

Development and Implementation of the National Geographic Names Database

Roger L. Payne

U.S. Board on Geographic Names

U.S. Geological Survey

In response to an acknowledged need and growing demand for a relatively complete list of geographic names, numerous attempts were made by governmental agencies and private concerns to develop a reliable, digital geographic names database that would include basic locative and description information for each name. Because of the scope and nature of the project, most users looked to the federal government to develop and maintain such a database. The U. S. Geological Survey, which is responsible for the National Mapping Program and related domestic geographic names activities, agreed to undertake the task of designing, developing and implementing what has become the nation's official automated geographic names repository.

Introduction

In addition to their intrinsic interest geographic names are necessary for spatial reference in order to make sense of an otherwise confusing and often unintelligible landscape. Geographic names are thus the means by which we describe and understand the landscape. Geographic names are initially connotative; however, in the course of their development they become denotative, labels which serve as referents to specific landmarks within a spatially defined area. The roles played by geographic names in referring and categorizing can be complex, employing personal, highly variable, and often idiosyncratic methods of conception and perception.

Background

The collection and publication of geographic names in the form of gazetteers and geographic indexes has been an expedient means of

308 Names 43.4 (December 1995)

identifying places, features and areas in the United States since at least the eighteenth century. Geographic names are needed as references and for identifying and relating locations. The growing number of names resulting from expansion and settlement during the eighteenth and nineteenth centuries made gazetteers even more essential than they already were; yet the ever-increasing number of names made impractical anything approaching completeness on even a regional or local scale, much less a national scale. Nevertheless, a large number of state and national gazetteers were published in the nineteenth century, but these were limited since they usually included only populated places and such larger physical features as mountains and rivers. The goal of completeness, although desired, was generally recognized as difficult or even impossible to attain. One extensive effort was begun in 1892, when the United States Geological Survey began cataloging geographic names as part of the national mapping effort then underway. Gazetteers of 12 states and three territories were published between 1894 and 1906. While accuracy and comprehensiveness were always goals, these gazetteers included only the names found on the topographic maps of the Geological Survey, which were usually at scales of 1:125,000 (a 30-minute area) or 1:63,360 (a 15-minute area). Thus, many local names and names of "minor" features were omitted. For these reasons, and since only a small percentage of the country had been mapped by the beginning of the twentieth century, the program was discontinued.

Standardization

A century of westward expansion, with subsequent naming by numerous groups and individuals, had created onomastic confusion on maps and in documents. The confusion was exacerbated by a proliferation of maps and scientific reports associated with exploration, mining and settlement of the American West, which accelerated after the Civil War. The accuracy of large numbers of names, their spelling and applications became serious problems to mapmakers and others who required uniform, nonconflicting geographic information. Because of the magnitude of the problem, and because it was most acute in lands still under federal administration, interested parties looked to the federal government for its solution. In response, on September 4, 1890, President Benjamin Harrison signed an executive order establishing the United States Board on Geographic Names. The Board was given

authority to resolve all unsettled questions concerning geographic names, and its decisions were to be binding on all departments and agencies of the federal government. Since its beginning in 1890, the Board has undergone periods of adjustments and changes in policies and procedures, the most significant in 1906 and 1947. In 1906 the Board was given authority to change names and approve new names for unnamed features as well as adjudicate conflicts concerning names. In 1947, the Board was abolished and re-established by the United States Congress under Public Law 80-242. This was a milestone in the history of the Board because it was mandated by this law to "provide for uniformity in geographic nomenclature and orthography throughout the Federal Government."

Automation

The need for an efficient means of processing and applying geographic names had been apparent for many years. Since the mid-twentieth century, the use of automation has been seen as a response to this need since it provided a potentially efficient means of processing and analyzing the nation's stock of geographic names. Thus from the mid-1950s through the 1960s, numerous attempts were made by various organizations to establish automated files of geographic names. For the most part, these files were limited in scope and content since the fledgling computer industry had not yet technically evolved to a point at which the great quantity of data could be stored efficiently and accessed with ease.

A single, and relatively complete, repository of geographic names containing basic locative and descriptive (e.g., type of feature, source of name) data about each name was needed at all levels of government and by diverse users in the private sector as well. The U. S. Board on Geographic Names was eager to cooperate with other agencies since such an automated database would be the perfect vehicle to aid names standardization throughout the federal government. In 1976, the U. S. Geological Survey, which provides support for domestic names activity to the U. S. Board on Geographic Names and is responsible for the National Mapping Program, agreed to undertake the task.

Perhaps the most important requirement of an automated names system is that it provide a means by which information can be stored and retrieved with no preconceptions about the data and thus with no

310 Names 43.4 (December 1995)

introduced bias. Functionally, a data management system for use with geographic names has to have this quality above all others. Other desirable qualities are rapid retrieval time, efficient means of partitioning data, flexible output and report formatting, and low cost. A further desideratum for a data management system is that it be user friendly, easy to learn and easy to access by a variety of users. Additional requirements include ease of addition of data and data categories and the ability to append other software for specialized and sophisticated data manipulation.

A database management system (DBMS) demonstrating these characteristics and suited to processing text of variable length was selected, tested and made operational by 1978. The arduous task of data compilation had been under study for about four years, during which names of features in Massachusetts were collected to serve as a pilot project. After considerable adjustment and modification of compilation methodology, names information in Kansas, Colorado and Rhode Island were compiled from 1976 to 1978 as a test of the new data compilation procedures and the data were entered into the National Geographic Names Database (NGNDB), the nation's *official* automated geographic names repository.

The initial phase of data compilation spanned a four-year period during which geographic names and locative attributes were recorded, encoded, and entered into the NGNDB from the large-scale topographic maps published by the U. S. Geological Survey. These maps were selected because they represented the official base map series of the United States, and they provided the single most comprehensive source of geographic names. Upon completion of the initial phase of data compilation, work was begun to supplement the basic inventory by adding additional names and attributes from the U. S. Forest Service maps and the National Ocean Service charts, thus completing data collection from the three primary map and chart sources created and maintained by the federal government.

It was clear from the outset that projected and actual use of the names information would require a database as complete as possible for all categories of feature types (mountains, streams, populated places, etc.). It was further recognized from intensive investigations of small areas that only about 40% of the known names were available from the Geological Survey and Forest Service maps and the National Ocean

Service charts. Thus, in order to meet the goals and requirements of the NGNDB, and to assist the national mapping effort, the U. S. Geological Survey agreed to fund a second phase of names compilation which would incorporate the established methodology and procedure and the use of automation to examine official state and local maps and documents as well as historical materials to complete insofar as possible the entries in the NGNDB. Checks built into the examination procedures ensured that the principles, policies, and procedures of the U. S. Board on Geographic Names were implemented. This phase of compilation was begun in 1982, and is being accomplished on an individual state basis. A national completion date of about 2005 has been projected.

The information contained in the National Geographic Names Database and the products derived from it assist in standardizing geographic names in the United States and provide support to the National Mapping Program. Other uses of the NGNDB materials include information for use in other mapping and charting programs, emergency preparedness, local and regional planning, routing requirements, marketing, site selection and analysis, and genealogical research. The database is available in a wide variety of conventional and magnetic media, including CD-ROM and over the Internet.

Maintenance

The integrity and viability of any data system are dependent upon the quality control and assurance instituted during system development and maintenance. A maintenance program for the NGNDB was first begun in early 1982, shortly after the first phase of data compilation was completed. Procedural and policy issues at this time unfortunately limited the maintenance effort to participation by the U. S. Geological Survey only. The effort was further hindered by hardware and software problems. Thus, the first attempt at a maintenance program met with limited success.

By 1985, a concerted effort was begun to completely redesign the maintenance program. The approach centered on the development of data entry software that would be universally distributable in order to allow participation by a large number of agencies and organizations. The extent and nature of names data and the variety of participants would render complete data verification impossible. Therefore, to ensure data integrity, software design proceeded with the goal of incorporating data

312 Names 43.4 (December 1995)

checking and validation routines. Another goal of the software design was to provide the versatility necessary to allow expansion and alteration of the database as required by system changes and enhancements.

Research and re-design of the maintenance program was completed in early 1987, and the revised program was tested for six months by U. S. Geological Survey mapping center personnel and the domestic names staff of the U. S. Board on Geographic Names. After a successful period of testing, the maintenance software became operational at the U. S. Geological Survey and the U. S. Board on Geographic Names on October 1, 1987.

The task remaining to fully implement the maintenance program was to establish rules governing the disposition of the data collected. Geographic names conflicting in form, application, or usage among maps, charts and documents produced by the U. S. Geological Survey, the U. S. Forest Service, the National Ocean Service, and other federal agencies are referred to the U. S. Board on Geographic Names for resolution according to the procedures established by that body. Other names may be submitted directly to the geographic names staff, which reviews them for overall adherence to established policy before they are entered into the database. Other federal agencies provide information, but because their data are irregular or infrequent, they provide it directly to the domestic names staff for whatever reformatting and processing may be necessary.

Several state mapping agencies have expressed an interest in participating in the database maintenance program. Also, since the data compilation program associated with the development of the NGNDB will largely complete the initial task of the Place Name Survey of the United States (PLANSUS), that organization may assist with the maintenance of the database. For the present, however, participation is limited to civilian federal mapping agencies. Because the federal agencies often work closely with state agencies, participation by both could create redundancy. The program will be monitored and analyzed over the next two years to determine if participation by individual states and PLANSUS is warranted.

Modernization

In March of 1994, the National Mapping Division (NMD) of the U. S. Geological Survey, which administers the National Geographic Names Data Base, undertook a project to identify and migrate the

database to a modern and more reliable data management system. The division determined that all of its various databases, including the NGNDB, should be managed and maintained by one common data management system. Oracle® was chosen as the Division's standard and therefore the NGNDB's new data management system. A team consisting of NGNDB experts and Oracle applications developers designed, developed, and implemented a version of the NGNDB under Oracle. It was decided that a rapid prototyping approach would be used to design the new system and that a subset of the database would first be converted. An Oracle version of features classified as "populated place" has been operational and available for testing since April of 1995. Full conversion of the NGNDB has now been completed. This conversion allows considerably more efficient data management and integration as well as greater flexibility in data manipulation and product development. Projected future enhancements include a graphic user interface with geographic area search capabilities.

Summary

Geographic names are applied to landmarks or specific natural and cultural phenomena on the landscape. These landmarks are essential in navigating the spatial environment, and it is by naming or labeling these landmarks that we are able to reference them and place them into geographic and social perspective.

At the end of the nineteenth century, the western part of the United States was being heavily mapped by several — unrelated — agencies of the federal government, causing much conflict and confusion in the use and application of geographic names. A need for standardization was apparent. In 1890 the U. S. Board on Geographic Names was created by President Benjamin Harrison. In 1906, the Board's responsibilities were extended from simply standardizing names to implementing name changes, and approving new names as well as adjudicating name controversies. In 1947 the Board was reorganized under public law mandating the standardization of geographic nomenclature in the United States.

The use of automation has been an obvious boon to the processing, analysis and standardization of geographic names. Automation and the development of the NGNDB has allowed large quantities of existing names to be processed and analyzed rapidly and efficiently, while

314 Names 43.4 (December 1995)

providing a vehicle for standardization. Automation has also made possible the long range program of a national names survey and data collection program as well as the revitalization of the National Gazetteer Program with both conventional and digital products. The 25-year data compilation project that will establish the most comprehensive and complete national geographic names inventory ever attempted is on schedule and should be completed by 2005, after which the NGNDB will be in a maintenance environment with active participation from the primary civilian mapping agencies as well as the Board on Geographic Names. Additional participation in the maintenance program by the Place Name Survey of the United States and various state agencies is under evaluation.

In addition to completing the extensive data compilation program, the modernization of the automated aspects of the database is mostly complete. The National Geographical Names Data Base of the United States is considered one of the most comprehensive and efficient automated toponymic tools in the world.

Note

Any mention of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. government.

References

- Burrill, Meredith F. 1990. *1890-1990: A Century of Service — United States Board on Geographic Names*. Washington, D.C.: U. S. Department of Agriculture.
- Orth, Donald J. and Roger L. Payne, 1985. *Principles, Policies and Procedures: Domestic Geographic Names*. Reston, VA: U. S. Geological Survey.
- Payne, Roger L. 1987. *Geographic Names Information System: Data Users Guide*. Reston, VA: U. S. Geological Survey.
- _____. 1995. "Geographic Names Activity in the United States of America." *Namenforschung Name Studies Les noms propres*. Ernst Eichler, Gerold Hilty, Heinrich Löffler, Hugo Steger and Ladislav Zgusta, eds. Berlin: Walter de Gruyter.
- Randall, Richard R. 1990. *United States Board on Geographic Names*. Washington, D.C.: Defense Mapping Agency.