Search | Back Issues | Author Index | Title Index | Contents

———— **A R T I C L E S** ————

The Alexandria Digital Library Project

Review, Assessment, and Prospects

[Michael F. Goodchild](#)
Department of Geography
University of California, Santa Barbara
<good@geog.ucsb.edu>

Abstract

The Alexandria Digital Library (ADL) was established in the late 1990s as a response to several perceived problems of traditional map libraries, notably access and organization. By 1999 it had evolved into an operational digital library, offering a well-defined set of services to a broad user community, based on an extensive collection of georeferenced information objects. The vision of ADL continues to evolve, as technology makes new services possible, as its users become more sophisticated and demanding, and as the broader field of geographic information science (GIScience) identifies new avenues for research and application.

A map library on the Web

Maps are perhaps one of the earliest forms of stored, sharable human knowledge. They may have originated as drawings in the mud of cave floors, but by the 16th century had evolved into sophisticated, scaled representations of the surface of the Earth. By the late 20th century the task of mapping the Earth's surface was essentially complete, in the sense that no areas of continents remained blank, though mapping was much less detailed in inaccessible and less populated regions. National mapping agencies such as the U.S. Geological Survey (USGS) maintained and distributed vast numbers of maps: for example, the coverage of the lower 48 states at 1:24,000 scale consists of approximately 55,000 individual map sheets.

The map library evolved as an intermediate storehouse in this system of creation and dissemination of geographic information. On a university campus, for example, it is clearly much more efficient if the map user can consult or borrow a map from a campus map library, rather than having to have one shipped from the USGS warehouse in Denver. In that

sense map libraries are analogous to retail stores, acting as central facilities that provide a well-defined service to a local community (Berry, 1967; Goodchild, 2001). On the campus of the University of California, Santa Barbara the map library is unusually extensive, largely because of the strong focus of the campus on research in geography and related disciplines. By the early 1990s it had accumulated several hundred thousand map sheets, plus several million air photos, and a large collection of atlases, globes, and other forms of geographic information. The map library was thus an important and prestigious component of the UCSB library system.

While map libraries often function well within larger library systems, they are distinct in a number of important ways. First, they are more specialized, and not every library has or can afford a large map collection; where they exist, map collections are therefore highly valued. Second, maps, images, and globes present assorted problems of storage, due to their size, their cumbersome physical form, and issues of preservation.

Third, they are notoriously difficult to catalog in the traditional author/title/subject paradigm of classification (for an early article on automating map catalogs see Goodchild and Donkin, 1967). The most obvious basis for search and retrieval of maps and related objects is geographic coverage: a user is typically looking for a map of some*where*. But geographic space is continuous rather than discrete, and an assortment of methods are used for defining geographic location, including coordinates (latitude and longitude), placenames, and various indexing schemes. Information about the City of Goleta may appear on a map titled "Santa Barbara County", and will not be found in a title search for "Goleta" unless the nesting relationship between the two geographic entities has been coded or can somehow be inferred.

The Alexandria Digital Library Project began in 1994 as an attempt to address these three problems, building on earlier work on automating the UCSB map library catalog. If both the catalog and the content of the map library could be automated, then all three problems could be addressed: users would be able to use the library remotely, leveraging the library's investment by extending access to it globally; digital storage would resolve issues of preservation and the management of physical media; and an automated catalog would be capable of finding information by geographic location. A five-year grant was obtained from the National Science Foundation under its first Digital Library Initiative, and work started in earnest in late 1994. Off-the-shelf GIS (geographic information system) software was used to construct rapidly a proof-of-concept prototype, guided by concepts that had been explored in a previous geoinformation project sponsored by the Research Libraries Group (1989).

Implicit from the start of the project was the notion that research in digital libraries must occur simultaneously with implementation, because ideas evolve much more rapidly when thinking is combined with doing. Several new and powerful ideas evolved almost immediately as this process began, though they were not part of the original vision of the proposal to NSF.

First, the Web arrived and was popularized, providing a far more effective vehicle for access, search, and dissemination, so plans were rapidly switched to an HTTP-based approach.

Second, distributing the prototype to potential users led to a stream of useful feedback, and

guidance on the appropriate design of the user interface (Hill *et al*., 2000). It became quickly apparent, for example, that it was not enough to present users with a map and to expect them to identify an area of interest on it, because in many cases users simply lacked the necessary skills to work with maps. Scale or level of detail also turned out to be a difficult concept for many users.

Third, the team recognized that its concept of search based on geographic location could be generalized to apply not only to maps and images, but to any information objects that possessed geographic references or *footprints*, by being associated with points or areas on the surface of the Earth. This includes collections of georeferenced photographs (e.g., Microsoft Research, 2004), reports relating to specific areas, news stories about places (e.g., MetaCarta Inc., 2004), or even pieces of music (e.g., Mussorgsky's *Pictures at an Exhibition*, which contains musical references to Limoges and the Tuileries and Catacombs in Paris). The project coined the term *geolibrary*, for a library containing georeferenced objects and with a search mechanism based on geographic location as the primary search key. The National Research Council was persuaded to sponsor a workshop to flesh out the concept, and a book reporting on the workshop was subsequently published (NRC, 1999).

The concepts pioneered by ADL in its early years were later adopted in a large number of other projects, and many sites now have the characteristics of a geolibrary. The most prominent of these is, of course, the ability to search by geographic location as represented by latitude and longitude coordinates, a concept that is entirely new to the library community. Geographic space is continuous and multidimensional, and an infinite number of possible locations can be referenced, making it impossible to create a discrete, unidimensional search key analogous to author, title, or subject. Complex relationships such as containment, overlap, and adjacency exist between locations, and while they can be deduced from formal, coordinate specifications they are mostly invisible in informal, placename-based specifications. A geolibrary can only exist in a digital world, therefore, and it remains one of the most powerful concepts to have come out of digital library research. Three methods for specifying location are in common use in geolibraries: visually, by interacting with an onscreen map; formally, by specifying location in some appropriate coordinate system, typically latitude and longitude; and informally, by placename. In the last case, the library must then invoke a *gazetteer service* to convert the placename to coordinates, and will normally do so in a dialog with the user because of the need to resolve inherent ambiguities (multiple places with the same name, spelling variations, diacritical marks, variants in Romanized placenames from Arabia or China, etc.).

The current operational version of ADL offers a Web-based service to users worldwide. The catalog of the central UCSB collection currently contains over 2 million entries, and many of the UCSB catalog entries point to online objects that collectively amount to several terabytes. The catalog is based on a defined set of characteristics compatible with the U.S. Federal Geographic Data Committee's *Content Standard for Digital Geospatial Metadata* (FGDC, 1998) and with MARC. The gazetteer contains over 4 million entries compiled from Federal and other sources. ADL middleware has been constructed to be interoperable with other collections and catalogs, and the trend in recent years has been to move ADL away from a single, monolithic collection to a complex of smaller, more homogeneous, and distributed collections that can be accessed and searched transparently in a globally effective digital library. In turn, this decentralization has allowed more sites to participate in the

collection-building process.

ADL has added very significant value to the UCSB map library, particularly to its collection of historic air photos, and is used worldwide for a dazzling array of applications. Lawyers, for example, frequently use the collection to prove or disprove theories about the historic use of land. In a recent Southern California case, air photos were used to establish the amount of water used for irrigation in an area that had triggered a sudden landslide, causing substantial property damage. The photos were able to show that irrigation had been common practice for years before the landslide and that landslides had been common both before and after irrigation. ADL allows users to pay a virtual visit to the library without spending the time to travel, and to search its collection much more effectively than would have been possible in the pre-digital era. These user services now provide a substantial and welcome revenue stream to the library, and users benefit from efficient access to georeferenced materials previously available only by visiting the library in person, searching manually through the collection, and physically copying items of interest.

Updating the vision

ADL implemented a vision of the map library not as a physical resource to which users must travel to obtain geographic information, but as a virtual resource accessible and searchable from anywhere. While it vastly improved access to the library, and hence the library's value, it did little to escape the conceptualization of the library as a warehouse of information objects, and it perpetuated the notion that user needs can be satisfied by describing and providing access to such objects. Georeferenced objects range in size from Kbytes to Gbytes, but are essentially treated as packages with descriptive labels in this model—in no way does the library interact with the contents once they have been described.

One consequence of this approach is that no effort was made in ADL to register objects to an Earth coordinate system, beyond capturing the bounding coordinates to sufficient accuracy to enable search. Accurate *georegistration* is a requirement if datasets are to be overlaid automatically on other datasets, and this is one of the strengths of GIS. But users of ADL datasets whose applications require them to combine the ADL data with other data would have to engage in a potentially difficult and lengthy process of accurate registration before the data could be used. Lack of accurate georegistration is, of course, not a problem for someone whose application is limited to examining a simple display or printed version of the data.

In the late 1990s, a more advanced vision began to emerge of the role of a digital library. Instead of a simple warehouse supporting search, discovery, and package-handling functions, the digital library was seen as providing a much richer suite of services based on its holdings. In the context of ADL, these might include the ability to generate a user-defined window of data derived from a coverage that the user interprets as continuous, any internal segmentation into map sheets or tiles being hidden from the user's view. A user-defined window might cross the boundaries of these internal objects, but any technical problems created by such instances would be addressed internally and would not in any way impact the user. This clearly requires the ability to open and process the objects in the library before delivery. The vision of ADL could include the provision of a suite of GIS-like services (for a recent introduction to GIS see Longley *et al*., 2001), using ADL's holdings as

the base layers, but processing them to obtain user-defined products. The vision might also include the ability to accept, parse, and resolve plain-language queries based on ADL holdings, such as "Tell me the names of the cities on the Mississippi River".

In this advanced vision, a digital library takes on the general role of an information service. ADL's special niche would be to provide a service based on geographic information; that is, information linking locations on the Earth's surface to the properties present at those locations. The internal organization of the library's holdings would be largely irrelevant to the user. The vision raises a host of questions, from user-interface issues (could the interface be simple enough to empower an average library user, rather than an expert in GIS?) and technical issues (what internal structures and indexing schemes could support such an advanced design?), to institutional issues (how should intellectual property be addressed when products are obtained from original sources by complex processes?). In this new role, the library becomes a source of assurance of data quality, a brand that is automatically attached to all of the information that flows from the library. The library's acquisition policy must provide the basis of such assurance, and documentation of the lineage of outgoing information, of the processes by which such information is obtained, becomes critical.

To date, ADL's user-interface designs have assumed access through a standard browser, with a screen of typical desktop size. Increasingly, however, geographic information services are being delivered through much smaller devices, including PDAs (personal digital assistants) and cellphones. The growing field of *location-based services* is concerned with providing information through mobile devices that 1) know where they are, and 2) modify the information they provide accordingly. So, for example, a "yellow-page" service allows the user of a cellphone to generate a map of the businesses nearest to his or her current location, to display the map on the cellphone's screen, and to interact with it (e.g., by zooming to greater detail). This is enabled by systems that track the location of the cellphone, using GPS (the Global Positioning System) or by various kinds of triangulation from cellphone towers. Location-based services (LBS) include those provided by in-car navigation systems, and rely on wireless services to provide such systems with real-time information on congestion. For a recent review of LBS and related technologies see Peng and Zhou (2003).

Such systems constitute a form of *augmented reality* (AR) that contrasts sharply with the virtual reality offered by standard desktop environments, which visit distant places only through the medium of geographic datasets. In AR, information from the database is provided on-site through some kind of mobile device, and serves to augment the information directly available to the user through the senses and from memory. AR systems are particularly powerful when the senses fail—when a construction crew digs blindly in a street and hits an invisible, buried pipe whose location is known through a GIS database, or when a driver using an in-car navigation system finds a business located out of sight on a side street. The most advanced AR systems are wearable and *heads-up*, superimposing visual information directly on the user's field of view.

In this context, the library becomes a much more important player in the information society than is suggested by its traditional role of delivering information contained in pre-assembled packages such as books or map sheets. To a degree this has always been true—the library never had a monopoly on the provision of information to society. But it becomes a critical issue in any discussion of the library's future: what kinds of information should a library

provide in a digital world, and to what extent will the metaphor of the traditional library continue to dominate the world of digital libraries? These issues are clearly much broader than ADL, but projects like ADL have done much to draw attention to them, and to focus thinking on the evolving vision of the digital library.

Impediments and encouragements to progress

Michael Wegener once wrote that "Everything that happens, happens somewhere in space and time". The location of an event establishes its context and allows it to be linked to other events that have occurred at the same place. In that sense, it is important that an information source—a digital library—be able to answer questions of the type "what information do you have about *there*?" Construction workers need to know about pipes buried under a specific location; vacationers need to know about hotel options or to know what books in the library are about a given area of interest; researchers need to know about factors that might be causing a disease outbreak; and citizens need to know about proposed projects that might impact their neighborhood.

ADL and similar systems are able to answer these types of queries by serving digital representations of maps, images, reports, and other information objects that cover the location of interest, and whose other catalog properties (theme, date, etc.) suggest relevance to the query. But breaking open the objects, and selecting only the information directly relevant to the query, still lies beyond the capabilities of most current systems. More specialized services are able to address specific queries, but only by limiting them to well-defined domains, such as retail businesses or public transit services. The primary impediments appear to be institutional and cultural, as they often are in times of rapidly changing technology, since the prevailing conceptualization of the digital library is still dominated by the image of the traditional, physical library and its role in assessing, cataloging, storing, and distributing discrete, indivisible information objects. At the beginning of the 20th century, the automobile was described as the "horseless carriage", a way of conceptualizing a new technology by comparison to an earlier one. Similarly the term "digital library" connotes a continuing dominance of the traditional, physical library metaphor; its replacement with a new, distinctive term will only come when the metaphor of the old technology is no longer useful and when the vision has advanced to the point where the old technology is no longer relevant. At the same time, it is important that the new world not lose sight of the multitude of useful functions that libraries perform, in addition to the simple one of supporting search and distribution of information objects.

In this regard, it is significant that the digital world has adopted a model of description or *metadata* that is similarly fixed on the object, rather than on its component parts or on the properties of entire collections. In the traditional world of physical access to map libraries, there was little interest in formalizing the description of entire collections, since there was usually little practical alternative to the local library as a source of information. But in the digital world, where it is equally easy to visit any digital library, collection-level metadata (CLM) is an essential component of any technique for searching across collections to find the one most likely to contain a given object (Goodchild and Zhou, 2003). ADL has developed a CLM structure and implemented it as a major feature of its architecture.

Maps and images are by definition static once produced, and the information delivered

through maps tends to be concerned with relatively static properties of the Earth's surface, such as the locations of landforms, roads, and built structures. Maps are expensive to update, since they must be recompiled and reprinted, and the average topographic map sheet produced by the USGS is now over 25 years old. It is difficult to find much attention being paid to time in map libraries, and ADL has, of course, inherited this problem. Although it is possible to search ADL for information objects by date of production, in no way do those information objects inform the user about changes through time, or about dynamic processes operating on the Earth's surface. Nevertheless, the geolibrary concept is easily generalized to time ("what information do you have about *there then*?"), and digital representations of change through time could be stored and disseminated through ADL. Time is also a critical factor in gazetteer services, since placenames change through time, and historic placenames are sometimes lost.

Geolibraries stand to benefit from trends in society that are giving greater importance to place and to geography. The concept of *place-based* research is becoming popular in many disciplines: instead of searching for principles that apply anywhere at any time, this paradigm emphasizes the inherent heterogeneity of environmental and social processes, and the need to understand specific cultures and physical environments. The homogenizing effects of globalization are being countered by trends to localization, and the resolution of issues at neighborhood and community levels. Local information is essential in public administration when general policies and principles must be applied in local conditions. Marketing is increasingly local, with advertising adapted to the needs and profiles of local markets. In the early days of the popularization of the Internet, it was fashionable to declare the death of location—the global reach of electronic communication would make place irrelevant (Cairncross, 1997). The tide has definitely turned, however: "It was naïve to imagine that the global reach of the Internet would make geography irrelevant. Wireline and wireless technologies have bound the virtual and physical worlds closer than ever." (Economist, 2003).

Acknowledgments

References

Berry, B.J.L., 1967. *Geography of Market Centers and Retail Distribution*. Englewood Cliffs, NJ: Prentice Hall.

Cairncross, F., 1997. *The Death of Distance: How the Communications Revolution is Changing Our Live*s. Boston, MA: Harvard Business School Press.

Economist, 2003. The revenge of geography. *Economist* March 15.

Goodchild, M.F., 2001. Towards a location theory of distributed computing and e-commerce. In T.R. Leinbach and S.D. Brunn, editors, *Worlds of E-Commerce: Economic, Geographical and Social Dimensions*. New York: Wiley, pp. 67-86.

Goodchild, M.F. and K.M. Donkin, 1967. A computerized approach to increased map library utility. *The Cartographer* 4.

Goodchild, M.F. and J. Zhou, 2003. Finding geographic information: collection-level metadata. *GeoInformatica* 7(2): 95-112.

Hill, L.L., L. Carver, R. Dolin, J. Frew, M. Larsgaard, M.-A. Rae, and T.R. Smith, 2000. Alexandria Digital Library: user evaluation studies and system design. *Journal of the American Society for Information Science*, Special Issue on Digital Libraries 51(3): 246-259.

Longley, P.A., M.F. Goodchild, D.J. Maguire, and D.W. Rhind, 2001. *Geographic Information Systems and Science*. New York: Wiley.

Microsoft Research. (2004). *World Wide Media Exchange*. Available: <http://wwmx.org> [2004, April 5].

MetaCarta Inc. (2004). *MetaCarta Geographic Text Search Demo*. Available: <http://www.metacarta.com> [2004, April 5].

National Research Council, 1999. *Distributed Geolibraries: Spatial Information Resources*. Washington, DC: National Academy Press.

Peng, Z.-R. and M.-H. Zhou, 2003. *Internet GIS: Distributed Geographic Information Services for the Internet and Wireless Networks*. Hoboken, NJ: Wiley.

Research Libraries Group, 1989. RLG enters new sphere with geoinformation project. *The Research Libraries Group News* 19 (Spring): 3-9.

U.S. Federal Geographic Data Committee. (1998). Content Standard for Digital Geospatial Metadata (Version 2). Available: <http://fgdc.er.usgs.gov/metadata/contstan.html> [2004, April 5].