The new trends for librarians in management of Geographic information

Ifigenia Vardakosta* and Sarantos Kapidakis

*Laboratory on Digital Libraries and Electronic Publishing
Department of Archive and Library Sciences, Ionian University
Ioannou Theotoki 72, Corfu 49100, Greece

Abstract

The purpose of this study is to examine the tools (e.g. CSDGM – FGDC, Gazetteers) used for the subject description and organization, of materials (maps, aerial photos, geospatial data, remote sense images etc) related to the geographic information in today's hybrid library environment. Furthermore we aim to identify changes (e.g. RDA) and determine the new forms (e.g. Linked Data) that have arisen in the management and use of geographic information in order to further enrich access to digital map collections and geographic databases. The intention of this work is to provide professionals in library world with an overview of information regarding terms and tools, issues that are essential in developing geographical collections.

This paper is structured in three parts. Initially, concepts related to geographic information and geographic collections in libraries are presented while the literature review in the specific field reveals the continuously interest of diverse scientific communities such as librarians, technicians, geographers etc on one hand and the continuous evolution of these tools as a result of technological development on the other. In the second part we attempt to provide examples of how librarians are handling geographical information as our research so far discerns while in the third part certain tools, that are most in use, are being described. Finally, the several aspects of these issues will be discussed.

Keywords: geographic information, geospatial collections, academic libraries, information management, metadata

*Corresponding author. Tel: +30-210-9549170; Fax: +30-210-9560161
E-mail address: ifigenia@ionio.gr
1. Introduction

Libraries traditionally have always been involved in providing geographical information (GI) through organized collections in both the conventional (e.g. printed maps, journals, books, atlases etc) and the digital environment (e.g. remotely sensed images, digital mapping data, aerial views etc) in recent years. Literature on the development of services related to GI is quite rich and supports the disposal of libraries for constant request so as to achieve improvable and innovative services for their users by adopting specific strategic policies. These developments, regarding to production, organization and distribution of information lead not only to new tools and ways that a librarian has the ability to use as information scientist, but mainly create a new and completely different perception and attitude on these issues. The purpose of this article is to raise and address a number of issues related to organization and manipulation of geospatial data in the library context and highlight the importance of GI and its proper management, while it will demonstrate the latest changes and developments in this area according to the international literature.

It is commonly accepted that GI the last few years, assisted by the technological developments as well as economic, social, and political conditions has gained particular importance to citizen everyday life. Governmental changes in many countries, immigration (political, economic, climatic) the environment and the global economy are issues that determine and affect the inhabitants of the planet and relate to GI. As the number, complexity, and diversity of geographic datasets grow, a method for providing an understanding of all aspects of this data grows in importance. To ensure that data is not misused, the assumptions and limitations affecting the creation of data must be fully documented [20]. Since interoperability is a central aim of any policy regarding the organization of information, librarians will always seek for the way that this main functionality will be achieved in order to secure the ability to communicate and exchange machine readable data, both on a data and system level. For instance, Library of Congress made it a requirement to include geographic coordinates in library catalog records despite the inability of many automated catalogs to search for coordinates [20].

The information obtained from this work could help librarians’ involving with GI to address the following questions: 1) Which are the latest news regarding the description of geographic information, and 2) What are the tools most commonly used for GI management.

2. Terminology

2.1 Geographic Information (GI)

GI must be seen as a special case of information as a whole [28]. According to Longley [26] GI is “the Information about objects or phenomena that are associated with a location relative to the surface of the Earth” while Goodchild [15] exceeds the term adding the “knowledge about where something is and what is at a given location”. As assumed by the above terminology, location is clearly essential in order to make information geographic.

GI is important to everyone’s life because (a) almost all human activities contain a geographic element and (b) the GI processing involves complex and difficult choices which are to a large extent, unique [26].

The economic significance of GI lies in the general referencing framework that it provides for integrating large numbers of different data sets from many application fields in both the public and the private sector [28]. The increasing commercialisation of GI supply and the free exchange of data among scientists working on global problems are two conflicting tendencies on the international scene nowadays [32]. At this point some clarifications should be made regarding the terms “geographic”, “spatial” and “geospatial”. The term “geographic” has to do with the earth its two-dimensional surface and its three dimensional atmosphere, oceans, and sub-surface. The term “spatial” has to do with any multi-dimensional frame and not only in the area at the surface of the earth (e.g. medical images are referenced to the human body, architectural drawings are referenced to a building). Geographic is a subset of “spatial” and often the terms are used interchangeably [15, 26] when in
the majority of scientific papers the term “geospatial” is used in order to assign the digital spatial information with a geographic component. Geospatial data can be “numbers, images, video or audio streams, software and software versioning information, algorithms, equations, animations, or models/simulations” [29] which have a spatial referent. They also include geographic coordinates (e.g., latitude and longitude) that identify a specific location on the Earth; and data that are linked to geographic locations or have a geospatial component (e.g., socio-economic data, land use records and analyses, land surveys, environmental analyses) [12]. A georeferenced digital library is an information system that stores georeferenced resources, and moreover provides a spatial orientation to those resources in terms of discovery, browsing, viewing, and access [22]. The main characteristics that bibliography [23, 13] reports for the geospatial data are: 1) they have no uniform data model, 2) vary widely in the amount of information they show, 3) these data are being stored in relational geodatabases requiring sophisticated storage and archiving schemes, 4) geospatial imagery datasets are often quite large and metadata may be voluminous as well, 5) may be produced over time, 6) are subject to versioning because of updated information being made available or to correct past errors in the data. Since [14] in his report for the Congress sets that “geographically referenced information is also known as GI” for the needs of the present paper be covered the term GI will be used.

2.2 Geographic Collections (GC)

A GC consists of materials such as books, serials, maps, atlases, aerial photos, remote sense images, geospatial data, software etc which deal with the study of the impact of people upon the earth. The GC is designed to support instruction, recreational needs, and research in the organization that library serves. Our research so far comes to add new data to the field of GC in academic libraries and expand previous researches [34, 35, 36, 37]. Thus, academic libraries worldwide are familiar with GC since: 1) collect geographical data in a variety of type, format and subject 2) sustain geographic collections and 3) develop GIS services for dealing with their users needs. More specifically, by drawing a profile of academic libraries with GC we can demonstrate that in their majority are established in developed countries and their parent institution sustain department/s whose curricula are based on the use of geospatial information and GIS e.g. Geography, Geology, Environmental Studies etc. These libraries are collecting a variety of different types of digital data such as aerial views, atlases, data series, remotely sensed images city foreign maps, topographic profiles etc. while the subject categories that library choose to offer to their users include both physical and human geography. The most useable formats are cd’s, dvd’s, and raster but vector, microforms, ArcInfo and cd roms are selected as well.

Librarians taking user needs into consideration provide GIS services such as user education, technological infrastructure and remote access. As international literature indicates, the above three factors plays a significant role in an integrated and efficient access to and use of GIS that library provide.

3. How Librarians Manage GI

Over the past ten years libraries have seen an increase in the amount and types of cartographic materials requested and used. This growing demand calls for skilled professionals equipped with specialized knowledge of maps, geographic information systems (GIS) and all other cartographic resources, whether in hardcopy or digital form, and the cataloguing of, or metadata creation for these same resources [4]. GIS Librarianship was routed as a new area of librarianship and numerous academic libraries (e.g. University of Texas Arlington, Princeton University, University of North Carolina at Chapel Hill and many others) that have developed GIS collections emerged to occupy GIS Librarians so to select, acquire, and manage geographical information and numeric data, assisting students and faculty in the use of appropriate GIS technology and supporting that way campus research and educational needs. Weimer and Reehling [39] define Geographic Information Librarianship “as the profession of providing geographic information resources and services in a library setting” while GIS Librarian
“is the librarian who knows the community and serves the clientele with appropriate maps as well as appropriate technologies”. So, the GIS Librarian is the main person who has the role to: acquire data, extract geographic information, extract semantic and ontological information and present in a form that allows easy exploration by users [38]. The way that librarians manage GI usually has two components: 1) by contributing in the development of interactive geo-services in their libraries, and 2) by using the proper descriptive standards. These two aspects will be discussed below.

3.1. Interactive Geoservices

Interactive Geo-Services in Libraries are information services for the public usually for a specific subject e.g. health, environment, archaeology etc. Some of them allow user to enter postcode, place name or address details to retrieve information and a map displays specific to their locations [28].

An example of interactive geoservice is the “Digimap” [11]. It is a collection of EDINA services that deliver maps and map data of Great Britain to UK tertiary education. Data is available either to download to use with appropriate application software such as GIS or CAD, or as maps generated by Digimap online. There are a number of collections available (Digimap-Ordnance Survey Collection, Historic Digimap, Geology Digimap and Marine Digimap). Harvard University Library developed “Harvard Geospatial Library” (HGL) a catalog and repository of data for Geographic Information Systems (GIS). HGL allows users to search the descriptive information of thousands of GIS layers using text as well as geographic coordinates. In addition, many of the layers are available for download, and are in a consistent, open format so they can easily be used by many different software packages without the need for translation. The benefits of HGL are: a) that makes GIS data collections more accessible and b) its search tools improve discovery of relevant GIS data.

3.2. Use of the proper descriptive standards

The library community has been very active in cataloguing various forms of information for many years (AACR, USMARC, UNIMARC, Z39.50, ISBD etc). Work relating to the cataloguing of spatial data in digital form began in the library community in 1990 [2]. The proliferation of GI the last decades of the private sector and the Federal Depository Library Program have pushed the creation and distribution of government information through electronic channels [21], a fact that should be consider as one of the reasons that in US Executive Order 12907, signed be President Clinton, requires agencies to base the geographic data documentation on federal metadata standards, post the data electronically and join in industry standard activities [1]. Geospatial metadata commonly document geographic digital data such as GIS files, geospatial databases, and earth imagery but can also be used to document geospatial resources including data catalogs, mapping applications, data models and related websites. Metadata records include core library catalog elements such as Title, Abstract, and Publication Data; geographic elements such as Geographic Extent and Projection Information; and database elements such as Attribute Label Definitions and Attribute Domain Values (http://www.fgdc.gov/metadata). Their role can be briefly described in the management of nested collections of resources, in the interoperability between metadata schemas, and in the integration of information retrieval techniques to the discovery services of geographic data catalogs (contributing in this way to avoid metadata content heterogeneity) [30]. It is apparent to above researchers that the sheer amount of geographical data that has been created without associated documentation makes it difficult to identify groups of related resources and in the same framework [3] illustrates that the value of geographic data depends, first on the awareness by potential users of the data’s existence and, second, on the data’s suitability for analysis and solving of spatial problems.
4. Geographical Tools and Concepts

In this section we will outline some of the tools that in our opinion play a significant role in establishing policies and procedures for developing a GC and that in the near future the above will be discussed extensively for the changes they will provoke in the metadata manipulation as literature indicates [2, 16]. Consequently the specific listing is not exhaustive.

4.1. GML

GML (Geography Markup Language) is an open industry standard that uses XML encoding for the transport and storage of GI including both the geometry and properties of geographic features, such as feature, geometry, coordinate reference system, time, dynamic feature, coverage (including geographic images), unit of measure, and map presentation style [16]. GML enables Internet based tools, such as Google maps to access GI such as merchant locations and traffic conditions. Implements concepts found within the ISO series (ISO 19136:2007) to support spatial and not spatial properties of objects. There is even a GML profile for RSS.

4.2 CSDGM – FGDC

Spatial data collection in USA is heavily dependent upon data issued by Federal agencies and that’s why the Geography and Map Division of Library of Congress began working with the FGDC, the agency tasked to generate a metadata standard [2]. The CSDGM – FGDC is similar to MARC and AACR2 and provides guidelines to develop geospatial profiles, allows the use of user-defined metadata entities and elements, creates short names/tags for all entities/elements, includes spatial reference, allows the use of free text and includes a glossary.

4.3. ANZLIC

The “Australia New Zealand Land Information Council” (ANZLIC) [5] is worthy of mention because of its uniqueness. ANZLIC’s strategic plan was to develop and implement a national land and geographic data directory system for Australia and New Zealand. The FGDC standard influenced the early stages of the development of the ANZLIC Guidelines.

4.4. Dublin Core

Dublin Core is well known to librarians and researchers related to digital libraries. Consists a simple general purpose metadata standard and nowadays, has become an important part of the emerging infrastructure of the Internet. Many communities are eager to adopt a common core of semantics for resource description, and the DC has attracted broad ranging international and interdisciplinary support for this purpose. Its simplicity (15 elements) is considered as an efficient choice for exposing CSDGM or ISO 19115 data to other communities through conversion [30, 16].

4.5. ISO 19115

There are approximately 28 standards specific to GI in the ISO 19100 family [24]. ISO 19115 standard is intended to facilitate the search and harvest of metadata records for all portal catalogues. Also allows for GI related groups to create profiles or guidelines for their respective communities e.g. INSPIRE, UK GEMINI 2 [28].
4.6 INSPIRE Directive

INSPIRE Directive has a direct effect on the EU –GI community as every member state of the EU must transpose this Directive into law. The remit of the INSPIRE Directive is to “establish an infrastructure for spatial information in Europe to support Community environmental policies, and policies or activities which may have an impact on the environment”. This in effect, is a mandate for the creation of metadata and data sharing across the EU public sector.

4.7. Gazetteers

Gazetteers are key components of georeferenced information systems, including applications such as Web-based mapping services. Digital gazetteers are directories containing triples of Place names (N), geographic footprints (F), and feature types (T) for named geographic places. Some gazetteers provide information about places and features. Some lists of geographic names are available as hierarchical term sets (thesauri) designed for information retrieval [19]. Some of the well known gazetteers are Thesaurus of Geographic Names from Getty Museum, Alexandria’s Digital Library Gazetteer, the GeoRef Thesaurus (American Geological Institute), Geographic Names Information System (GNIS), Geographic Names Processing System (GNPS). Researchers [25] that are involved with gazetteers moved on to point out a next generation gazetteer infrastructure (NGGI) through a focus on a distributed approach for the integration of volunteered GI as well as semantics-based retrieval and navigation.

4.8. RDA

AACR2r facilitates data sharing among disparate systems and frees the cataloger from having to reinvent the rules for each dataset. That’s what we used for the print material. Joint Steering Committee for Revision of AACR are working on a new standard RDA Resource Description and Access which has the ability to better describe and embed FGDC data elements and description into the bibliographic and authority records [16]. One of the changes is in how map elements are named, described and structured. As the RDA is finalized the impact of increased access and description will better serve the geographic and GIS communities.

4.9. FRBR

IFLA promotes the Functional Requirements for Bibliographic Records (FRBR) model, a conceptual one and develops new descriptive standards and standards for access points. More technically, FRBR uses an entity-relationship model of metadata for information objects, instead of the single flat record concept underlying current cataloging standards. The FRBR model includes four levels of representation: work, expression, manifestation, and item (http://www.oclc.org/research/activities/frbr.html).

4.10 Linked Data

Proposed 2006 by Berners-Lee: “data (its web representation) can contain links to data located elsewhere on the Web”. "Linked Data" (LD) refers to data published in accordance with principles designed to facilitate linkages among datasets, element sets, and value vocabularies. LD uses Uniform Resource Identifiers (URIs) as globally unique identifiers for any kind of resource. In LD, URIs may be Internationalized Resource Identifiers (IRIs) – Web addresses that use the extended set of natural-language scripts supported by Unicode. LD is expressed using standards such as Resource Description Framework (RDF) which specifies relationships between things -- relationships that can be used for navigating between, or integrating, information from multiple sources.
LData is sharable, extensible, and easily re-usable [31] and have been recognized as one of the first success stories of the emerging Semantic Web and the cloud they reside is growing with tremendous rate. For geospatial applications, the key linked open data hub is the GeoNames service, which provides persistent URIs and metadata about a variety of geographic locations [7].

All changes to new technologies and formats require changes in workflow processes and adding new tools to developer’s kits. On a positive note, these changes will allow users to create more consistent, robust metadata in multiple formats (txt, doc, html etc) so to better manage data [8, 16].

5. Conclusions-Discussion

Because of the importance of geospatial data, both public and private sectors have invested huge amounts of resources and money in collecting, managing, archiving, and distributing geospatial data [10]. GI is increasing rapidly and both public and private sector recognizes that an understanding of location and place is vital component of effective decision making. Technological evolution will continue to accelerate as already provokes significant challenges in today’s digital environment: Technological advances and software engineering, Reuse of bibliographic records by multiple agents in the www, Information retrieval from different heterogeneity sources, Proliferation of volunteered geographic information, Semantics-based retrieval and navigation, The time and cost factors for anonymising, tranforming and delivering data remains as a major obstacle for most data providers, Changes in data policies and revisions to standards also impose additional concerns on data developers, Issues regarding privacy and security [25, 28]. UN Committee highlights that humanity currently suffering from data overload since 2.5 quantrillion bytes of data created every day, a significance amount of which will have some kind of location reference. That challenges the data management moving forward so a number of primarily technology-driven trends in the professional world of geospatial data collection will continue to improve both the quality of data collected and the efficiency with which it is collected. In the above context of continuous changes we depict some key concepts (GI and GC) and aspects of a GIS Librarians every day work as well, (manage of GI) while furthermore we underline the constant need of being informed on technological evolutions that will support his mission: help the user to find the information he wants.

References