Geographical Representation of Library Collections in WorldCat: A Prototype

Lynn Silipigni Connaway*
Consulting Research Scientist III
Research
OCLC Online Library Computer
  Center Inc.
6565 Frantz Road
Dublin, OH 43017
Email: connawal@oclc.org

Clifton Snyder
Software Engineer
Research
OCLC Online Library Computer
  Center Inc.
6565 Frantz Road
Dublin, OH 43017
Email: snyderc@oclc.org

Lawrence Olszewski
Director
OCLC Library
OCLC Online Library Computer
  Center Inc.
6565 Frantz Road
Dublin, OH 43017
Email: olszewsl@oclc.org

*All correspondence should be directed to Lynn Silipigni Connaway.

Abstract
In today's world, people can be inundated by an overwhelming amount of information. Library and information science professionals attempt to provide information systems that are capable of retrieving precise and accurate information. One method for the organization and retrieval of geographically-based information is to develop a system to visually represent the data. A prototype for an interactive map, the OCLC WorldMap, was developed to provide a visual tool for the management and representation of geographically-based library collections and library data. These data are used to provide information for decision-making in regard to remote storage, collection management and cooperative collection development, preservation, and digitization. The collection data for the map were generated from WorldCat, the OCLC Online Computer Library Center bibliographic database. WorldCat contains approximately 55 million records, and not only serves as an aggregator of bibliographic data, but also identifies a billion holding locations by type of library for library resources. Additional data, gathered from more than thirty other sources, such as library expenditures and number of libraries, are represented on the OCLC WorldMap.

Introduction

The development of the Internet has provided a means for the creation and distribution of more data than ever has been available before. One of the principles of library and information science is to provide systems for the organization and retrieval of information and to provide users assistance with the evaluation of this information. One way to organize large datasets is the creation of visual representations of data. Visual displays are often more appealing and more easily understood than tables of numbers. Another method is to pare the data, carving off large chunks of extraneous facts and figures in an attempt to get to the important information. This project combines these two concepts into a single visual interface.

The goal of the OCLC WorldMap is to create a visual tool for the management and representation of geographically-based library collections and data. Most of the data were mined from WorldCat, the OCLC bibliographic database, which contains more than 55 million records. Additional data were gathered from the Association of Research Libraries (ARL) (http://www.arl.org) and the National Center for Educational Statistics (NCES) (http://nces.ed.gov). The analysis of these statistics can provide useful information for remote storage, collection management and cooperative collection development, preservation, and digitization. Previously, the data existed in spreadsheets with thousands of rows and columns, which made it difficult to review and analyze. The interactive OCLC WorldMap was created to geographically represent these data.

Literature Review

Discussion of geographic information systems (GIS) dates back to the early 1990s; the American Society for Information Science devoted a special issue to spatial information, edited by Gluck (1994), in an attempt to bring this topic to the attention of library and information scientists. Lamont (1997) discusses the management issues involved in collecting, describing, and accessing spatial data. Fraser and Gluck (1999) discuss how users determine the relevance or potential value of geospatial objects from metadata in an update to Gluck's work with the OCLC Office of Research (Gluck, 1997). Gluck and Yu (2000) provide an introduction to the topic of GIS uses in libraries in their discussion of standards for geospatial data, description of several library implementations of GIS, and analysis of the role of GIS as a library management tool.

Koontz (1996) addresses how GIS software facilitates library market analysis. Ottensmann (1997) discusses how geographical information systems can be employed to analyze patterns of library utilization in public library systems. Jue et al. (1999) analyze the distribution of poverty areas relative to public library outlets in order to assess the best funding and development policies for those residents. Koontz and Jue (2004, 2004a) focus on how public librarians can use data to help them fulfill their roles as agencies in equitable information provision. They also have made available a Public Library Geographic Database (PLGDB) that geographically represents United States public library census data (http://www.geolib.org/PLGDB.cfm).

Tufte's pioneering study (2001) explains how to communicate information through the simultaneous presentation of words, numbers and pictures. According to Shneiderman and Plaisant (2004) there are five factors for benchmarking the usability of an interface. Two factors, *time to learn* and *speed of performance*, are measured by carefully timing a set of tasks provided to the tester. Users are carefully watched to determine their *rate of errors*, and *retention over time* is judged by the tester's ability to complete similar tasks throughout the course of testing. Finally, the user's *subjective satisfaction* is gauged by both the spoken comments during testing as well as a follow-up interview and questionnaire.

OCLC WorldMap – Development and Specifications

The OCLC WorldMap does not attempt photographic accuracy; rather, it is a relatively simple means of displaying geographic data. The system allows the user to select a dataset of interest from several options provided on the map. The user is able to display library collections, a group of libraries' collections, and all holdings in WorldCat by country of publication and date, or by library data. See Figure 1. The results are displayed on the map by variations of gradation to represent the data for the selected geographic regions (see Figure 2) or in a data table (see Figure 3), which the user is able to sort by selected column headers.

Many different technologies are available for creating tools of this kind. In an attempt to create an entirely open source/open standards prototype, the first version of the map was created using Scalable Vector Graphics (SVG), an open standard maintained by the W3C. SVG supports the rendering of basic shape elements, i.e., circle, rectangle, line, ellipse, polyline, polygon, a text element, and complex paths. It supports styling for these elements and cascading style sheets (CSS), allowing the developer to manipulate various painting attributes as well as clipping, masking, and filtering. The specification also provides the means for describing various transformations and animations of these elements that can be automatically started or triggered by user

interaction, and allows for custom scripting using ECMAScript, a scripting language based on JavaScript. Because it is scalable, any amount of zooming in or out of an SVG document can be done without pixilation or distortion. SVG is a very young technology browser; therefore, browser support is currently limited. The most popular of the SVG plugins, produced by Adobe, only is supported by Windows with Internet Explorer or Netscape Navigator. The Mozilla Foundation is currently working on an SVG project, with the eventual goal of providing native support for their Mozilla and Mozilla Firefox browsers. At this time most browsers support Macromedia Flash. In order to make the map available to a larger audience and to provide features not supported by SVG, Flash was used to implement a new version of the OCLC WorldMap.

Usability Testing

Usability testing for the OCLC WorldMap was done internally in the OCLC Usability Lab using OCLC employees. The tests provided a diverse set of users from a range of different educational backgrounds, disciplines, ages, and levels of technical experience. Informal requests for revisions also were provided by staff within the Office of Research and staff from other OCLC divisions.

The initial testing of the OCLC WorldMap returned a conclusive verdict: that the map was interesting, but the interface design was cumbersome and difficult to use. All of the users had difficulty locating small, obscure countries, which was attributed to a poorly implemented search feature. This problem was solved by providing the user with a textual list of country names, organized hierarchically by continent. The prototype map used only primary colors. According to Tufte (2001, pp. 153-54), a color-coded topical map should not "overload" the use of coloration as a means of communicating information to the user; it tends to be overwhelming and negatively impacts the user's understanding of the data. For example, the map that was tested used colors to represent the dataset selected, whether or not the user was hovering over a given country or had selected (clicked on) a given country. Lighter tones were used to replace the primary colors and user interaction is now indicated by changing the color of the outline of a country.

The users also noted the lack of sufficient "Help" functionality and confusion about how to use drop-down boxes effectively to access information; they found the labeling of the boxes misleading and uninformative. Despite these trouble spots, the testers saw the potential of the map and offered many helpful suggestions for improving its usability.

Conclusion and Future Development

The OCLC WorldMap is a prototype, with many planned additions and revisions. The development of a system that enables the addition and revision of datasets with greater ease will facilitate future updates. Although it is not practical for a system to represent every possible dataset, the map in its current form is much too static. The map is being revised to use varying shades of a single color to represent the data so that variability within the datasets will become more distinct and understandable to the user. Other methods for displaying geographic data also are being developed, such as cartograms. One interesting example represents countries as circles as opposed to shapes defined by political boundaries. These circles vary in size depending upon the number of materials represented in WorldCat by country of publication. This tool makes it possible to show a user at a glance where the greatest number of WorldCat records is published. Other visual tools for information display are being explored to represent data in non-traditional ways, which will have the potential to make datasets more accessible to the user.

The data represented in the OCLC WorldMap can be utilized by several different user groups. Internally, marketing and sales staff can use the library data to target potential areas of growth by segment in global expansion and to assess current market penetration. Externally, the place, date and language of publication data can be used by collection development staff to identify strong and weak area studies collections and to determine collection overlap between and last copies held by individual and groups of libraries. Librarians can use the datasets to plan for the integration of paper and digital collaborative collections, suggest candidates for deaccessioning and remote storage, and identify areas for preservation and digitization.

REFERENCES

Association of Research Libraries. Retrieved June 29, 2005 from http://www.arl.org.

Fraser, B. & Gluck, M. (1999). Usability of geospatial metadata or space-time matters. *American Society for Information Science Bulletin, 25*, 24-26.

Gluck, M. (Ed.). (1994). Spatial information [Special issue]. *Journal of the American Society for Information Science, 45* (9).

Gluck, M. (1997). A descriptive study of the usability of geospatial metadata. *Annual Review of OCLC Research.* Retrieved June 29, 2005 from http://digitalarchive.oclc.org/da/ViewObject.jsp;jsessionid=36543e159640472aad5d9c6dc5e3ccc4?fileid=0000002652:000000058927&reqid=22096.

Gluck, M. & Yu, L. (2000). Geographic Information Systems: background, frameworks, and uses in libraries. In F. C. Lynden & E. A. Chapman (Eds.). *Advances in Librarianship*, *23*, 1-38.

Jue, D. K. et al. (1999). Using public libraries to provide technology access for individuals in poverty: a nationwide analysis of library market areas using a Geographic Information System. *Library & Information Science Research, 21,* 299-325.

Koontz, C. (1996). Using Geographic Information Systems for estimating and profiling geographic library market areas. In L. C. Smith, L.C. & M. Gluck, M. (Eds.), *Geographic Information Systems and libraries: patrons, maps, and spatial information* (pp. 181-193). Clinic on Library Applications of Data Processing. Champaign: Graduate School of Library and Information Science.

Koontz, C. & Jue, D. K. (2004). Customer data 24/7 aids library planning and decision making. *Florida Libraries, 47*, 17-19.

Koontz, C. & Jue, D. (2004a). Unlock your demographics. *Library Journal, 129*(4)*,* 32-33.

Lamont, M. (1997). Managing geospatial data and services. *Journal of Academic Librarianship, 23,* 469-73.

National Center for Educational Statistics. Retrieved June 29, 2005 from http://nces.ed.gov.

Ottensmann, J.R. (1997). Using geographic information systems to analyze library utilization. *Library Quarterly, 67*, 373-95.

Public Library Geographic Database. Retrieved June 29, 2005 from http://www.geolib.org/PLGDB.cfm.

Shneiderman, B. & Plaisant, C. (2004). *Designing the user interface.* (4th ed.). Boston, MA: Pearson/Addison-Wesley.

Tufte, E. R. (2001). *The visual display of quantitative information.* (2nd ed.). Cheshire, CT: Graphics Press.

UNESCO Institute for Statistics. Retrieved June 29, 2005 from http://www.uis.unesco.org.

Appendix A
Tables

Figure 1. OCLC WorldMap search screen

Figure 2. Results displayed by variations of gradation.

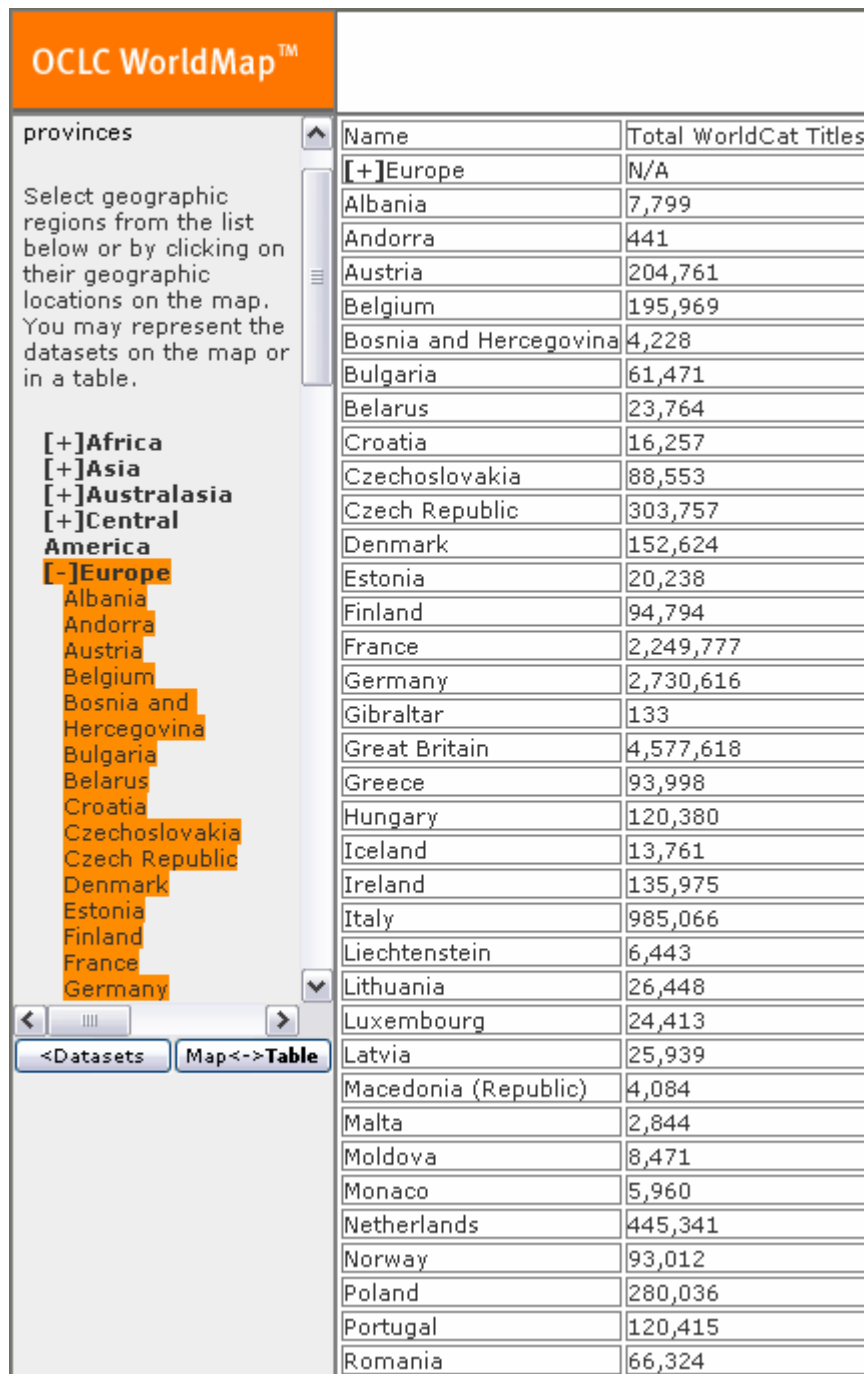| Name	Total WorldCat Titles
[+]Europe	N/A
Albania	7,799
Andorra	441
Austria	204,761
Belgium	195,969
Bosnia and Hercegovina	4,228
Bulgaria	61,471
Belarus	23,764
Croatia	16,257
Czechoslovakia	88,553
Czech Republic	303,757
Denmark	152,624
Estonia	20,238
Finland	94,794
France	2,249,777
Germany	2,730,616
Gibraltar	133
Great Britain	4,577,618
Greece	93,998
Hungary	120,380
Iceland	13,761
Ireland	135,975
Italy	985,066
Liechtenstein	6,443
Lithuania	26,448
Luxembourg	24,413
Latvia	25,939
Macedonia (Republic)	4,084
Malta	2,844
Moldova	8,471
Monaco	5,960
Netherlands	445,341
Norway	93,012
Poland	280,036
Portugal	120,415
Romania	66,324

Figure 3. Results displayed in a data table.