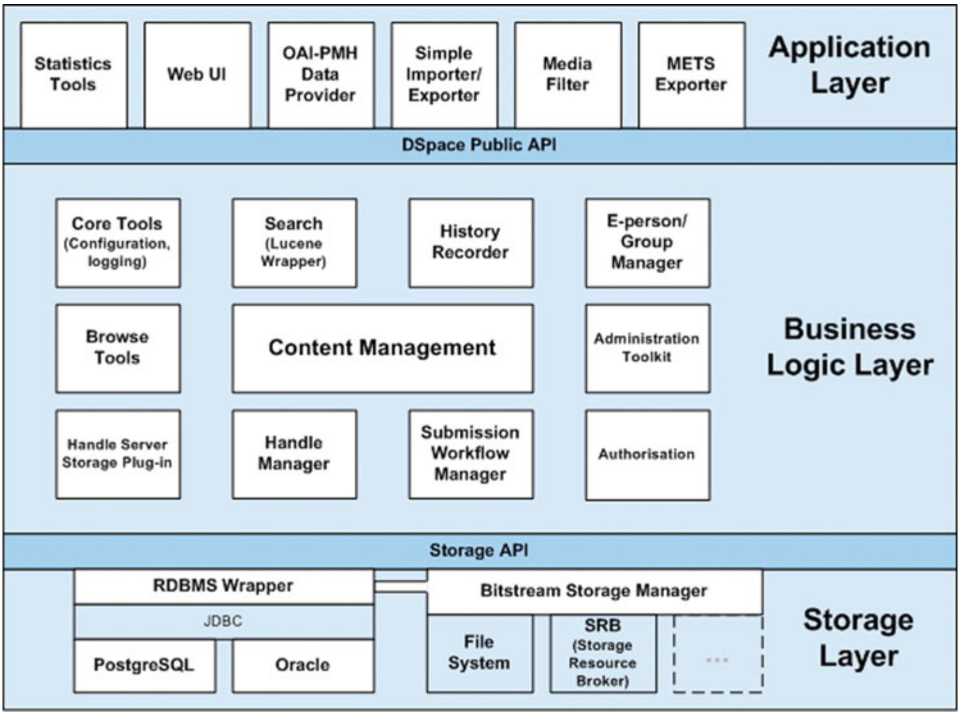

FIGURE 2 The DSpace Data Model, accessed from (https://wiki.duraspace.org/download/attachments/32473991/architecture-600×450.gif?version=1&modificationDate=126204313178x0&api = v2). This image is licensed under a Creative Commons Attribution-Share Alike 3.0 Unported License.

backup, it is not built upon an explicit archival model focused on ensuring long-term preservation of the data assets available through the system. This is where the archival focus of UNM's institutional repository, LoboVault, is able to provide a complementary suite of capabilities to those of GSToRE.

LoboVault: The University of New Mexico's Institutional Repository

The University of New Mexico Libraries' institutional repository, LoboVault, is an implementation of the widely adopted DSpace application and is the university's designated repository for administrative records, theses, dissertations, and scholarly content. As a format agnostic preservation platform, DSpace provides bit-level preservation of content files with associated metadata and supports discovery and federation of stored content via search engine indexing, including Google Scholar, and an Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) interface.

Developed to conform with the Open Archival Information System[29] specification, the DSpace data model (Figure 2 above) adheres to a threetiered architecture consisting of independent storage, business logic, and application layers across which communication is supported by a storage and a public API. Per the recommendations of the OAIS model, compliant systems support the long-term preservation of digital assets by identifying the characteristics and minimum required knowledge base for a designated community of content users (CCSDS 2012). The systems also establish metadata and content packaging profiles and processes sufficient to monitor the integrity of stored assets, together with sufficient contextual information to satisfy the minimum requirements of that designated community for reuse and comprehension of the information. To satisfy the information and knowledge base requirements of specialized communities such as GIS professionals, DSpace can be customized to support multiple metadata extensions and quality-assurance work flows.

Although natively configured to connect to a PostgreSQL database, DSpace is platform independent and can be used on top of an Oracle/MySQL or other enterprise database system. With regard to file storage, DSpace can likewise be configured to use any of an array of established and emerging technologies, including various network attached storage systems, iRODs based architectures, and CIFS and NFS file systems in combination with or as an alternative to native DSpace file system resources. This flexibility allows for multiple options when archiving large, disparate, and heterogeneous resources such as spatial data sets. In particular, the ability to register remotely managed assets within a DSpace repository offers promising efficiency and scalability potential. Metadata for registered assets is stored locally and the presentation and functionality of their item records are the same for users as locally stored bitstreams. While some configuration is required, the ability to grow the storage layer in this fashion can mitigate issues related to large file transfer and local storage capacity.

Communicating with the storage layer via a storage API, the business logic layer provides a comprehensive suite of repository tools for user- and access-control administration, content management, indexing, and browsing. Of particular interest in an archival and preservation context are tools for documenting provenance, curating assets via checksum and hyperlink validation, and flexible utilities for creating permanent identifiers, including Handles and Digital Object Identifiers (DOI). A recently added version control system offers some promise in managing relationships between current and historical or superseded versions of individual data sets.

The application layer provides multiple interfaces through which users can access or manage repository content. From the standpoint of archiving content from networked resources such as GSToRE, the batch import functionality allows repository managers to harvest, normalize, and ingest content through an efficient, streamlined, but still mediated, process. Additionally, the OAI-PMH interface facilitates federation and export of metadata to third party resources. While the default descriptive metadata element set is based on the general purpose Dublin Core metadata schema, the OAI-PMH interface can be extended by repository managers to expose an arbitrary number of metadata standards for discovery and harvest.

Additionally, the University of New Mexico Libraries has investigated multiple means of extending the capabilities of DSpace through the development of parallel services and features. As reported by Olendorf and Koch (2012), to mitigate the limitations of DSpace for storing and archiving large and complex data sets—characteristics relevant to geospatial data—extensions to the native storage architecture and metadata registry within DSpace have been implemented together with ingest and workflow routines designed to capture robust, standards-compliant metadata. In particular, data sets that are too large or that contain too many files to be efficiently exposed or represented by default item records are stored in alternative directories on the DSpace server. Per directory, a customized metadata page and file manifest is generated and provided as a means to navigate complex directory structures, with file level metadata included. A link to the alternative directory is provided in the corresponding item record.

The DSpace metadata registry is accessible to system administrators and extensible with regard to ingest and data entry processes, as well as enhanced federation and discovery capabilities. Customized metadata entry fields are created by registering an appropriate namespace, for example the ISO 19115 application schema.[30] Once registered, fields and qualifiers within that namespace can be individually added to build a customized profile for associated assets. However, it should be noted that the DSpace metadata model is relatively flat compared with complex standards such as ISO 19115. In practice, descriptors that are logically the offspring of higher level "parent" elements cannot be defined as such, and ingested metadata are limited to atomic elements with optional qualifiers. Although these limitations result in some compromise, the overall impact on preservation and discovery can be mediated through archiving the full metadata representations as item-level bitstreams and developing robust XSLT dissemination crosswalks. More detail on these processes is provided below. Once incorporated, data entry routines utilizing the extended terms can be implemented using XSLT, while crosswalks between an added schema and Dublin Core can be imported to facilitate federation and discovery via the OAI-PMH interface.

Likewise, the system includes a format registry for specifying institutional support for custom formats. Although this does not in itself address the need for a comprehensive body of documentation and representation information as expressed by Janée et al. (2009), it does provide some capability for monitoring bitstream formats and triggering processes for format migrations as needed. Specifically, the format registry provides an important administrative function by documenting the level of organizational support available to preserve specific formats. For example, an institution may designate PDF (portable document format) as a fully supported format, confirming a commitment to monitor changes in the PDF specification and to globally validate and migrate PDF files as needed. Conversely, formats can be registered as known or unknown, with a corresponding reduction in the level of service or overall capabilities to manage files in the specified formats.

A final relevant feature of the repository is the set of curation tasks that can be scheduled or run on demand and provide a ready assessment of metadata and bitstream integrity as well as metadata link resolution. In the context of Open Geospatial Consortium services, this link-resolving capability offers the potential to monitor multiple potential host services over time, as the GetCapabilities links of both original provider and alternative services can be routinely validated.

Functional Requirements for Integrating Services-Oriented Architecture Assets into an Archival Framework

In addition to the complexities described above, further issues adhere to the data as assets maintained within a SOA. As noted in particular by Morris (2010) and McGavra et al. (2009), geospatial SOAs offer multiple advantages to consumers of spatial data. By providing a centralized means to store and manage large and complex data sets, spatial SOAs increase the impact and utility of the data and enable use by various third parties and nonexpert stakeholders. Notable examples include Web map mashups and volunteered spatial information (Bishop et al. 2013), which capitalize on a growing body of OGC or similar services to support a number of social and commercial functions. Importantly, by broadening the potential user base, geospatial SOAs enable efficient decision making by potentially resource-constrained localities and governmental, public, or other civic organizations.

However, this growing development and impact of spatial SOAs creates an archival imperative on two fronts. In the first case, as essentially stateless applications from the front-end user's perspective, SOAs do not preserve or maintain processing and analysis information supporting future reuse (Morris 2010). This is especially pertinent as underlying data changes over time, and there is generally no standard means by which users can return to a service and generate the same map or report that supported a previous decision. Because the products of geospatial SOAs are ephemeral, any decision to archive or preserve the products in themselves may necessarily have to be made by the users. The question of data provenance and data processing that produced the products is, however, of issue to service administrators. The ability, at a minimum, to archive a snapshot of the underlying data would go some way toward addressing the legal and historical rationale for spatial data preservation as described above, particularly with products of significant archival or public value.

While it may generally be that the long-term preservation and curation of the products of geospatial SOAs is primarily the concern of the user, as noted, this does not minimize the archival imperative for service providers for whom additional significant concerns must be addressed. In particular, services such as GSToRE perform multiple enhancements and provide value additions to held data assets that can and ought to be preserved. Minimum provenance information to be documented in this case include not only data sources but also any relevant documentation of system processes such as metadata enhancement, integrity validation, file format migrations, and so on. As described above, additional metadata enhancements provided by GSToRE include available data access services, metadata representations, and alternative data formats. Because it may not be practical or possible to archive all the available data and metadata formats associated with a single item, it is necessary to identify in each instance the canonical version of data as well as metadata and to provide reference to the alternatives within the metadata record published by the archive.

Likewise, while SOAs create efficiency for users by minimizing storage and format requirements, issues of size, format, and versioning are of primary importance when transferring assets to external archival systems. Although a general purpose archive such an institutional repository may have multiple terabytes of space, as is the case with UNM's LoboVault, managers will have to concern themselves with the percentage of available storage that can be allocated to a single project or collection in addition to the potential impact of a large ingest on indexing and other system functions. As noted by Hoover (2012), substantial storage costs and systems maintenance requirements provide a strong rationale for libraries to build some resource sharing into producer-archive partnerships (p. 69). Ultimately, as with any archival acquisition, decisions must be made concerning prioritization of assets to be archived. In the case of GSToRE, which maintains assets in multiple formats, determination must also be made about which among multiple options to ingest. Finally, as versions of source data are updated, policies must be in place to address issues of accessioning newer versions of superseded data as well as whether to deaccession older versions. Beyond storage constraints, maintaining and documenting relationships between different versions of the same asset is an issue that may be unique to the archive: If the practice or policy of the host or originating service has been to simply overwrite superseded data, there may not be any existing mechanisms for establishing or documenting relationships between versions to which the archive can refer.

The development of sufficient metadata for the long-term documentation, preservation, and potential future transfer of assets into alternative architectures is another overarching concern. The archive will need to capture and make available any provenance information describing the original source of the data as well as actions performed by any intermediary services. Additionally, various pieces of information about the context and access and use requirements specific to the data that may be implicit in the SOA will have to be enhanced or made explicit upon transfer to the archive. Finally, to support decisions regarding required future format migrations, as well as whether to accession or deaccession different versions of the same data set, sufficient metadata must be captured to allow for triggering mechanisms to be specified and acted upon.

The above issues translate into functional requirements for archiving and preserving geospatial assets published through SOAs, several of which can be satisfied using the available features and supported extensions within DSpace.

With specific regard to the ingest, discovery, and long-term management of GSToRE assets within LoboVault, the following functional requirements can be described according to administrative processes, the logical and physical data models, and the metadata model.

Administrative Processes

Specification of the nature of expectations and the relationship between the archive and the producer is fundamental to archival practice, and with regard to digital assets is quantified to some extent by the OAIS and Producer-Archive Interface Methodology[31] standards, among others. Administrative actions and processes have been further formalized as essential characteristics of trustworthy repositories through both the Center for Research Libraries Trustworthy Repository Audit Checklist[32] and the Data Seal of Approval[33] criteria. They are accounted for via repository risk-assessment kits including the Data Asset Framework[34] and, the Digital Repository Audit Method Based on Risk Assessment.[35] Accordingly, even in the absence of certification, there is ample rationale for libraries and archives to develop work flows and policies that adhere to recommended practices and support transparency. Without addressing the full scope of recommended administrative policies, a minimum set of requirements includes the specification of the rights and responsibilities of the archive and the producer, the specification of copyright, use and access requirements pertaining to the data, and the method by which assets will be prioritized for transfer between GSToRE and LoboVault.

Specification of the Rights and Responsibilities of the Archive

In order to manage, disseminate, and preserve data assets, institutional repository (IR) managers require authorization from GSToRE administrators to perform any necessary actions or manipulations of the data including but not limited to virus scanning, checksum validation, replication and backup, and format migration of images, data files, and metadata.

Additionally, as applications, formats, and organizations change over time, long-term preservation of the data depends upon active monitoring by archive staff of the individual data assets as well as the organizational and technological context of the University of New Mexico Libraries and broader geospatial data management practices. Relevant trigger events and required follow-up actions may include changes in vendor and application support for image and data formats, development and succession of metadata standards, and institutional loss of funding.

Functional mechanisms for addressing these requirements include standardized forms and clerical tools for elaborating transfer agreements, curation plans, and potential format migration paths. Form elements and processes will map clearly to specific trustworthiness criteria and recommended risk management practices, and will be processed and maintained in a centralized, networked database.

Specification of Copyright, Use, and Access Requirements

Because LoboVault is in principle an open access repository and data assets maintained in GSToRE are already available via anonymous public access, there is no requirement to develop new or enhance existing access mechanisms within the University Libraries' DSpace instance. However, because copyright, use, and access requirements may be implicitly enforced by underlying processes within a given SOA or expressed as repository- or collection-level policies, rights and access metadata should be provided at the more granular, item level record within the IR.

Processes for harvesting assets will therefore identify and capture copyright and access requirements from the appropriate field in the GSToRE-supplied metadata for individual items, and the information will be mapped to the corresponding Dublin Core field in LoboVault. If no item level copyright information is available, the item will be flagged for follow-up review and missing copyright information will be determined according to the broader collection level or the most generally applicable GSToRE policy.

Prioritization of Asset Transfer

As noted in the discussion above, storage and resource constraints will likely prevent the wholesale transfer of assets from any spatial data service into a single, general purpose archive such as LoboVault. To support the assessment and preservation of the most valuable assets, mechanisms for supporting the automated assessment and identification of priority assets per the objectives of an established collection development policy are required.

In the case of a GSToRE harvest into LoboVault, a combination of item-level statistics and archive readiness and specific designation for integration into LoboVault will be retrieved via the GSToRE API. Items thus identified for potential archiving will be further assessed per the mission and collection policies of the GSToRE and the University of New Mexico Libraries.

Logical and Physical Data Model

Perhaps the most difficult requirement to satisfy when archiving spatial data relates to the physical format. Format concerns are significant because they are multidimensional and cut across data types and granularities, from published cartographic representations to packaging and container formats to the individual files and datum that make up the components of both dissemination and container files. Any decision to maintain or convert files between formats will carry some risk, and as noted above, documented migration paths and environmental triggers must be specified in policy.

Storage Allocation and Cross Registration

Regardless of which data will be preserved in particular formats, before an archival harvest takes place, policies and procedures necessary to provide sufficient storage must be in effect. Further, because storage capabilities in excess of available DSpace space may be required, processes for ingesting assets into external storage services and cross-registering their associated metadata will be developed.

Image and Data Format

Images and tabular data will be harvested from GSToRE according to available resolutions of the identified version of record and transformed as needed to preservation formats described by the Sustainability of Digital Formats Planning for Library of Congress Collections, Geospatial Information Recommendations.[36] An inventory of files and formats will be maintained by archive administrators, together with relevant available documentation and standards as identified by PRONOM[37] and the Unified Digital Format Registry.[38]

Within LoboVault, the DSpace format registry feature will be updated to specify levels of support provided for individual formats.

Container or Packaging Format

The GSToRE platform provides a variety of data/metadata packaging options, ranging from a package containing the source data and documentation used to create the GSToRE data objects, to packages that include specifically requested combinations of data formats and metadata representations. While these published packages (as ZIP archives) are potentially reusable within LoboVault, the data files, metadata representations, and available data packages provided by GSToRE will be treated as atomic data objects from which archival packages maintained by LoboVault will be built.

Metadata Model

A third general set of requirements relates to metadata. From the standpoint of preserving the value additions and characteristics of service-oriented architectures, at least in the case of GSToRE, it is primarily through extended metadata profiles that the dependencies between the data and the SOA will be expressed and maintained. Because of the constraints created by the "flatness" of the DSpace metadata model as described above, in addition to the mappings described here, it will be necessary to include the complete ISO 19115 metadata record in XML format as a content bitstream. This enables the archive to not only preserve the metadata as provided, but to also provide a serialized HTML transformation which can be accessed from the item record. Although archiving the metadata in this way does not expose the complete attribute set to search engines or the archive's discovery layer, it provides a means for users who have discovered the data to make a full assessment of the context of an item and its fitness for a specific purpose, and enables automated processing of the stored metadata by external systems that interact with LoboVault via the available DSpace API.

Also, whereas the flattening of a complex metadata standard for inclusion within DSpace is a factor of the system's data entry, ingest, and discovery functions, in support of federation by external resources via the OAI-PMH interface the archive will create dissemination crosswalks which more accurately represent the structure of metadata profiles such as ISO 19115 and FGDC. This capability is demonstrated within the default DSpace configuration by the inclusion of both METS and MODS profiles among the available OAI-PMH metadata sets. While still largely oriented toward document description and the information science domain, both standards are considerably more complex than Dublin Core and offer useful examples for developing additional dissemination crosswalks that map to geospatial metadata standards.

Descriptive Metadata

Descriptive metadata will be mapped from ISO 19115 or FGDC metadata provided by GSToRE and crosswalked to appropriate DSpace Dublin Core fields in support of discovery and browsing.

Spatial Metadata

The DSpace metadata registry feature in LoboVault will be extended to include key spatial elements in the ISO 19115 format, including coordinates and projection information as well as ISO 19115 keywords.

GetCapabilities

In order to monitor the status of the data source provider, links to the GetCapabilities features provided by OGC services will be captured as actionable URLs for routine validation by the metadata link resolver within LoboVault. While essentially a curation feature, by automating communication between the archive and the provider, this addition effectively adds the OGC service to the DSpace application layer and provides archive managers with a means to perform timely preservation or migration actions on archived data in the event of a change in the status or capabilities of the OGC service provider. Specifically, reported broken GetCapabilities links serve as trigger events and identify data sets requiring action.

Further, the harvest of selected GSToRE assets will include a download of their associated GetCapabilities responses for archiving as XML bitstreams. As data provider and OGC services evolve over time, these captured responses will provide important technical context as snapshots of the SOA environment at the time of harvest.

Administrative and Rights Metadata

As noted above and with particular regard to rights metadata, information regarding the ownership, use and access requirements, and copyright of GSToRE assets will be captured during harvest and mapped to DSpace Dublin Core "rights," "publisher," and appropriate "contributor" fields for each individual item.

Provenance Metadata

A final set of requirements pertains to the capture and expression of provenance metadata. This information is maintained by GSToRE at the item level within the FGDC and ISO 19115 fields. Processing instructions will be mapped to the Dublin Core description field to provide contextual information, while chain-of-custody information will be mapped to a specific provenance description field.

DISCUSSION AND CONCLUSIONS

As front end services and platforms evolve, the inherent complexities of geospatial data remain as barriers to the long-term archiving and preservation of geospatial data assets. In addition to extensive storage capabilities, the preservation of geospatial data requires the capture and maintenance of sufficient metadata to support the discovery, understanding, and use requirements of future researchers. Aside from information describing file format dependencies, map projections, and other technical details, preservation metadata can be expensive and time consuming to compile and necessarily includes documentation about the lineage, chain of custody, and access and use policies attached to the data. Finally, the packaging of geospatial data and images into predominantly proprietary file formats creates a risk of information loss if mappings to alternative formats are not clearly specified.

As the business case for preservation actions may not be articulated or else may exceed the capabilities of service providers primarily focused on the dynamic discovery and use of the most current data, a complementary role exists for general purpose repositories such as LoboVault to periodically harvest, assure, and archive geospatial assets held by GSToRE or other providers. By working with providers to identify collection priorities and characterize the value additions created by services such as GSToRE, repository managers and digital archivists can enable the archiving of geospatial data within preservation environments offering metadata quality control, bit-level file validation, and format migration capabilities over a period of decades.

NOTES

1. http://gstore.unm.edu
2. http://www.dspace.org/
3. http://datapub.cdlib.org/2014/01/30/institutional-repositories-part-1/, http://datapub.cdlib.org/2014/02/20/institutional-repositories-part-2/
4. http://www.opengeospatial.org/
5. https://repository.unm.edu/
6. http://www.ngda.org/
7. http://www.digitalpreservation.gov/
8. http://www.opengeospatial.org/standards/gml
9. http://www.iso.org/iso/catalogue_detail.htm?csnumber=26020
10. http://mcmcweb.er.usgs.gov/sdts/
11. http://rgis.unm.edu
12. http://www.nmepscor.org/data_portal/browse-data
13. http://westernconsortium.org/
14. https://www.fgdc.gov/standards/projects/GeoProfile
15. National Science Foundation (Track 1 [Awards: 0447691, 0814449, 1301346] and Track 2 awards [0918635, 1329470])
16. ACCESS-110018 NNX12AF52A
17. supported by funding from FGDC in 2001 as part of their WebMap CAP solicitation
18. http://catalog.data.gov

19. http://www.iso.org/iso/catalogue_detail.htm?csnumber=32557
20. http://dublincore.org/documents/dces/
21. http://www.elasticsearch.org/overview/
22. documentation for available GSToRE services may be found at the GSToRE API documentation website - http://gstore.unm.edu/docs/index.html
23. https://www.dataone.org/what-dataone
24. https://www.cuahsi.org/About
25. http://www.w3.org/TR/prov-o/
26. http://www.unidata.ucar.edu/software/netcdf/
27. http://www.w3.org/TR/soap/
28. Stable: http://gstore.unm.edu/docs/stable.html & experimental: http://gstore.unm.edu/docs/experimental.html
29. http://public.ccsds.org/sites/cwe/rids/Lists/CCSDS%206500P11/CCSDSAgency.aspx
30. http://www.isotc211.org/2005/gmd/applicationSchema.xsd
31. http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=39577
32. http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying/iso16363
33. http://www.datasealofapproval.org/en/
34. http://www.data-audit.eu/
35. http://www.repositoryaudit.eu/
36. http://www.digitalpreservation.gov/formats/content/gis.shtml
37. http://apps.nationalarchives.gov.uk/PRONOM/Default.aspx
38. http://www.udfr.org/

REFERENCES

Bethune, A., B. Lazorchak, and Z. Nagy. 2009. GeoMAPP: A Geospatial multistate archive and preservation partnership. *Journal of Map & Geography Libraries 6*(1): 45–56. doi:10.1080/15420350903432630

Bishop, B. W., T. H. Grubesic, and S. Prasertong. 2013. Digital curation and the GeoWeb: An emerging role for geographic information librarians. *Journal of Map & Geography Libraries 9*(3): 296–312. doi:10.15420353.2013.817367

CCSDS, June 2012. Reference model for an open archival information system (OAIS). CCSDS 650.0-M-2, Magenta Book.

Clinton, W. J. 1993. *Executive Order 12906—Coordinating Geographic Data Acquisition and Access: The National Spatial Data Infrastructure*. Washington, DC: Federal Register.

Crompvoets, J., A. Bregt, A. Rajabifard, and I. Williamson. 2004. Assessing the worldwide developments of national spatial data clearinghouses. *International Journal of Geographical Information Science 18*(7): 665–689.

de Andrade, F. G., C. de Souza Baptista, and C. A. Davis. 2014. Improving geographic information retrieval in spatial data infrastructures. *GeoInformatica: An International Journal on Advances of Computer Science for Geographic Information Systems 18*(4): 793–818.

de la Beaujardiere, J. (ed.) 2006. OpenGIS Web Map Server Implementation Specification, Version 1.3.0. Vol. OGC® 06-042: Open Geospatial Consortium.

Duraspace. DSpace 4.x Documentation. Last modified July 21, 2014. https://wiki.duraspace.org/display/DSDOC4x/DSpace+4.x+Documentation

Erickson, J., and K. Siau. 2008. Web services, service-oriented computing, and service-oriented architecture: Separating hype from reality. *Journal of Database Management 19*: 42.

Federal Geographic Data Committee (FGDC) Metadata Ad Hoc Working Group. 1998. Content Standard for Digital Geospatial Metadata. https://www.fgdc.gov/standards/projects/FGDC-standards-projects/metadata/base-metadata/v2_0698.pdf

Federal Geographic Data Committee (FGDC). 2000. Financing the NSDI—National Spatial Data Infrastructure aligning federal and non-federal investments in spatial data, decision support and information resources. Reston, VA: Federal Geographic Data Committee. http://purl.access.gpo.gov/GPO/LPS33018

Federal Geographic Data Committee—Standards Working Group. 2002. *Content Standard for Digital Geospatial Metadata: Extensions for Remote Sensing Metadata. FGDC-STD-012-2002*. Washington, DC: Federal Geographic Data Committee. https://www.fgdc.gov/standards/projects/FGDC-standards-projects/csdgm_rs_ex/MetadataRemoteSensingExtens.pdf

Fielding, R. 2000. *Architectural Styles and the Design of Network-based Software Architectures*. PhD Diss., University of California, Irvine.

Fielding, R., J. Gettys, J. Mogul, H. Frystyk, P. Leach, & T. Berners-Lee. 1990. Hypertext Transfer Protocol–HTTP/1.1 (RFC 2616). Retrieved 5/20/2011, from http://tools.ietf.org/html/rfc2616.

Friis-Christensen, A., N. Ostländer, M. Lutz, and L. Bernard. 2007. Designing service architectures for distributed geoprocessing: Challenges and future directions. *Transactions in GIS 11*(6): 799–818.

Fu, P., and J. Sun. 2011. *Web GIS: Principles and Applications*. Redlands, CA: ESRI Press.

Hoover, J. 2012. GIS collaborations in Saskatchewan: SGIC and the University of Saskatchewan Library. *Journal of Map & Geography Libraries 8*(1): 68–79. doi:10.1080/15420353.2011.622601

Janée, G., J. Mathena, and J. Frew. 2008. "A data model and architecture for long-term preservation." *Proceedings of the 8th ACM/IEEE-CS joint conference on Digital Libraries*. New York: Association for Computing Machinery.

Janée, G., J. Sweetkind-Singer, and T. Moore. 2009. "National Geospatial Digital Archive (NGDA)."

Kiehle, C., K. Greve, and C. Heier. 2007. Requirements for next generation spatial data infrastructures—Standardized Web based geoprocessing and Web service orchestration. *Transactions in GIS 11*(6): 819–834.

Locher, A.-E., and M. Termens. 2012. Exploring Alternatives for Geodata Preservation. Paper presented at the Sistemas y Tecnologías de Información: Acta de la 7th Conferencia Ibérica de Sistemas y Tecnologías de Información, Madrid, Spain.

McGavra, G., S. Morris, and G. Janée. 2009. Technology watch report: Preserving geospatial data. *Digital Preservation Coalition Technology Watch Series* Report 09 1 (2009).

Morris, S. P. 2006. Geospatial Web services and geoarchiving: New opportunities and challenges in geographic information service. *Library Trends 55*(2): 285–303.

Morris, S. P. 2010. Preservation of Geospatial Data: The Connection with Open Standards Development. In *Preservation in Digital Cartography: Archiving Aspects*. Markus Jobst, ed. (Urheberrechtlich geschutztes Material) Berlin, Heidelberg: Springer, 129–146.

Morris, S., J. Tuttle, and J. Essic. 2009. A partnership framework for geospatial data preservation in North Carolina. *Library Trends 57*(3): 516–540.

National Digital Stewardship Alliance. 2013. "Issues in the Appraisal and Selection of Geospatial Data: An NDSA Report." Chief author, Steven Morris. Retrieved from http://hdl.loc.gov/loc.gdc/lcpub.2013655112.1

National Research Council. 2006. *Preliminary principles and guidelines for archiving environmental and geospatial data at NOAA: Interim report*. Washington, DC: National Academies Press.

Nebert, Douglas D. 1999. "Z39.50 Application Profile for Geospatial Metadata or 'GEO' Version 2.2." Last Modified: 02 Oct 1999. Accessed May 20, 2011. http://www.fgdc.gov/standards/projects/GeoProfile

Nedovic-Budic, Zorica, Gerrit-Jan Knaap, Nama Raj Budhathoki, and Branko Cavric. 2009. NSDI building blocks: Regional GIS in the United States. *Journal of the Urban & Regional Information Systems Association 21*(2):5–23.

Olendorf, Robert, and Steve Koch. 2012. Beyond the low hanging fruit: Data services and archiving at the University of New Mexico. *Journal of Digital Information 13*(1). https://journals.tdl.org/jodi/index.php/jodi/article/view/5878/5882 (accessed on 7/7/15).

Rajabifard, A., M. E. Feeney, and I. P. Williamson. 2002. The Cultural aspects of sharing and dynamic partnerships within an SDI hierarchy. *Cartography 31*(1): 21–32.

Vretanos, Panagiotis A., ed. 2005. *Web Feature Service Implementation Specification, Version 1.1.0. Vol. OGC 04-094*: Open Geospatial Consortium.

Whiteside, Arliss, and John D. Evans, eds. 2006. *Web Coverage Service (WCS) Implementation Specification, Version 1.1.0. Vol. 06-083r8*: Open Geospatial Consortium.