Olha Buchel  
Faculty of Information and Media Studies,  
University of Western Ontario, London, ON, Canada

Uncovering Hidden Clues about Geographic Visualization in LCC

Abstract: Geospatial information technologies revolutionize the way we have traditionally approached navigation and browsing in information systems. Colorful graphics, statistical summaries, geospatial relationships of underlying collections make them attractive for text retrieval systems. This paper examines the nature of georeferenced information in academic library catalogs organized according to the Library of Congress Classification (LCC) with the goal of understanding their implications for geovisualization of library collections.

Introduction:

Recent advancements in geovisualization and geographic information retrieval have a potential to transform information systems into highly interactive tools for learners and information seekers. Google Local\(^1\) and MSN Virtual Earth\(^2\) are good examples of such transformations. The visual learning power of cartographic displays makes them highly desirable not only for information systems designed as geographic information systems, but also for systems containing geographic references in the form of text. Academic library catalogs are among information systems rich on georeferences that can be represented cartographically. What a visualization of academic library collections should look like largely depends on the categories of georeferences contained in library metadata records, classifications and on their connections with other subjects. In this study we propose to analyze georeferences in LCC with the goal of improving our understanding of a cartographic visualization of library materials.

We assume that the majority of records that have geographic subject headings have georeferences in the LCC-based call numbers. Commonly, geographic references recorded in subject headings are recorded in call numbers as well, unless the geographic aspect is not important. According to the report from the University of California MELEVYL catalog (Petras, 2004) a large share of library records (53.87%) contain geographic subject headings. 70.56% of 832,108,482 OCLC records in 2002 had LCC call number in 050 MARC field and 16.43% of records had this number in 090 MARC field (OCLC, 2002). These numbers suggest that the number of georeferences in call numbers is statistically significant to facilitate geographic retrieval and visualization. Furthermore, the number of geographic references in call numbers is even higher, if we count georeferences like languages, literatures, religions, and ethnic groups.

Literature review:

This study builds upon findings in information visualization, geovisualization, cognitive psychology, and geographic information retrieval. We begin our discussion with an introduction to the notion of representation, crucial for understanding visualization.

A graphical representation is a building block of any visualization. Graphical representations are used both in cartography and information visualization. They are defined as “an interpretable graphic summary of spatial information” (MacEachren, 1995) and “the way of representing abstract things” (Spence, 2001). A well-established form of representation in geography is a map. There is a great variety of map types: thematic,
analogous, choroplethic, scattered dot, proportional circle maps, and so on. Other representations used in presentations of geospatial visualizations are timelines, map legends, and various representations of underlying collections: 3D spatial histograms of dataset counts, footprints of maps and images (Ancona, 2002), iconic stacks, differently shaped colored blocks (Ahonen-Rainio, 2005), multidimensional icons (Spence, 2001). Together these representations facilitate data exploration and knowledge discovery.

The difference between graphical representations and representations used in library and information science is important to clarify as well. Library representations are associated with surrogate metadata records that have to be exact copies of the original documents, so that users could unequivocally recognize a collection item described in a record. This is not always true in case of graphical and cartographic representations. Some representations with high degree of abstraction (like Beck’s London Underground map) can be even better than representations that have very high degree of fidelity (i.e., those that are accurate replicas of originals). The reason is simple – because “abstraction (that is schematization) and omission of information … reduces the otherwise unmanageable glut of information to an amount that can be processed by mental computing equipment” (Card, 1999, 11). Abstractions highlight the salient features of information and make them easy to comprehend.

Cartographic representations are data dependent (Fairbairn, 2001). Geographic data has certain properties. First, it can be very precise or may lack precision. For some map related tasks accuracy can be safety-critical: for instance, the task of aeronautical navigation (Peterson, 1996). For other tasks, categories may lack well-defined boundaries: for instance, when we talk about folklore in Carpathian mountains, it does not always matter to what specific location in Carpathian mountains we refer to. Second, geospatial data are inherently structured in two (longitude and latitude), three (position above or below the Earth’s surface) or four (time) dimensions, while they are often unstructured in others. Third, geospatial data are typically collected at multiple scales, with fundamental differences in entities and their semantic structure across scales. For example, things defined as objects at one scale may be conceptualized as fields at another, or not represented at all (MacEachren, 1995, 4). And lastly, we should remember about geographic classifications. Geospatial classifications are not library classifications. They are sometimes designed specifically for visualization purposes to designate areas with specific attributes. An example of a classification suitable for visualization can be a classification of languages, a classification of religions, a classification of countries. Each category in such classification fits into a nested hierarchy based on a container relationship and does not overlap with other categories in terms of space. Library classifications, however, often have overlapping (in terms of space) categories.

Until the 1970s geographic classifications included only mutually exclusive non-overlapping categories (MacEachren, 1995; Peuquet, 2002). Nowadays cartographic animations, cartographic movies and interactive maps not only allow the display of non-overlapping classifications but also overlapping and competing classifications by overlaying multiple representations containing different classifications. For example, an interactive map (as well as, an animation or a movie) may include a series of cartographic representations showing changes in administrative political boundaries, changes in linguistic territories, or climatic zones.

Another important property of cartographic representations in geographic information systems that we should be aware of is that cartographic representations may have composite structure and may include representations of space and representations of underlying collections. Cartographic representations of space show how people divide the space, for example into countries, counties, provinces, biomes, soil zones, linguistic zones,
physiographic features and so on. They are typically present in the base maps. A base map is “the framework layer upon which other layers of data are displayed” (Hill, in press). Graphical representations of collections provide summaries of collections. It can be any other pictorial representation mentioned above (a timeline, a legend, a multidimensional icon and so forth).

Despite their virtues, representations are not without limitations. The application of each form of representation is limited. “Just as there is no perfect screwdriver which optimally satisfies all purposes, users and circumstances, so there is no perfect form of representation” (Peterson, 1996, 13). One representation may be computationally efficient for dealing with one part of a problem-solving, reasoning, or concept acquisition task, while another representation may be more advantageous for another. Thus, users in anthropology may find a map showing the location of ethnic groups more useful than a political map. A historian searching for materials on some specific battle in World War II may find a historical map more intuitive than a contemporary political map. A student studying a foreign language, may prefer a linguistic map to an administrative map, because it conveys more information. A topographic map may assist better in a road finding task than a city map (Ahonen-Rainio, 2005). Therefore, representations should receive a significant attention in the design of visualizations for academic library collections.

**Methodology:**

The choice of LCC for this analysis was not accidental. Unlike LCSH where georeferences are grouped in arrays in one facet, LCC presents georeferences in context. LCC contains various snapshots of geospatial knowledge in various domains. Each snapshot contains a linguistic representation of geographic space. These representations convey information about divisions of space. It appears that various disciplines do not have the same worldview. Moreover, views change with tasks and types of data. Geographers and psychologists (Peuquet, 2002) assert that linguistic and graphic representations are interlinked and can be translated from one form of representation into another. In this study we intend to explore the possibility of translating linguistic representations into pictorial representations.

This approach is different from statistical content analysis that we usually carry out in information retrieval. We argue that for cartographic visualization, many other aspects are important besides statistical analysis. Cartographers, for example, look at the areal coverage, map scale, density of observations, why the data was compiled, and user tasks (Dodge, 2001).

Linguistic representations of space can be found in georeferences in LCC. A georeference is a reference to a geographic location. Two major types of georeferences are differentiated in geographic information retrieval: explicit and implicit. In library catalogs explicit references are recorded in metadata and gazetteers as coordinates. Implicit or indirect georeferences can be found in placenames in subject headings, titles and bibliographic notes, geocodes, ISBN numbers, place of publication, language codes, call numbers, classifications. Indirect references require additional computational steps for them to become explicit (e.g., the system should be able to extract geographic names and assign coordinates to placenames) (Hill in press). Whenever it comes to geographic retrieval in library catalogs, we usually think of geographic subject headings as a place to start. Geographic subject headings include names of countries, cities, continents, physiographic features and so forth. However, both LCSH and LCC are imbued with a number of other concepts that bear geographic connotation and can be represented in terms of coordinates. These are languages (like Greek, Ukrainian, Russian, Slavic and other), religions (Russian
Orthodox, Ukrainian Orthodox, and so forth), ethnic people (Russians, Basques, so on). In this study we looked at various geographic indicators.

To analyze linguistic representations we consider time, scale, types of georeferenced data, associations with certain geographic classifications, and overlapping categories in classifications. Three of these categories (time, scale and types of georeferenced data) require more detailed explanation.

Time. Time can be defined as “an interval, especially a span of years, marked by similar events, conditions, or phenomena; an era” (Lexico Publishing Group, 2006). Time in LCC is measured in periods. Periods are not merely convenient collections of years. They are thematic categories of time that require substantiation by the historian, literary critic, or some other specialist (Frommeyer, 2004, 200). In this study we look at temporal aspects of geographic arrangements of LCC and their implication for cartographic visualization.

Scale. Scale is an important aspect of geographic categories. Library classifications allow indexing with the concepts of varying specificity. For example, items can be indexed at the level of an hemisphere, a continent, a country or a region, a county or a province, a city, or a more specific place (like a museum, or any other place of interest). The reason why we think it is important to consider scale is because all these categories may have associations with different representations and therefore should be linked to different representations at different scales.

Types of georeferenced data. Some georeferences may be associated with aboutness of specific entities (e.g., museums, laboratories, scientific institutions, periodicals and so forth). We are trying to get an insight onto what types of data is referenced geographically in library catalogs.

Discussion of observations:

In this paper we present only a few examples of geospatial snapshots and offer possible visualization solutions.

The most interesting sections for visualization can be found in schedules D3, E and F4. They are interesting because they combine temporal, geospatial, and topical aspects. These schedules include general History (D1-2009) and history of individual parts of the world: like Great Britain, Germany, Eastern Europe, United States, and so on.

In General History we can distinctly identify snapshots of historical periods and events. An example of an event is a section on World War II (D741-809). In this part of the schedule you will find a number of subjects with geographical arrangements by countries. For example, military, naval, submarine, aerial, engineering operations are organized by period and country or region. Furthermore, georeferences can be found not only in the names of geographic locations, but also in the names of the battles that took place in specific locations (e.g., Leningrad, Siege of, 1941-1944, Stalingrad Battle of, 1942-1943). This snapshot presents a generalized linguistic representation of the world during the World War II and provides references to all places that participated in operations. Progressions of war through time and changes in borders are not well represented in this example. The temporal aspect is often missing too. While some battles have temporal dates and ranges, the time when countries entered the war is not clearly stated.

LCC records geographic knowledge in a static format. Whenever a new category appears, the new concept is added to the existing structure but does not reflect the changes in the shape of the regions (which is the main defining criteria for geographic classifications). For this reason, some geographic categories overlap in the section on World War II (for instance, Yugoslavia and Slovenia, Ruthenia and Soviet Union) making it
difficult to represent them cartographically. Such overlaps are not numerous in this part of the schedule, however, and could probably be resolved with adequate representations.

Classification schedules allow the linking of resources at different geographic scales: countries and regions, continents (e.g. D766.5 Africa: general works), as well as individual cities and places. It is different from a collection on Google Local, where all items are georeferenced with the highest specificity and accuracy and therefore can be linked to the most detailed representation.

To represent the world in World War II graphically, we should carefully examine the territories of the countries in World War II, changes in territories and decide which political map will be able to represent this part of LCC classification and whether one representation can offer a viable solution at all.

In history schedules that are more focused on the history of individual countries, linguistic representation of space and time looks different. These snapshots suggest that visualization will be more interactive, because classifications are arranged not only geographically but also chronologically and therefore cartographic representation should also include a timeline. Each country will have its own representation of time. Moreover, moving the slider along the timeline should invoke changes in cartographic representations. Consider, for example, “History of Russia. Soviet Union. Former Soviet Republics” (DK1-949.5). The changes in names of Russia Kievan Rus’, Muscovy, and time periods linked to the times of individual tsars remind us of geospatial transformations of Russian territories. This part of the LCC classification (D70.A2 - D293) can be linked to a series of historical cartographic representations, showing transformations. A presentation of library collection with historical maps may provide users with better understanding of library collections.

Besides individual histories for various countries, history schedules also include georeferences denoting local history, where one can find information about a specific place (country, county, province, city, historical place). This information is not time sensitive. DK500, DK508 include such references to places in Russia and Ukraine. Georeferences within these classes typically point to one location. If the name of the place changes, the new names are added to a class, but the call number stays the same. It is possible that a contemporary political map with various scales may well represent these parts of LCC.

Language and literature schedule P5 has interesting sections suitable for geovisualization as well. These are sections on individual literatures and languages: e.g., Russian language and literature (PG2001-2826, PG2900-3698), Ukrainian language and literature (PG3801-3987). They are organized by topics and literary time periods. Time periods are different for each literature. Periodization in literature serves to demarcate recognizable contours of style, trends, shifts in literary creations, or criticism. Within each time period, LCC subjects are organized by individual authors and their works. Authors are listed alphabetically, but also can be rearranged chronologically, because each author’s name is followed by the dates of birth and death.

Since literatures and languages are listed by language or literature, perhaps, the most suitable cartographic representation for all languages will be a linguistic map. LCC classification of languages resembles the classification of languages described in (Ruhlen, 1987). Such classifications serve as foundations for the design of linguistic maps as shown in Figure 1.
It is also possible that we should think of multiple linguistic maps. Contemporary linguistic maps include only existing languages. LCC includes topics related to extinct languages, like Old Church Slavonic. Extinct languages could be represented on historical linguistic maps.

Caution should be observed when selecting a proper linguistic map because not all languages are always present on linguistic maps. Each cartographic representation is derived statistically and some categories may be omitted due to the low statistics, while linguistically they are present in library classifications. For example, Csángó language (Hungarian or Romanian dialect) is spoken in villages in Romania and Moldova. Library collections have resources and subjects about this language and folklore, but it is often hard to find this language on linguistic maps, because its usage is restricted to a small territory.

Another snapshot from LCC may provide clues to the visualization of archeological collections (GN778.52 -885). They are organized by periods: Stone age, Paleolithic, Mesolithic, Neolithic, Copper and Bronze ages, Iron age. Within these periods the schedule has divisions by special cultures and peoples, as well as geographical division by region or country. There exist many historical maps that show locations of ancient cultures and peoples and that could possibly provide solutions to the visualization of these schedules. These representations can be used alongside with historical maps or political maps showing administrative units. Each presentation should include a time scale.

Many other snapshots of linguistic representations of worldviews can be found in other LCC schedules. Besides snapshots, one may notice that geographic arrangements are used for geographic distributions of invertebrates, insects, vertebrates, and local plants in schedule Q, for arrangement of books about individual diseases, hospitals, clinics, hospices and nursing homes in schedule R6, corporations, land use, vital events, dictionaries, societies, congresses in schedule H7, handbooks, sports, weapons, arms, military life,
customs, ceremonies in schedule U\textsuperscript{8}. Throughout all the schedules, georeferences refer to
directories, services, boards, history, laboratories, institutes, research and experimentation,
study and teaching, schools, statistic and surveys, practices (e.g., medical practice), catalogs
and collections, publications (encyclopedias, periodicals, bibliographies, and so forth) and
especially often to museums/exhibitions. Knowing about the types of georeferenced entities
may allow us to think of pictorial representations for each individual type and how to filter
the resources in the visualizations.

Conclusion:
We explored the nature of georeferenced information in LCC and may conclude that
each domain of knowledge in LCC might have different geographic representations, since
divisions of space and references to space vary from one discipline to another. The
advantages of having multiple representations in geovisualization of library collections are
numerous. Multiple representations can be used for filtering the collections according to the
topic, time, scale and document type, design of effective and meaningful summaries of
collections, and knowledge discovery.

Notes:
1  http://maps.google.com/
2  http://virtualearth.msn.com/
history of Europe, part 1 : Gale's Library of Congress classification schedules combined
with additions and changes through ... Detroit: Gale Research Inc.
Service.
Congress classification schedules combined with additions and changes through ... Detroit: Gale Research Inc.
Congress classification schedules combined with additions and changes through ... Detroit: Gale Research Inc.

References:
utility of sample maps as an element of geospatial metadata.” Cartography and
graphic information science. 32(2): 101-12.
Alexandria Digital Earth Prototype. In K. Boner, and Chen, C. Visual interfaces to
digital libraries. Lecture notes in computer science. 2539: 199-213.

Frommeyer, J. (2004). “Chronological terms and period subdivisions in LCSH, RAMEAU, and RSWK: Development of an integrative model for time retrieval across various online catalogs.” *Library resources and technical services*. 48(3).


