

EXPLORATORY ACCESS TO DIGITAL GEOGRAPHIC LIBRARIES[†]

Vincent F. Schenkelaars

TNO Physics and Electronics Lab

Oude Waalsdorperweg 63, NL-2597 AK The Hague, The Netherlands

Schenkelaars@fel.tno.nl

Max J. Egenhofer

National Center for Geographic Information and Analysis
and

Department of Spatial Information Science and Engineering

Department of Computer Science

5711 Boardman Hall, University of Maine, Orono, ME 04469-5711, USA

max@spatial.maine.edu

ABSTRACT

Users of digital geographic libraries face the challenge of discovering the dataset they are interested in. When locating and selecting spatial datasets, different scenarios may occur with varying levels of knowledge about the datasets desired. For example, some users may have detailed information about what they want ("I need the most recent Spot image covering Orono, Maine"), while others may be more vague in their descriptions ("To test my hydrological runoff model, I need a dataset that includes a terrain with steep slopes as well as some flat areas with sandy clay"). Usually, the collection of datasets available is by far too large to be examined one-by-one. The sheer size of geographic libraries poses a performance problem for the digital library – how to retrieve enough data within a short time so that users can make decisions – as well as a cognitive overload for the users – how to select from among all datasets available, those datasets that are worth a more detailed examination. We propose an interactive geographic browser with which users can explore a geographic library by examining query results. The browser, based on the magnifying glass metaphor, allows users to move a filter over datasets displayed against a background map, while on the fly changing non-spatial parameters that determine what datasets will be visible in the magnifying glass.

[†] This research was partially supported by Intergraph Corporation and TNO-FEL under AIO-contract number 93/399/FEL. Max Egenhofer's work is further funded by the National Science Foundation under grant numbers SBR-8810917, IRI-9309230, and IRI-9411330; and grants from Space Imaging Inc., Environmental Systems Research Institute, and the Scientific Division of the North Atlantic Treaty Organization.

INTRODUCTION

The number of digital geographic datasets available either through the Internet or on CD-ROMs is increasing dramatically. The critical aspect is not anymore to find smart ways to process data efficiently, but to determine which data are relevant and find what datasets are available for the task. This is not finding the famous needle in the haystack, but finding a straw with some desired properties. Similar to the process of checking sequentially one straw after another, it would be impractical to ask users to retrieve the geographic datasets one after another until a dataset matches their expectations. Such access would be tedious and expensive – both in terms of time and in money, particularly if charging for access was based on the number of datasets retrieved. To locate the “right” spatial dataset for a project, a digital spatial library has to offer some *access tools* tailored towards the working habits of its users.

Querying a digital geographic library for the purpose of retrieving a dataset is not much different from querying a geographic information system; therefore, querying in digital spatial libraries is well supported by spatial query languages, where SQL dialects are the most prominent ones (Egenhofer 1992). If one assumes that every user of a digital geographic library will have detailed knowledge about the particular dataset he or she is interested in, access becomes a mere query that results in the desired dataset. No further searching is necessary. The important aspect with this access method is the size of the result – a small enough answer, which can be presented to the user so that he or she can understand it and exploit it without further questions. Such a scenario may fit the behavior of users retrieving a particular dataset they had accessed before and whose key characteristics they remembered, such as the type of the dataset, its geographic location, and its recording time. More frequently, however, is the scenario when users lack such specific and precise knowledge about the datasets they want to find and their query will result in a fairly large subset of the entire library. The subsequent interaction with the query result is the challenging effort, because now the user has to choose those datasets that are more promising to his or her task than the others. This is the moment when users *explore* what is available in the library, and feedback they get from the datasets may trigger new demands, or make them decide to drop a particular line of thought in their search.

We distinguish three types of interactions with a geographic library:

- *spatial querying* (Egenhofer and Herring 1993) allows users to get answers to particular questions provided they have enough knowledge about the target objects. In addition, users need some knowledge about the way the data are structured in the database.
- *Spatial browsing* (Clementini *et al.* 1990), on the other hand, is going through the answers of a spatial query and finding the interesting items. This can be considered as a human executed query process.

- *Exploratory spatial access* (Egenhofer and Richards 1993) can be considered as a special type of browsing, in which users do not know in advance what they are looking for. Users go out into the unknown and come back with whatever they think is interesting or appropriate for their tasks. Exploratory access is far more interactive than browsing through a query result. It is this demand for interactivity that creates the need of a special tool for exploratory browsing.

When concerned with finding datasets, exploration is most often the users' choice. A variety of implementations can be envisioned for exploratory access to spatial datasets.

- One could present the users with a (prioritized) list of resulting datasets together with some of their key properties. By scrolling through this list, users examine the characteristics and if close to their intent, they download the dataset. While such a prioritized list may lead to a good hit list if non-spatial properties are irrelevant, it lacks a connection of the datasets' spatial properties.
- To provide users some idea of the content of the datasets, small representative subsets may be used to assess spatial properties such as dispersion, density, and pattern (Flewelling 1997). Still users need supporting tools to evaluate subsets.
- Another method that exploits spatial properties of the datasets such as location and extent is the display of the datasets' outlines over the background of a reference map or an aerial photograph. Users can make spatial choices by directly examining extent and location.

This paper focuses on the latter scenario and describes a browsing method that is based on displaying spatially the datasets and allows users to filter interactively those datasets that match some non-spatial criteria. It stresses the user interface aspects of such a method and its primary concern is how to deal with the visual clutter that may result from the many overlapping and nested geographic datasets. Investigations of such a browsing tool are related to the Alexandria project, which focuses on the development of a Digital Spatial Library (Smith 1996).

The remainder of this paper is structured as follows: The next section describes an example scenario of a users task. Section 3 describes EAGLE's user interface and desired functionality. Section 4 deals with architectural aspects of the intended system. Section 5 shows an example user session of the system. Finally, section 6 discusses future research and implementation aspects.

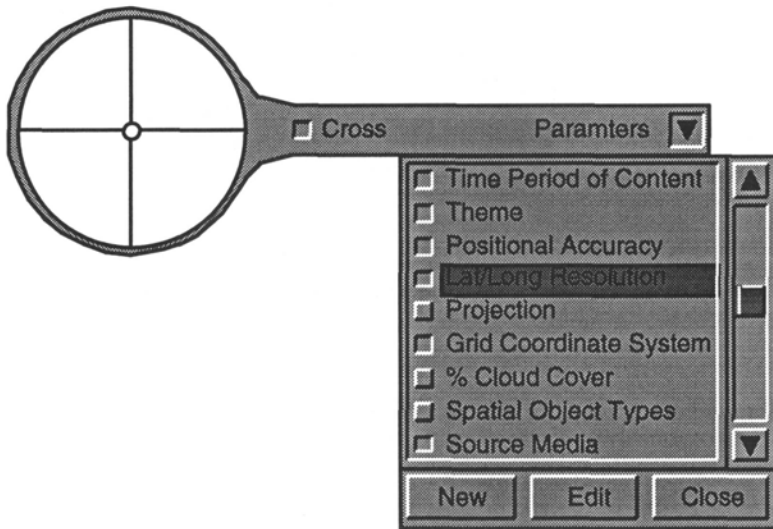


Figure 1: The magnifying glass with its list of parameters.

EXAMPLE SCENARIO

Consider a user who wants to find a dataset that can be used in some scientific project. The user has some knowledge about what the data must contain. First, this user tries to define a query and process it on a database containing the metadata information of the datasets. After execution of the query, the user is likely to end up with a very large number of datasets that match the initial requirements. After examining the metadata of a small number of the resulting datasets, the user realizes that there is another constraint that can be put into the query. So the query process has to be started again. This iterative process of adjusting a query and processing the query again is often inefficient. We argue that when the user has a tool that gives him or her direct and dynamic control over what datasets will be selected, the task of getting an appropriate dataset can be accomplished more efficiently. For this purpose, we use the magnifying glass metaphor and extend it with having filter parameters inside it. This is a combination of the movable filter (Stone *et al.* 1994) and dynamic query filtering (Ahlberg and Shneiderman 1994) concepts.

THE MAGNIFYING GLASS

The magnifying glass metaphor is the central object in the user interface. One of the important properties of the magnifying glass is that it restricts the operation area. This will dramatically improve the efficiency of the selection process. With the magnifying glass, the local operation area is moved over a

larger “overview” map, such that the user can explore a smaller area with more detail.

For the exploratory access to a geographic library, we also need a kind of filtering mechanism that will show only the qualifying target objects. This filtering mechanism needs to be very flexible. It should be possible to change the filtering parameters in an instant. Those filtering parameters are conceptually connected to the magnifying glass and we therefore connect them physically with the magnifying glass (Figure 1). The magnifying glass contains a button to switch a crosshair on and off. The crosshair may be used to focus the magnifying glass on a specific object. When the user moves the crosshair over an object on the map, the parameter values of that object are shown in a separate *attribute window*, which reflects the magnifying glass parameters, so the user can compare the actual object attribute values with these parameters. This window remains visible until the user decides to close it. This way, a user can compare the parameter values of a number of objects.

All the parameters in the magnifying glass parameter list are combined with a conjunction (logical and). There is little use for an or combination of parameters, because in that case the user is better off with an SQL-like query language. The magnifying glass is meant to be an exploratory browsing tool, so it does not need all the expression power of a spatial query language.

There is also a button on the handle of the magnifying glass which, when pressed, will show a list with possible filtering parameters. Each parameter has a check box with which it can be turned on and off. A turned-off parameter will not be considered in the filtering process. Attached to the list are three buttons: **New** to add a new parameter; **Edit** to change the values of the search parameters; and **Close** to close the parameter list. The **New** button activates a parameter creation dialog. In this dialog, the user can enter a name for the parameter, a type of parameter, and, depending on the type of parameter, a value range.

We distinguish four of different parameter types: Tri-State, Threshold, Interval, and Enumeration. These parameter types are discussed in the next subsections.

Tri-State set

This type of parameter is used to group a number of related parameters (Figure 2). A geographic dataset contains a number of spatial object types. A user might not be interested in datasets containing, for example, uniform B-splines, or the user might only be interested in datasets that contain at least point, line, and area features. Another possibility is that a user does not care if a dataset contains a certain object type. Initially, all the elements of the parameter set are in the *don't-care* state. The user moves the element to one of the other states by selecting it and dragging it via direct manipulation to the desired column.

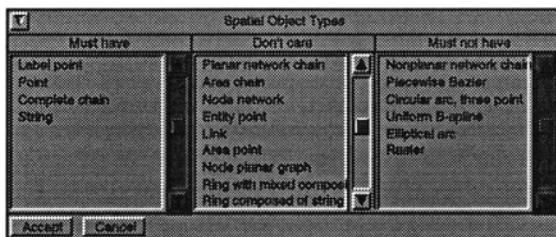


Figure 2: The Tri-State Set to select properties of objects.

Threshold

This type of parameter defines a threshold value. The target objects must have a parameter value that is larger or smaller than the threshold value. The dialog box in figure 3 shows an example of this type of parameter. The units, and ranges (maximum and minimum values) of the slider are defined when the search parameter is created.

Interval

This type of parameter defines the interval in which the parameter value is allowed. An example is a date value. A user can require that, for example the time period described by the geographic data lies between January 1 1990 and December 1 1993. The begin and end value of an interval parameter are set by moving two sliders. It is also possible to put some constraint on the sliders, e.g., end date is later than begin date (Sleezer 1994). Again, the units and ranges (maximum and minimum values) of the sliders are defined when the search parameter is created.

Enumeration Subset

This parameter type allows users to select a set of discrete values from a small set of choices. The user can select which subset of the initial set is allowed. The dialog box in Figure 4 shows an example of the data media on which the geographic dataset can be purchased. In this example the user can only use the Internet, a CD-ROM, a floppy disk, or a 9-track tape. Only geographic datasets that are available in one of those media are selected in the magnifying glass.

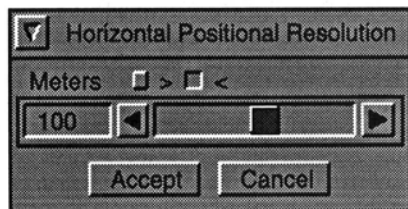


Figure 3: Dialog box to select a threshold

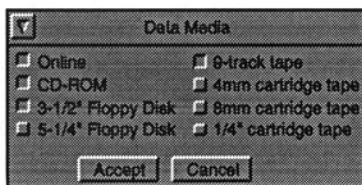


Figure 4: A dialog box to select from an Enumeration

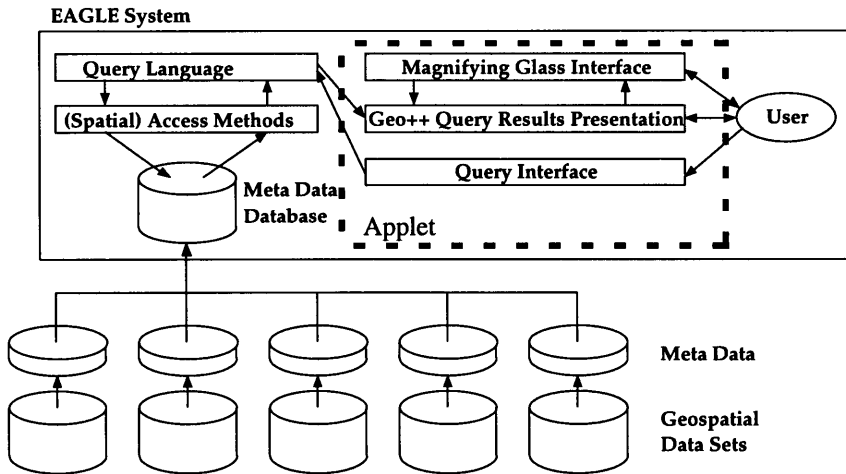


Figure 5: EAGL Architecture

ARCHITECTURE

This section describes an architecture of an environment that would support the browsing through datasets. Figure 5 shows an overview of the architecture of the EAGL (Exploratory Access to Geographic Libraries) System. The architecture is somewhat similar to the architecture of the Alexandria Project (Smith 1996).

The geographic datasets and their metadata information may be stored on sites all over the world. Since the EAGL system is intended to be a highly interactive tool, it is not possible to use the information stored on these remote sites, even if the required data are on-line. The time to collect the metadata of all the geographic datasets and transfer it to user's site would be too long. Therefore, the metadata has to be collected in advance in a metadata database. Not only is all the metadata available at the fastest speed, but it also allows to install some auxiliary access methods on top of the data. Again, EAGL is designed as an interactive system and needs all the speed it can get. The database must allow for efficient storage of spatial and non-spatial data. Furthermore, it must be able to define indices on both types of attributes. In our implementation we are using Illustra (Illustra 1995) as our metadata database. Illustra is an extendible object relational database on which TNO-FEL has developed a GIS 2D/3D extension similar to the extensions build on top of Postgres (Oosterom 1991).

On top of the database we have developed a front-end application in Java (SUN 96). This applet implements a number of concepts that can be found in GEO++ (Vijlbrief 1993). The applet contains a Main Map View, a Overview Map, buttons for starting the "Layer Manager" and a "Tool Box" in which the magnifying glass can be activated. The applet is being further extended in the

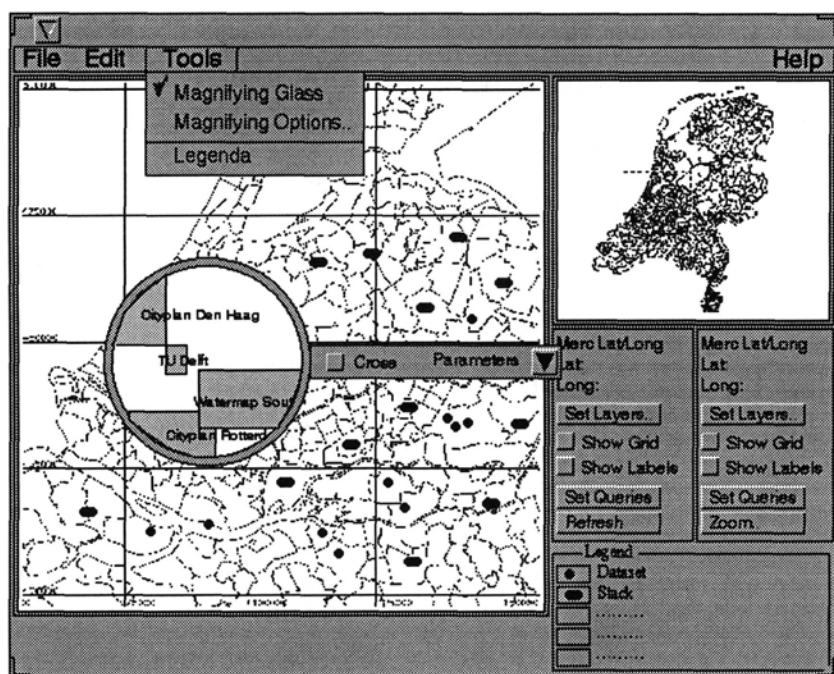


Figure 6: Browsing EAGL with the magnifying glass.

GeoMed project. A project description and design documents can be found at the GeoMed site (GeoMed 1996). Figure 6 shows a browsing session with EAGL.

EXAMPLE SESSION

This section describes a user session with EAGL. The user wants to find a geographic dataset describing a wetlands area in the United States. After starting EAGL, a coarse background world map is shown. Since the user has some constraints on the spatial location of the geographic dataset, zooming in to that area is the first action. Zooming in with EAGL is simply drawing a smaller rectangle on the map. After the right area is on the map window, the user can create a first, probably rough, query. This can be done with the *Set Query* button, which activates a *Query Composer*. The results of this query are presented on top of the background map. Each object in the geographic library represents a dataset covering a certain area. This area is used to display the object on the map. If more than two objects occupy the same area, both objects are replaced by a *stack-symbol*.

The result of the initial query has probably returned a huge amount of objects. The user can now activate the magnifying glass and *browse* through the

result. The user can either decide to exploit the parameters values of the magnifying glass or the user can activate the crosshair and examine the individual datasets.

In this example session, the user decides to use the crosshair and gets information about the digital data format of the datasets. Since the user's Geographic Information System cannot deal with all data formats, the corresponding filter parameter in the magnifying glass is switch on. When the user moves the magnifying glass over the map, with the crosshair switched off, only the objects that qualify the parameter values will be shown. Again, the user gets information about the datasets and is able to set new filter parameters. This step of browsing, interpreting, refining, and again browsing, eventually leads to a very small number of datasets. With the crosshair switched back on, the user can compare the attribute values of each individual object with the magnifying glass parameters and finally decide which datasets to purchase.

CONCLUSIONS

We have presented a tool for exploratory access to geographic libraries. We used the magnifying glass metaphor and combined it with the dynamic filter. So far, there is only a system design and a prototype implementation is under way. A number of open problems will have to answered by the release of this first prototype.

The first problem is what are useful filter parameters. For the design phase we have taken the FGDC metadata standard (FGDC 1997) as a source for filter parameters. In a first selection, we ended up with about 25 possible filter parameters. It is likely that they are not all as useful in the exploring process.

EAGL is designed as a highly interactive system. Some measures have to be taken to reach the required speed. For instance, implementing an advanced search mechanism is necessary (Oosterom 1995).

When a lot of datasets describe the same area, a stack symbol is presented on the map. A problem is how to access each individual element of the stack. One solution might be to move the magnifying glass over the stack and allow the user to flip through it. However, when the stack is very large, this might not be a good solution.

An exploratory access system can become very useful to the user community when public access is allowed. Connecting EAGL to the Internet by means of a World Wide Web (WWW) server seems to be a good idea. A Java implementation would make this possible.

REFERENCES

- C. Ahlberg and B. Shneiderman, Visual Information Seeking: Tight Coupling of Dynamic Query Filters with Starfield Displays, *CHI 94*, 1994, Boston, MA, 313-317.
- E. Clementini, A. D'Atri, and P. Di Felice, Browsing in Geographic Databases: An Object-Oriented Approach, *Workshop on Visual Languages*, Skokie, IL, 1990, 125-131.
- M. Egenhofer, Why not SQL! *International Journal of Geographical Information Systems* 6(2): 71-85, 1992.
- M. Egenhofer and J. Herring, Querying a Geographical Information System, *Human Factors in Geographical Information Systems*, D. Medyckyj and H. Hearnshaw (ed.) Belhaven Press, London, 124-136, 1993.
- M. Egenhofer and J. Richards, Exploratory Access to Geographic Data Based on the Map-Overlay Metaphor, *Journal of Visual Languages and Computing* 4(2): 105-125, 1993.
- FGDC, Metadata Standards Development, <http://www.fgdc.gov/Metadata/metahome.html>, 1997.
- D. Flewelling, Comparing Subsets for Digital Spatial Libraries, *Ph.D. Thesis, Department of Spatial Information Science and Engineering, University of Maine*, 1997.
- The GeoMed project, <http://illusion.fel.tno.nl/geomed/geomed.html>, 1996.
- P. van Oosterom and V. Schenkelaars, The development of an interactive multi-scale GIS, *International Journal of Geographic Information Systems*, 9(5): 489-507, 1995
- P. van Oosterom and T. Vijlbrief, Building a GIS on top of the open DBMS Postgres, *Second European Conference on Geographical Information Systems*, 775-787, 1991.
- A. Sleezer, Direct Manipulation of Temporally Constrained Activities from Geographic Modeling, *Master's Thesis, Department of Surveying Engineering, University of Maine*, 1994.
- T. Smith, A Digital Library for Geographically Referenced Materials, *IEEE Computer* 29(5): 54-60, 1996.
- M. Stone, K. Fishkin, and E. Bier, The Movable Filter as a User Interface Tool, *CHI 94*, 1994, Boston, MA, pp. 306-312.
- T. Vijlbrief and P. van Oosterom, The GEO++ system: An extensible GIS, *Spatial Data Handling*, Charleston, 1992.