One store has all?

– the backend story of managing geospatial information toward an easy discovery

Ningning Nicole Kong[1]

Abstract

Geospatial data includes many formats, varying from historical paper maps, to digital information collected by various sensors. Many libraries have started the efforts to build a geospatial data portal to connect users with the various information. For example, GeoBlacklight and OpenGeoportal are two open-source projects that initiated from academic institutions which have been adopted by many universities and libraries for geospatial data discovery. While several recent studies have focused on the metadata, usability and data collection perspectives of geospatial data portals, not many have explored the backend stories about data management to support the data discovery platform. The objective of this paper is to provide a summary about geospatial data management strategies involved in the geospatial data portal development by reviewing related projects. These data management strategies include managing the historical paper maps, scanned maps, aerial photos, research generated geospatial information, and web map services. This paper focuses on the data organization, storage, cyberinfrastructure configuration, preservation and sharing perspectives of these efforts with the goal to provide a range of options or best management practices for information managers when curating geospatial data in their own institutions.

Keywords

Geospatial information, data management, geoportal

Introduction

In United States, academic libraries began to provide geospatial datasets to their users from the 1990s when the census materials were given to depository libraries as TIGER (Topologically Integrated Geographic Encoding and Referencing) files (Gabaldón and Repplinger, 2006). The way how these datasets were managed and distributed by libraries depends on how they were acquired. In the early years, as these government datasets were distributed as part of the depository program, they arrived libraries as CD-ROMs (Abbott and Argentati, 1995). The datasets can be managed by libraries from their cataloging systems as same as any other electronic resources. As the ARL Geographic Information Systems (GIS) literacy program introduced more librarians GIS skills, library's GIS service expanded and reached more information users (Association of Research Libraries, 1999). Since then, academic libraries started to include more geospatial datasets into their collection development. The management and access to these geospatial data was one of the key challenges that librarians face. Many academic libraries have purchased computer hardware and software to store and manage these datasets, and provided users data access via library visits (Boissé and Larsgaard, 1995).

During the past decade, as technology evolves and the library's geospatial data collection expands, libraries began to explore new ways to provide access to geospatial information. Several reasons have contributed to this information access change. First of all, library's geospatial data collection has expanded to include a great variety of datasets, from the government datasets to proprietary datasets, including generic GIS files, scanned or georeferenced maps, regional collected aerial photos, as well as research data collection (Longstreth *et al.*, 1995; Bennett and Nicholson, 2007; Newton, Miller and Bracke, 2010). Since these data are scattered in different server spaces, it is difficult for geospatial information users to find their interested datasets. There is a need to have an inclusive geospatial information catalog serving for an easy data discovery. Secondly, the datasets previously distributed from the depository program became available online as public domain data. Often these data were difficult to find and difficult to use, so libraries began to provide links and general guides of using these datasets instead of directly managing them (Morris, 2006). Finally, the development of GIS industry has made it possible to distribute maps and geospatial datasets via web services instead of on-site visits.

As a logical response, many academic libraries started to develop geodata portal to facilitate federated search across different data provider services. Examples include the INSIDE Idaho project at University of Idaho, the Scholars Geoportal project at Ontario Council of University Libraries, and the Open Geoportal federation project initiated at Tufts University (Kenyon, Godfrey and Eckwright, 2012; Florance *et al.*, 2015; Trimble *et al.*, 2015). Geoportals provide a gateway to discover and access geographic information. With an effective design, users can discover their interested geospatial datasets from multiple data providers by a map, as well as a keyword or faceted search. Geoportals can improve data access and sharing between libraries, public domain platforms, as well as any other data providers, and greatly foster geospatial data discovery and usage.

However, academic libraries are facing many challenges in creating the geoportal. These challenges include how to choose and create the cyberinfrastructure for the geoportal development; how to effectively organize the geospatial metadata and develop the data catalog to support the geoportal; how to manage the library's datasets; and the portal usability design. Many studies have explored different aspects of these challenges, such as the geospatial data catalog, metadata schema, portal federation, and usability (Kollen et al., 2013; Hardy and Durante, 2014; Florance et al., 2015; Blake et al., 2017). Among these challenges, my research interest is particularly focused on how to manage the ever increasing datasets in the library's environment, so that the information can be well organized and preserved to serve as a reliable node for the geoportal. In this paper, I collected data management information from two different sources, including existing geoportal systems and OpenGeoportal participating institutions, in order to learn about the data management practices in different cases. With this information, I hope this paper will be a reference to help librarians to understand the current status, technology and major challenges about managing geospatial data.

Background

Geospatial data refers to the wide variety of datasets that have a geographic component, and that can typically be viewed as representing a portion of the Earth's surface in some way. Different with other electronic resources in library's collections, geospatial data are notoriously difficult to manage due to the following reasons. First, geospatial data doesn't have a uniform data model. It includes both vector and raster datasets and might reside in complex, multi-file objects. Many file formats used to store this information are proprietary and are linked directly to the version of the program in which they were designed. Especially in recent years, many of these data are being stored in relational geodatabases requiring sophisticated storage and archiving schemes (Janée, 2009). Second, geospatial data are often quite large with datasets commonly having gigabyte granularities and with some datasets growing by terabytes per day. The geospatial data collection efforts are on-going and long-lived programs. Due to the improved sensor technologies, geospatial data are increasing exponentially during the last decade (Chuck Killpack, 2011). The GIS data libraries are challenged by the increased requirements to collect, curate, and make available more data and services than ever (Goldberg *et al.*, 2014). Third, geospatial metadata may be voluminous because the associated technology and how it has changed over time need to be documented, which is difficult to find or not included with other information about the file itself (Erwin and Sweetkind-Singer, 2009).

Geoportal provides an effective way to exchange and sharing the complicated geospatial datasets between people and organizations. Academic libraries have started their efforts to adopt the geoportal technology into their own discovery tools, such as the development of the Alexandria Digital Library project (Smith and Frew, 1995) and the G-portal project developed at Singapore (Lim et al., 2002). In recent years, these efforts were boosted by the open source development concept, which allows more institutions to participate and contribute into the development of a single geoportal, such as the Open Geoportal federation and the GeoBlacklight community (Florance et al., 2015; Battista et al., 2017).

Geoportal is developed and implemented using three distributed components: a discovery interface or portal connecting to the metadata catalog; a set of web services which comply with existing standards for describing, accessing and exchanging digital data; and a data management system which provides a managed environment for both raster and vector geographic contents as shown in Figure 1 (Maguire and Longley, 2005; Tait, 2005). The first component is related to geospatial metadata. There are existing metadata standards such as ISO 19115 and FGDC (Federal Geographic Data Committee) content standard to describe geospatial information. Although challenges exist when document metadata using these standards and transform metadata records between the standards and library's generic metadata formats, many studies have addressed these issues and provided best management practices or crosswalk (Nogueras-Iso, Zarazaga-soria and Muro-medrano, 2005; Batcheller, 2008; Hardy and Durante, 2014). Starting from 2014, a GitHub repository, OpenGeoMetadata, emerged as an online space for libraries to share their geospatial metadata in an open, standards-agnostic way and has become an essential piece of infrastructure for building cross-institutional catalogs (Battista *et al.*, 2017; *OpenGeoMetadata*, 2018). It greatly fostered the collaboration between individual libraries. The majority of OpenGeoMetadata represents geospatial data objects held within the respective institution's repository as well as records extracted from government open data portals.

The second component is about standard web services. Standards specify communication protocols between data servers that provide geospatial datasets. They are used to ensure interoperability of different datasets (Strain, Rajabifard and Williamson, 2006). Several technical standards defined by the Open Geospatial Consortium (OGC) and the World Wide Web Consortium (W3C) play an important role in the dissemination of geospatial data (Bocher and Martin, 2012). For example, OGC has specified geospatial data delivery standards including Web Mapping Service (WMS), Web Feature Service (WFS) and its transactional equivalent (WFS-T), and the Web Coverage Service (WCS) (OGC, 2013). Specifications developed by the W3C for data dissemination include HTML, XML, SVG, SOAP, WSDL, etc. Libraries should consider to adopt these standards as much as possible when designing their geospatial data management systems so that data could be shared in an interoperable environment.

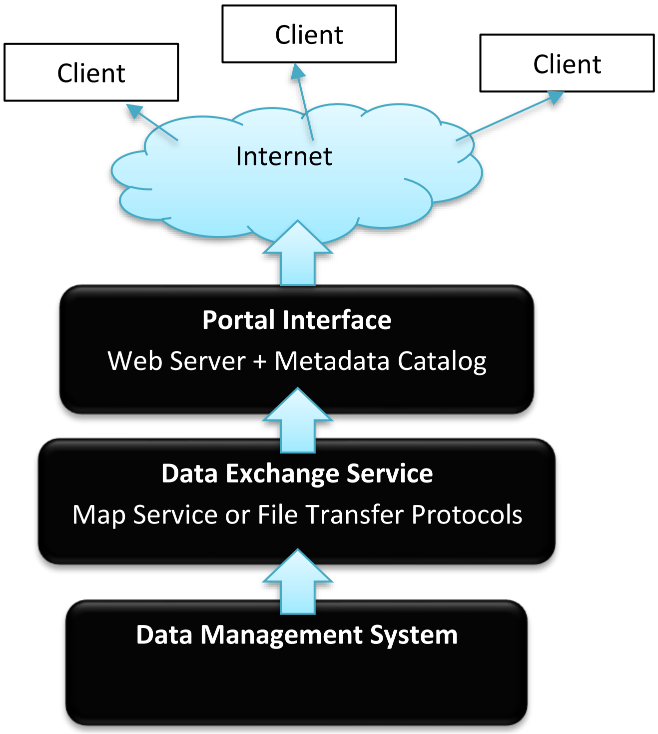

Figure 1. Components of a Geoportal infrastructure.

The third component of geoportal, which will be further discussed in this article, is a data management system. A geoportal is only as good as the information that is published through the site. Thus, it is essential to keep the datasets reliable, appropriate, and professionally maintained (Tait, 2005). In order to effectively share the datasets, we need to design the spatial data management infrastructure compatible with the existing standards. Many researches in the geographic information science has explored the efficiency of geospatial web services, which will provide great insights for building the data management system (Yang *et al.*, 2006; Li *et al.*, 2015). Academic libraries has also explored to manage geospatial data via their institutional data repositories (Newton, Miller and Bracke, 2010; Durante and Hardy, 2015). We will explore the data management options in various projects and analyze the capabilities offered by each option.

Geoportals have been existing outside of library's domain ever since 1990s (Tait, 2005). Government organizations or multi-institution projects have been using geoportal as a way to facilitate data access and sharing for various reasons. Initiated from government organizations, the United States National Spatial Data Infrastructure (NSDI) released the Geospatial One-Stop (GOS) geoportal in 2003, which aims to promote coordination and alignment of geospatial data collection and maintenance among all levels of government. In Europe, the INSPIRE geoportal intends to be Europe's Internet access point for geospatial data discovery, access and services (Crompvoets and Wachowicz, 2004). To facilitate multi-organizational project data sharing, different geoportals have been implemented for marine administration, forest management, countryside development, and biodiversity research (Askew et al., 2005; Strain, Rajabifard and Williamson, 2006; Flemons et al., 2007; Mcinerney et al., 2012). Examine the data management solutions in these projects as well as in the current library's practices would provide great insights for librarians to manage the expanding geospatial datasets.

Methods

In order to learn about the data management strategies in different geoportal projects, I have reviewed articles about geoportal project development. The data management portion of these projects has been analyzed and compared in order to generate useful references for librarians. Since different articles introduced their projects from different aspects, the data management solutions were not introduced in consistent details and procedures. Some of these papers focused on the planning stage of the project, while others focused on the geoportal outcome and assessment. I integrated the information extracted from planning concerns and assessment results, then summarized as the considerations that librarians should take in designing their geospatial data management system. I also listed possible options librarians could take, followed by specific examples in non-library and academic library settings.

In addition, I reviewed the metadata records contributed from OpenGeoportal community, and analyzed the online data link properties managed by different contributors in order to learn about the data management practices in each participating institution. With this information, I hope this paper will help readers to understand the current status, technology, and major challenges about managing geospatial data in academic libraries.

Results

Designing the Spatial Data Management Strategies

In the many existing geoportal development projects, two of them focused or introduced the planning stage of their projects. The needs assessment is essential as the first step to design a geospatial data management system. The purpose of such an assessment is to help understand the program goal, geospatial data properties, existing systems, expected GIS functionalities, etc., so that the data management system could be designed around these information. Smith *et al.* (2015) documented a needs assessment process that applied to one of the geospatial programs in National Park Service, which provided a comprehensive and logical workflow for the data system administrators and developers. In library setting, Shell Australia Technical Library has also conducted a very detailed initial assessment when designing the data management system, which includes the details about file size, data lifecycle, file formats, and management needs.

There are several considerations in designing the data management solutions for geoportal. First, how data will be organized, maintained and updated in the system? The data management portion of many geoportal projects intends to create a centralized system to store, describe, and manage the datasets from multiple sources. Planning for such a system requires to set up common practices for data formats, data organization rules, metadata schema and content standard. In addition, the different license agreements for various datasets need to be considered and access policy need to be designed in order to disseminate the data to appropriate user groups. Some projects also require the geospatial datasets to be maintained or updated on time to ensure the users can access the latest datasets. In those cases, a centralized enterprise infrastructure with version control is needed.

The second consideration is about the expected functionality for the data service, such as data downloading, visualization, and additional mapping functions. The most common function for many geoportal projects is downloading the datasets and simple metadata. Some portals were designed with a map preview function either as a thumbnail image or an interactive map so that users can have an idea about the datasets before downloading (Schwarb *et al.*, 2011; Trimble *et al.*, 2015). As the geographic analysis expands from its primary domain of geoscience to a broad range of disciplines, the new user groups need to access geospatial data collections in the way that reflects the current environment of web-based research and publishing activities (Durante and Hardy, 2015). So, web-based spatial visualization became a common requirement in the recent year's geoportal development. The Spanish National spatial data infrastructure has monitored the different types of visits on their geoportal. According to their reports between 2009 to 2012, the data visualization functionality was the most visits, which composed more than 80% of their yearly visits. Data downloading service was the second most visit type. In addition to data visualization and downloading, some geoportals also offer GIS functions such as location or attribute based query within datasets, data format conversion, projection conversion, etc. (Smith *et al.*, 2015; Trimble *et al.*, 2015). These additional requirements indicate that web-based map service is a must-to-have function in their portal, which requires a relational database management system (DBMS) and a spatial database engine to extend the functionality of the datasets and operate geospatial industry standards.

Finally, long-term preservation is another consideration for a data management system. Although it is not a frequently mentioned topic in most of the prevailing geoportal projects, it was studied in library-centered data management systems and the long-standing data programs such as the datasets managed by NASA (U.S. National Aeronautics and Space Administration) (Sweetkind, Larsgaard and Erwin, 2006; Durante and Hardy, 2015; Khayat and Kempler, 2015; Beaujardière, 2016). The long-term preservation requires a well-designed technical architecture which is usually independent with the routine geospatial data management system. Most of these projects have designated data repository as their preservation solution, and set up a set of preservation protocols, including format registry (to ensure the data format could be used for a long-term), metadata documentation, rights management and contracts, as well as collection management.

Spatial Data Management Options

Based on the current technology, there are many options that librarians could take in order to develop their data management system. Delivery of spatial data over the Internet can be realized in various ways ranging from data transmit via emails to OGC standard web services (Crompvoets *et al.*, 2004). While a simple file structure could serve for the purpose, many organizations have chosen a spatial DBMS because it offers extended visualization, mapping, and GIS capabilities. The major proprietary options of spatial DBMS include Oracle 10g Spatial and the Esri enterprise geodatabase which needs to be implemented on top of an enterprise DBMS choice, such as Microsoft SQL. These options provide most of the OGC standard web services and have been widely adopted in many U.S. government agencies and big organizations. The largest user base in free and open source software (FOSS) market is PostGIS, which adds spatial data types and analysis functions to the FOSS database PostgreSQL (Bocher and Martin, 2012). Comparisons with proprietary spatial DBMS show that PostGIS is a comparable alternative when considering functionality, robustness, support and price. Other FOSS spatial DBMS include MySQL Spatial extension which adds spatial support to MySQL, and the SpatiaLite project for SQLite.

Data Management Solutions in non-Library Geospatial Projects

In the non-library setting, there are two major kinds of geospatial data management approaches – file management system and spatial DBMS. The file management system is implemented by a metadata inventory which organizes the detailed information about each dataset in a centralized database. The file could be accessed by FTP, ODBC, or HTTP protocols. This approach is usually adopted in two extreme cases, either the datasets are simple and consistent enough to manage (Ganor, 2017) or the datasets are distributed on multiple servers in very different formats which is impossible to centralize them in one unified system (Tsontos and Kiefer, 2002; Johnson, 2017).

Spatial DBMS is used in most of the geospatial projects because of the additional spatial functionalities, especially the scientific visualization capability. One major kind of implementation is the FOSS option which includes PostGIS and Geoserver. Examples of this kind of implementation include the Forest Data Portal project at European countries (Mcinerney *et al.*, 2012) and the Tioga project which intends to provide data management for scientific visualization as part of the Global Change Research Project (Stonebraker *et al.*, 1993). The other major implementation is the adoption of Esri products including ArcGIS Server and SQL database as discussed by many U.S. and European geospatial portal projects (Maguire and Longley, 2005; Porcal-Gonzalo, 2015).

There are cases that more customized data management system needs to be developed in order to fit the projects' needs, especially in the case of handing big data. In one of the NASA's remote sensing data preservation projects, a customized system was developed based on Flexible Extensible Digital Object Repository Architecture (Fedora), which is a digital repository addressing the data management aspects of university and institutional libraries (Khayat and Kempler, 2015). NOAA managed its big data projects by exploring cloud-based infrastructure options including providers from Amazon, Google, IBM, Microsoft and Open Cloud Consortium. NOAA not only used these cloud service as storage spaces, but also explored additional functionalities from these platforms such as the Unified Access Framework that improved the discoverability and accessibility of regularly gridded data and the Dataset Identifier Project which assign DOIs to archived datasets (Beaujardière, 2016).

Data Management Solutions in Library Geospatial Data Projects

The data management system in academic libraries varies a lot, ranging from CD-ROMs, library maintained workstations to enterprise database. Serving for the purpose of geoportal, libraries mainly use spatial DBMS or institutional data repository as the options to manage their geospatial data collection.

More than a dozen of academic libraries have implemented OpenGeoportal project in their institutions (OpenGeoportal Community, 2017). OpenGeoportal provides data visualization and download functions for its users. Serving for spatial visualization purpose, libraries have the options to use either PostGIS and Geoserver technology stack, or Esri enterprise. Or, libraries can have the option of not providing the map preview. For the data download function, libraries can choose to provide data via Geoserver's processing function, or FTP/HTTP file transfer protocols, or through other data transfer format. Libraries can also build in their physical map collection into the system and provide the call number from the portal. In the next section, metadata shared from participating institutions will be further analyzed to review the data management options from each institution.

The INSIDE Idaho project uses SQL database for data management and ArcGIS server for web map visualization (Kenyon, Godfrey and Eckwright, 2012). In the Florida Geographic Data Library project, data quality assurance/quality control (QAQC) was a major concern in the project design. Spatial DBMS, more specifically, ArcGIS products and Oracle database were selected because the QAQC process is relatively complex and the DBMS provides an excellent environment to store, manage and relate information about the QAQC process (Goodison, Guillaume thomas and Palmer, 2016).

Drawing upon the increased needs of data management in academic libraries, digital repository is an excellent option for geospatial data management. There are several advantages of choosing data repositories for the geoportal purpose. First of all, many academic libraries already have digital repositories in their institution which accept geospatial datasets, so there is no additional requirement to create a new data infrastructure. Second, digital repository could include a wide range of dataset formats, which fits the requirements of geospatial data especially for collections from multiple sources. Third, digital repository manages dataset as distinct object which includes metadata, access policy and long-term preservation services. However, in order to visualize the spatial data from the portal, additional visualization function has to be developed and additional infrastructure such as a GIS server has to be added. Examples of this kind of option include the Geo-Hydra project at Stanford University and the geoportal developed by the Ontario Council of University Libraries in Canada (Durante and Hardy, 2015; Trimble *et al.*, 2015).

Spatial Data Management in OpenGeoportal Community

The participating institutions in the OpenGeoportal community shared their metadata database via the portal. I analyzed all the metadata records with an online link to understand the data management strategies in each institution. By the time we harvest the metadata for this analysis, there are 44,603 records with online link information. Among them, 75% records are geospatial information shared publicly, and the other 25% are restricted information only available to the specific institution or group of users. The participating institutions and their shared records are shown in Figure 2.

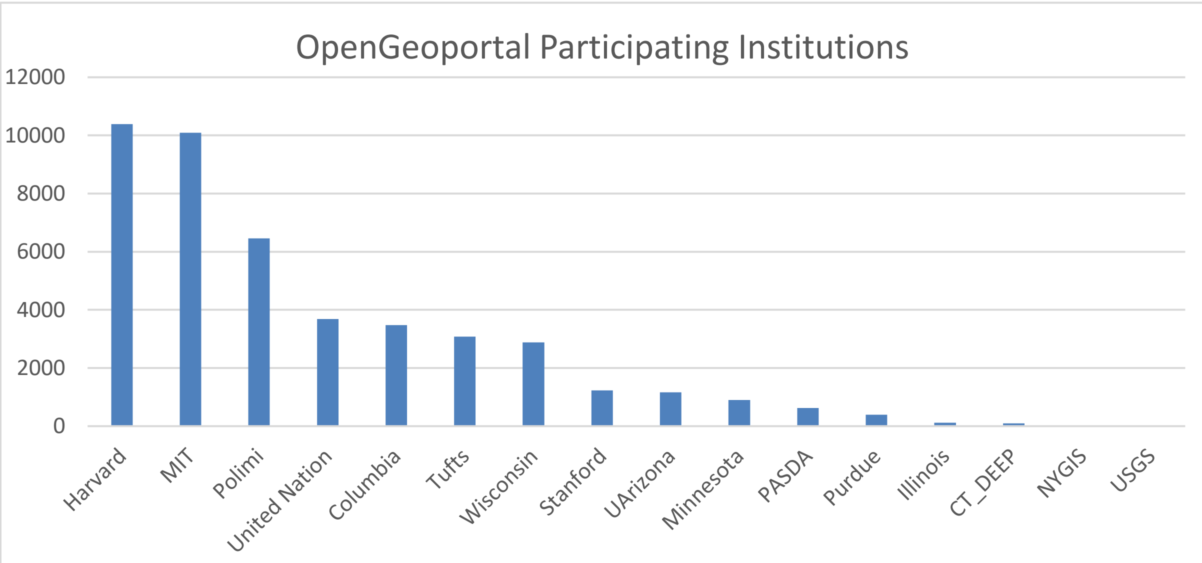

Figure 2. OpenGeoportal Participating Institutions and the number of records in their collections.

Figure 3 shows the data type of all the records discussed above . Overall, about half of the records are vector files saved as polygon, line or point type. Ten percent of the records are raster files, and another ten percent are scanned maps. Twenty-five percent are defined as paper maps. Five percent are undefined library records with an established web link. It must be pointed out that due to the limitations on metadata content standard, institutions might define their data type very differently. For example, "paper map" in Harvard University refers to those georeferenced paper maps which are available online as a web map; while in MIT literally refers to the library's physical paper map collection with metadata information existing in library's record system.

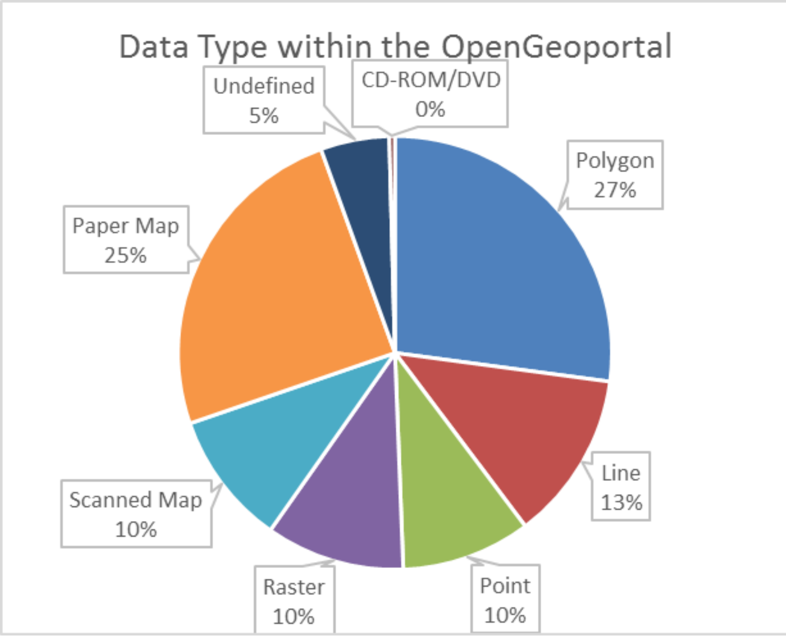

Figure 3. Different types data registered in the OpenGeoportal community.

In order to learn about the data management practices in each institution, I further analyzed the preview and download link for each vector and raster datasets (including the scanned or paper map records which refer to georeferenced datasets) managed by different institutions. Figure 4 shows the type of primary data links from each institution. The majority of the institution chose to use OGC standard web services, WMS, WCS, or WFS to disseminate their datasets online. WMS is especially the most popular choice and became the dominate type in many universities such as Harvard, MIT, Stanford and University of Arizona. Most of the institutions using these OGC standards manage their data using GeoServer and PostGIS database. Other than that, PASDA (Pennsylvania Spatial Data Access), Purdue, and a small portion of records from University of Minnesota use web services offered by ArcGIS server to manage their geospatial datasets, including Esri web map services, web feature services, and image services. A big portion of the records managed by United Nation's Food and Agriculture Organization (FAO) are managed by Geonetwork. GeoNetwork is a catalog application to manage spatially referenced resources, and was started in 2001 as a Spatial Data Catalogue System for the FAO. Without a spatial database engine, geospatial data can still be managed and transferred in a zip file format for users to download. The downloadable zip file via FTP or HTTP(S) is also a commonly used method within the community and it is at least used by five institutions.

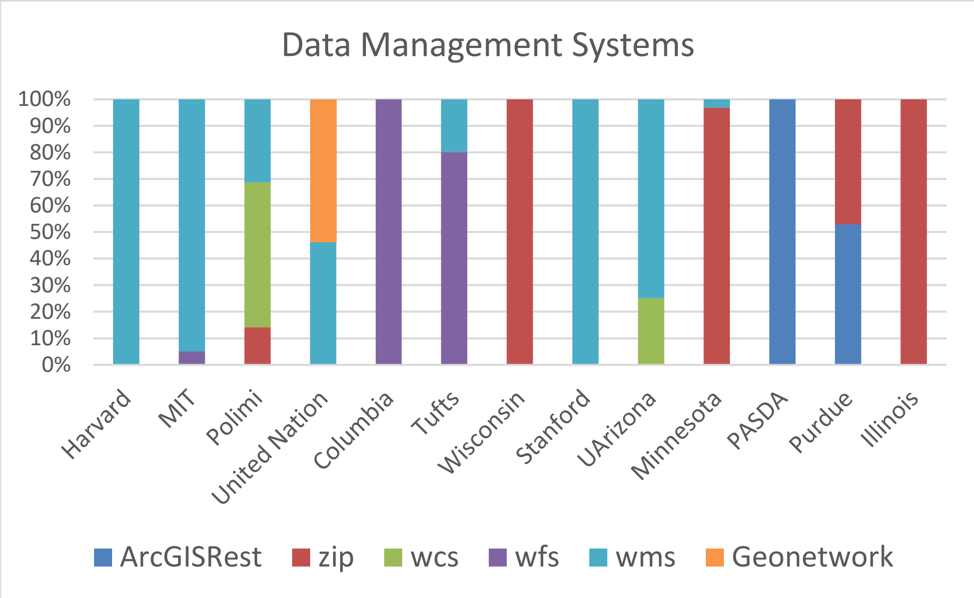

Figure 4. Data management strategies managed by different institutions.

Discussion

Managing geospatial data is never an easy task. A well-designed data management system along with good metadata practices are essential for academic libraries to manage the great variety of geospatial datasets. Visualizing the spatial data using a web interface will be the trend for general information users, as geospatial information is expanding across disciplines (Kong, Zhang and Stonebraker, 2014). This requires libraries to serve these datasets as web map services. To fulfil this need, libraries will need to maintain the geospatial datasets beyond their original formats and start to build spatial DBMS and spatial database engines such as GeoServer or Esri ArcGIS Engerprise to implement these services. In order to provide these map services in a stable, immediate, and up-to-date fashion, libraries will need a lot of management efforts to explore and compare different factors contributing to the map services, such as data formats, map service properties, and database versioning so that the data are managed with minimal storage space requirements while being delivered to the web browsers in an interoperable format with a satisfying speed (Kong, 2015).

Many libraries started to use digital repository as a platform to manage their geospatial datasets. Digital repositories is an excellent platform to centralize the datasets. However, many academic libraries are still in the early stage of employing data repository in their institutions and the collection policy might not include all the library’s geospatial datasets which were usually collected from different sources. More development needs to be made in order to combine the GIS server capability with the digital repository so that the geospatial datasets can be well maintained and disseminated.

Finally, it will need a long-term discussion and effort between librarians in order to preserve geospatial datasets. Although many agencies provide the current collected geospatial data, it is not in their responsibility to provide the previous decades data, even though these data could provide very important information for researchers. What is the mechanism to preserve these big datasets will be a challenging topic for libraries with increased data collection efforts and increased sensor technology.

References

Abbott, L. T. and Argentati, C. D. (1995) 'GIS: A new component of public services', *The Journal of Academic Librarianship*, 21(4), pp. 251–256.

Askew, D., Evans, S., Matthews, R. and Swanton, P. (2005) 'MAGIC: A geoportal for the English countryside', *Computers, Environment and Urban Systems*, 29(1 SPEC.ISS.), pp. 71–85. doi: 10.1016/j.compenvurbsys.2004.05.013.

Association of Research Libraries (1999) *The ARL Geographic Information Systems Literacy Project*, *SPEC Kit 238*. Washington, DC. Available at: http://www.istl.org/06-fall/refereed.html (Accessed: 1 May 2015).

Batcheller, J. K. (2008) 'Automating geospatial metadata generation-An integrated data management and documentation approach', *Computers and Geosciences*, 34(4), pp. 387–398. doi: 10.1016/j.cageo.2007.04.001.

Battista, A., Majewicz, K., Balogh, S. and Hardy, D. (2017) 'Consortial Geospatial Data Collection: Toward Standards and Processes for Shared GeoBlacklight Metadata', *Journal of Library Metadata*. Routledge, 17(3–4), pp. 183–200. doi: 10.1080/19386389.2018.1443414.

Beaujardière, J. D. La (2016) 'NOAA Environmental Data Management', *Journal of Map & Geography Libraries*, 12(1), pp. 5–27. doi: 10.1080/15420353.2015.1087446.

Bennett, T. B. and Nicholson, S. W. (2007) 'Research Libraries: Connecting Users to Numeric and Spatial Resources', *Social Science Computer Review*, 25(3), pp. 302–318. doi: 10.1177/0894439306294466.

Blake, M., Majewicz, K., Tickner, A. and Lam, J. (2017) 'Usability Analysis of the Big Ten Academic Alliance Geoportal: Findings and Recommendations for Improvement of the User Experience', *Code4Lib*, (38).

Bocher, E. and Martin, J. (2012) *Geospatial Free and Open Source Software in the 21st Century*, *Springer*. doi: 10.1007/978-3-642-10595-1.

Boissé, J. A. and Larsgaard, M. (1995) 'GIS in Academic Libraries: A Managerial Perspective', *The Journal of Academic Librarianship*, 21(4), pp. 288–291.

Chuck Killpack (2011) 'Geospatial content – Big data, Bigger opportunity', *Geospatial World*, 1(9), pp. 18–26.

Crompvoets, J., Bregt, A., Rajabifard, A. and Williamson, I. (2004) 'Assessing the worldwide developments of national spatial data clearinghouses', *International Journal of Geographical Information Science*, 18(7), pp. 665–689. doi: 10.1080/13658810410001702030.

Crompvoets, J. and Wachowicz, M. (2004) ‘Impact assessment of the INSPIRE geo-portal’, *Proc. of the 10th EC …*, (June), pp. 23–25. Available at: http://www.ec-gis.org/Workshops/10ec-gis/papers/23june_crompvoets.pdf.

Durante, K. and Hardy, D. (2015) ‘Discovery, Management, and Preservation of Geospatial Data Using Hydra’, *Journal of Map and Geography Libraries*, 11(2), pp. 123–154. doi: 10.1080/15420353.2015.1041630.

Erwin, T. and Sweetkind-Singer, J. (2009) ‘The National Geospatial Digital Archive: A Collaborative Project to Archive Geospatial Data’, *Journal of Map & Geography Libraries*, 6(1), pp. 6–25. doi: 10.1080/15420350903432440.

Flemons, P., Guralnick, R., Krieger, J., Ranipeta, A. and Neufeld, D. (2007) ‘A web-based GIS tool for exploring the world’s biodiversity: The Global Biodiversity Information Facility Mapping and Analysis Portal Application (GBIF-MAPA)’, *Ecological Informatics*, 2(1), pp. 49–60. doi: 10.1016/j.ecoinf.2007.03.004.

Florance, P., McGee, M., Barnett, C. and McDonald, S. (2015) ‘The Open Geoportal Federation’, *Journal of Map & Geography Libraries*, 11(3), pp. 376–394.

Gabaldón, C. and Repplinger, J. (2006) ‘GIS and the Academic Library : A Survey of Libraries Offering GIS Services in Two Consortia’, *Issues in Science and Technology Librarianship*, 48, pp. 1–8. doi: 10.5062/F4QJ7F8R.

Ganor, T. (2017) ‘An Integrated Spatial Search Engine for Maps and Aerial Photographs on a Google Maps API Platform’, *Journal of Map and Geography Libraries*. Taylor & Francis, 13(2), pp. 175–197. doi: 10.1080/15420353.2016.1277574.

Goldberg, D., Olivares, M., Li, Z. and Klein, A. G. (2014) ‘Maps & GIS Data Libraries in the Era of Big Data and Cloud Computing’, *Journal of Map & Geography Libraries*, 10(1), pp. 100–122. doi: 10.1080/15420353.2014.893944.

Goodison, C., Guillaume thomas, A. and Palmer, S. (2016) ‘The Florida Geographic Data Library: Lessons Learned and Workflows for Geospatial Data Management’, *Journal of Map & Geography Libraries*, 12(1), pp. 73–99. doi: 10.1080/15420353.2015.1038861.

Hardy, D. and Durante, K. (2014) ‘A Metadata Schema for Geospatial Resource Discovery Use Cases’, *Code4Lib*, (25).

Janée, G. (2009) ‘Preserving Geospatial Data: The National Geospatial Digital Archive’s Approach’, in *Archiving 2009 Final Program and Proceedings*. Arlington, VA: Society for Imaging Science and Technology, pp. 25–29. Available at: http://www.ingentaconnect.com/content/ist/ac/2009/00002009/00000001/art00007 (Accessed: 2 April 2018).

Johnson, V. (2017) ‘Leveraging Technical Library Expertise for Big Data Management’, *Journal of the Australian Library and Information Association*. Routledge, 158, pp. 1–16. doi: 10.1080/24750158.2017.1356982.

Kenyon, J., Godfrey, B. and Eckwright, G. Z. (2012) ‘Geospatial Data Curation at the University of Idaho’, *Journal of Web Librarianship*, 6(4), pp. 251–262. doi: 10.1080/19322909.2012.729983.

Khayat, M. and Kempler, S. J. (2015) ‘Life Cycle Management Considerations of Remotely Sensed Geospatial Data and Documentation for Long Term Preservation’, *Journal of Map & Geography*

Libraries, 11(3), pp. 271–288. doi: 10.1080/15420353.2015.1072122.

Kollen, C., Dietz, C., Suh, J. and Lee, A. (2013) 'Geospatial data catalogs: Approaches by academic libraries', *Journal of Map and Geography Libraries*, 9(3), pp. 276–295. doi: 10.1080/15420353.2013.820161.

Kong, N. (2015) 'Exploring Best Management Practices for Geospatial Data in Academic Libraries', *Journal of Map & Geography Libraries*, 11(September), pp. 207–225. doi: 10.1080/15420353.2015.1043170.

Kong, N., Zhang, T. and Stonebraker, I. (2014) 'Common metrics for web-based mapping applications in academic libraries', *Online Information Review*, 38(7), pp. 918–935. doi: 10.1108/OIR-06-2014-0140.

Li, W., Song, M., Zhou, B., Cao, K. and Gao, S. (2015) 'Performance improvement techniques for geospatial web services in a cyberinfrastructure environment - A case study with a disaster management portal', *Computers, Environment and Urban Systems*. Elsevier Ltd, 54, pp. 314–325. doi: 10.1016/j.compenvurbsys.2015.04.003.

Longstreth, K., Librarian, M., Library, M. and Arbor, A. (1995) 'GIS Collection Development, Staffing, and Training', (July), pp. 267–274.

Maguire, D. J. and Longley, P. a. (2005) 'The emergence of geoportals and their role in spatial data infrastructures', *Computers, Environment and Urban Systems*, 29(1), pp. 3–14. doi: 10.1016/j.compenvurbsys.2004.05.012.

Mcinerney, D., Bastin, L., Diaz, L., Figueiredo, C., Barredo, J. I. and San-Miguel Ayanz, J. (2012) 'Developing a forest data portal to support multi-scale decision making', *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5(6), pp. 1692–1999. doi: 10.1109/JSTARS.2012.2194136.

Morris, S. P. (2006) 'Geospatial Web Services and Geoarchiving : New Opportunities and Challenges in Geographic Information Services Brief Overview of Digital Geospatial Data Services in', *Geographic Information Systems and Libraries*, 55(2), pp. 285–303.

Newton, M. P., Miller, C. C. and Bracke, M. S. (2010) 'Librarian Roles in Institutional Repository Data Set Collecting: Outcomes of a Research Library Task Force', *Collection Management*, 36(1), pp. 53–67. doi: 10.1080/01462679.2011.530546.

Nogueras-Iso, J., Zarazaga-soria, F. J. and Muro-medrano, P. P. R. (2005) *Geographic information metadata for spatial data infrastructures*, *Springer*. doi: 10.1007/3-540-27508-8.

OGC (2013) *OGC Standards*, *Open Geospatial Consortium*. Available at: http://www.opengeospatial.org/standards/is (Accessed: 9 October 2013).

OpenGeoMetadata (2018). Available at: https://github.com/OpenGeoMetadata.

OpenGeoportal Community (2017) *Open Geoportal*. Available at: http://opengeoportal.org/ (Accessed: 26 March 2017).

Porcal-Gonzalo, M. C. (2015) 'A Strategy for the Management, Preservation, and Reutilization of Geographical Information Based on the Lifecycle of Geospatial Data: An Assessment and a Proposal Based on Experiences from Spain and Europe', *Journal of Map and Geography Libraries*, 11(3), pp. 289–329. doi: 10.1080/15420353.2015.1064054.

Schwarb, M., Acuña, D., Konzelmann, T., Rohrer, M., Salzmann, N., Serpa Lopez, B. and Silvestre, E. (2011) 'A data portal for regional climatic trend analysis in a Peruvian High Andes region', *Advances in Science and Research*, 6, pp. 219–226. doi: 10.5194/asr-6-219-2011.

Smith, J. W., Slocumb, W. S., Smith, C. and Matney, J. (2015) 'A Needs-Assessment Process for Designing Geospatial Data Management Systems within Federal Agencies', *Journal of Map and Geography Libraries*, 11(2), pp. 226–244. doi: 10.1080/15420353.2015.1048035.

Stonebraker, M., Chen, J., Nathan, N., Paxson, C. and Wu, J. (1993) 'Tioga : Providing Data Management Support for Scienti c Visualization Applications 1 Introduction 2 The Tioga Programming Paradigm', *Vldb*, pp. 1–12.

Strain, L., Rajabifard, A. and Williamson, I. (2006) 'Marine administration and spatial data infrastructure', *Marine Policy*, 30(4), pp. 431–441. doi: 10.1016/j.marpol.2005.03.005.

Sweetkind, J., Larsgaard, M. L. and Erwin, T. (2006) 'Digital Preservation of Geospatial Data', *Library Trends*, 55(2), pp. 304–314. doi: 10.1353/lib.2006.0065.

Tait, M. G. (2005) 'Implementing geoportals: Applications of distributed GIS', *Computers, Environment and Urban Systems*, 29(1 SPEC.ISS.), pp. 33–47. doi: 10.1016/j.compenvurbsys.2004.05.011.

Trimble, L., Woods, C., Berish, F., Jakubek, D. and Simpkin, S. (2015) 'Collaborative Approaches to the Management of Geospatial Data Collections in Canadian Academic Libraries: A Historical Case Study.', *Journal of Map & Geography Libraries*, 11(3), pp. 330–358. doi: 10.1080/15420353.2015.1043067.

Tsontos, V. M. and Kiefer, D. A. (2002) 'The Gulf of Maine biogeographical information system project: Developing a spatial data management framework in support of OBIS', *Oceanologica Acta*, 25(5), pp. 199–206. doi: 10.1016/S0399-1784(02)01209-4.

Yang, C. P., Cao, Y., Evans, J., Kafatos, M. and Bambacus, M. (2006) 'Spatial web portal for building spatial data infrastructure', *Geographic Information Sciences*, 12(1), pp. 38–43. doi: 10.1080/10824000609480616.

End-notes

[1] Nicole Kong is an assistant professor and GIS specialist at Purdue University Libraries. She can be reached by email: kongn@purdue.edu