The GNIS and the PC: Two Tools for Today's Toponymic Research

Lewis L. McArthur
Portland, Oregon

The personal computer has revolutionized geographic name research through its ability to store and manipulate vast quantities of data. Oregon has made considerable use of the information made available through the Geographic Names Information System and has combined the basic data with a program designed to aid in researching the origins and history of Oregon names. Procedures have been developed which allow the display of various types of names and naming patterns tailored to specific onomastic interests or the interests of researchers in other disciplines.

James Burke, the author of *Connections* and *The Day the Universe Changed,* was interviewed in Portland, Oregon, in March, 1995. He was asked: "You cite moments and developments in history that have altered civilization. Do you think the computer information age is one of those turning points?" Burke replied: "I think this is the most important turning point in history, since maybe the development of the alphabet. I say that advisedly because I think what it's going to do is change the way we think — to look for patterns rather than going on knowing more and more about less and less...." I will speak first to "more and more" and then to "patterns."

Learning More and More

The Geographic Names Information System (GNIS) is the basic digital national gazetteer.[1] It includes all names found on the most recent U.S. Geological Survey (USGS) quadrangle maps plus names from other maps by federal agencies such as those of the Forest Service and the National Ocean Service (NOS). The compilation has been expanded in some states where names on maps produced by state agencies have been added along with variant and historical names taken from older maps

Names 43.4 (December 1995):245-254
ISSN:0027-7738

and reference texts. The GNIS is on line at the USGS in Reston, VA and is available on CD-ROM, tape or floppy disk; it will soon be accessible on the Internet as well. It is not free but the cost is reasonable; the CD sells for about $65. The GNIS database allows a researcher with even a modest personal computer (PC) such as a 386 to examine, sort, manipulate, add and delete data for a vast array of research projects. The PC and GNIS are being used in Oregon to facilitate the collection and recording of name data and to expand the study of geographic onomastics.

The GNIS provides the data which for the first time permits the examination of a large corpus of geographic names, such as a complete state file. The Oregon file contains some 53,000 records (a record is a single geographic name and its associated information, such as the type of feature named and its location) and is currently being used for three primary purposes: first, researching and recording the origin and history of Oregon geographic names (we now have information on some 7,000 names); second, classifying and recording the types of names according to the system developed by PLANSUS, the Place Name Survey of the United States (Smith 1992); and third, analyzing the linguistic aspects of the names — their language(s) of origin, mode(s) of formation, and the like. Each of these areas is complex and will be discussed individually.

Oregon Geographic Names

Work on the origins of Oregon's names was begun prior to World War I by my late father, Lewis A. McArthur. He began with the materials at hand — lists of counties, cites, towns and post offices. To these he added major geographic features such as streams and mountains. At that time there was no comprehensive or reliable gazetteer; only a small part of the state had been adequately mapped and that generally at 1:125,000. However, since his work commenced rather early in Oregon's statehood, many primary sources, including people who had bestowed the names, were still alive. In addition, because of his extensive knowledge of early exploration and settlement, he was able to provide relatively complete accounts of the more important names including, in many cases, interesting background material.

After World War II, the National Mapping Program of the U.S. Geological Survey made considerable progress in mapping Oregon at

1:62,500. This work added innumerable names of creeks, hills, lakes and other landscape features. Father died in 1951 and by the middle 1960s, I had time to become involved. Coincident with other work for the Oregon Geographic Names Board, I made lists of all the names on each new quadrangle map and looked for their sources and their histories. Meanwhile, the computer was becoming an important tool in many fields. (Here I must digress for a moment because my first work on the revisions of Lewis A. McArthur's *Oregon Geographic Names* was directly linked to the use of this then new tool. Father had had considerable experience with typesetting and printing and his three editions of *Oregon Geographic Names* were all prepared using old-fashioned hot lead linotype. I need not comment on the time and trouble involved in the proofreading and correcting of endless galleys. The 4th edition of *Oregon Geographic Names*, published in 1974, was my first and I was lucky enough to find a man named Gordon Nelson who was both a journeyman printer and a computer programmer. He keyboarded the 3rd edition text and my considerable file of changes and additions on an IBM mainframe. Nelson was not only knowledgeable but a near perfectionist and his program was far superior to most present day desktop publishing software. Its draft printout showed the final justification, hyphenation, line length and leading as it would eventually appear. As the book was completely formatted on the computer, the three indices were produced with the correct ultimate printed page numbers rather than the page numbers of the usual draft printout. Nelson later prepared the camera ready copy for the 5th edition in 1982, but the advent of the PC doomed his business and forced him to seek other work. When the 6th edition was being prepared in 1991, Nelson's tape file was converted to WordPerfect® and subsequent changes and additions were done by me on a PC. The file still contains Nelson's codes for simple internal indexing. The names of people mentioned in the text are coded with *$rname$e*, variant placenames with *$gplacename$e* and American Indian names with *$nname$e*. These codes provide the biographic, geographic and American Indian name indices, respectively. The biographic index is kept on diskette and contains all the names of individuals mentioned in the text. The geographic index contains all names not headed but mentioned in the body of the text. These are usually obsolete or variant names. The American Indian index will be discussed under "languages,"

below. Unfortunately, the printed page numbers for the 6th edition index had to be keyboarded after typesetting.)

There is a wide variety of sources of information on name origins and their histories. Most states have a number of historical societies which collect information and publish county histories. Many diaries and recollections are already published and new ones appear regularly. In the western states the Bureau of Land Management has preserved the General Land Office original entry tract books. The Oregon tract books have been sent to the National Archives but they are available on microfilm. These books show by section, township and range the people who first filed on or purchased property. Many map names for creeks, lakes and terrain features match those in the tract books. Of course, such a match is not definitive but it provides valuable leads to other sources. Microfilm is also available for plats, the original surveys of each township. Often the plats show homesteads by owners' names. In the late 1920s, Chas. F. Metzger began a series of county maps for Oregon and Washington. These show the owner of each plot of land other than most city lots. They also show towns, roads and railroads, donation land claims and some subdivisions. Early census records are invaluable. Other good sources are local philatelic and railroad societies. Oregon has both and they often can provide the origins of long-gone post office and railroad station names.

Newly found name origins and histories are prepared as text for additions or corrections in subsequent editions of *Oregon Geographic Names*.

Recording and Classifying Names

After the basic information concerning a name has been collected, the next step is to construct a database formatted to show which names have been studied and where the information is recorded. PLANSUS has made a number of detailed recommendations concerning the classification of geographic names; these are summarized in Smith 1992. Here I will not address the many issues considered by Smith, but rather I will report on the system I use to divide the name corpus into smaller segments delimited by fields of interest or areas. This same system also allows a framework for discussing the patterns of names which appear in the data.

Figure 1: A Sample of the GNIS Oregon file, as received from ODOT

IDNO	NAME	FEATURE	COORD	COUNTY	MAPNAME	ELEV
41000151	Abbot Creek	stream	443413N1213713W	Jefferson	Prairie Farm Spring	
41000152	Abbot Creek Campground	park	443245N1214019W	Jefferson	Candle Creek	
41044678	Abbot Burn	area	451050N1214900W	Clackamas	Wolf Peak	4400
41032851	Acton (historical)	ppl	453555N1192505W	Morrow	Butter Creek Junction	1087
41000216	Adel	ppl	421039N1195351W	Lake	Adel	4542

Figure 2: A Sample of the GNIS Oregon file, after modification

IDNO	NAME	FEATURE	COUNTY	DATE	BIBLIO	USED	TYPE	RELAT
41000151	Abbot Creek	stream	Jeff	1930	OR-T1	PRIM	BIOG	COMM
41000152	Abbot Creek Campground	park	Jeff			SHFT	BIOG	
41044678	Abbot Burn	area	Clac	1890C	OR-T1	PRIM	BIOG	ASSO
41032851	Acton (historical)	ppl	Morr	1879	OR-T1	PRIM	UNKN	
41000216	Adel	ppl	Lake	1896	OR-T1	PRIM	COIN	

The U.S. Geological Survey provides the Oregon Department of Transportation (ODOT) with a full GNIS file on tape. This is kept on the mainframe since its primary use is for computer aided state mapping. ODOT provides me with the Oregon file on disk in Dbase® format, a sample of which is shown in figure 1. The information is given in seven fields:

IDNO: The eight digit number assigned by the USGS,

NAME: The name as it appears on a USGS or USFS map, or NOS chart,

FEATURE CLASS: The type of geographic area named (populated place, mountain, park, etc.),

COORDINATES: Latitude followed by longitude,

COUNTY: The county in which the feature is located,

MAP NAME: The name of the quad or quads from which the name and descriptive information was obtained,

ELEVATION: The height of the feature in feet.

This file takes about 6 megabytes of space and may be stored on either a hard drive or floppy disks. I modify the file I receive from ODOT by deleting the COORDINATES, MAPNAME and ELEVATION fields and adding the five PLANSUS fields of DATE, BIBLIO, USED, TYPE, and RELAT. These changes result in the file shown in figure 2, which contains nine fields:

DATE: The date the name was first used,

BIBLIO: Bibliography, the source of information,

USED: Whether the name is

PRIMARY: The original use of a name in an area, a

SHIFT: Other uses or combined forms of a primary name in a related area, or a

TRANSFER: A name brought from outside the area, often the name of a previous residence.

TYPE: whether the origin of the name is

BIOGRAPHIC: Refers to a personal name or surname or the name of a group,

PHYSICAL: Denotes or describes a physical characteristic,

BIOLOGIC: Refers to any flora or fauna or subdivision thereof.

ACTVITY: Refers to any activity, human or natural.

COINAGE: Combines phonemes or letters without standard lexical meaning or reflects errors in transmission.

MISCELLANEOUS: Does not readily fit into another category.

UNKNOWN: Researched names whose origins remain unknown.

RELATION:

ASSO: An individual or event more or less directly associated with the name.

COMM: An individual or event not directly involved in the naming, but commemorated by the application of the name.

CONT & INCI: Usually concerned with activities which are continuous or incidental.

FLOR & FAUN: A name from the plant or animal kingdoms.

For the five added fields, I have keyboarded the data from the 6th edition of *Oregon Geographic Names*. It took about a week to enter some 6000 names. Oregon has three other excellent placename books, one covering a single county (Clarke 1977), one a single National Forest (LaLande 1995) and one a major river basin (Williams 1985). Data entry from these three works is now partly completed. The nine-field file now becomes the basic research tool. When Oregon receives an updated GNIS file, this data entry need not be repeated as there is a simple program that will combine an updated GNIS master with the added fields. The only requirement is that both files must have matching IDNOs and NAMEs.

As the number of names with known or probable origins increases, the number with no information diminishes. The file can be sorted so that all names with no data are listed together. A visual inspection will give a good indication of which names fall into particular categories; all those which are probably biographic, for instance, are coded **BIOG** but without any **Biblio** reference. These can be sorted by county, creating a list that is small enough to be manageable. I have given copies of such lists to county historical societies and other interested groups, where I have often found people eager to help. When and if origin and history information is found, it is recorded and a **Biblio** entry made. This process separates facts from conjectures.

Languages

The linguistic aspects of the names present numerous, often serious, problems. The 1st edition of *Oregon Geographic Names* in 1928 listed the language origin(s) of the names. This practice was continued with minor modifications through the 5th edition, but the divisions were arbitrary and not satisfactory. I have asked the help of experts in the discipline and I hope to promote discussion of the subject at future PLANSUS and American Name Society gatherings.

American Indian names in the 6th edition are coded and indexed. I am not certain all names are included and the coding will be checked for completeness when the next edition is proofed. These names present several problems, the most important being, "what are American Indian names?" In Oregon we have at least the following categories of names which might be considered "Indian" to a greater or lesser degree:

- English approximations of names actually used by indigenous tribes. The tribes had no writing systems and the approximations are often confusing and contradictory. There are 12 different forms of *Willamette* used by explorers from Lewis and Clark in 1805 to Wilkes in 1841. Every important feature with an English approximation has two or more recorded forms.
- Chinook Jargon names, usually applied by whites. The Jargon was a trade language made up of words taken from different tribal dialects combined with French and English.
- Imports. Due largely to patterns of exploration and settlement, Oregon has numerous American Indian names which have been imported from across the country (e.g., *Illinois, Oswego, Oneonta, Oceola).*
- Others. How should *Indian, Squaw* and similar names be classified?

A great deal of work is needed in the area of languages. I cannot predict the direction this work will take, but some rationalization of linguistic classes would permit, I am sure, the recognition of patterns currently obscure.

Patterns

As our database of knowledge increases, we will undoubtedly refine its format. Electronic indexing will facilitate origin and history studies and new computer hardware and software will allow the rapid creation and testing of suspected and presently unsuspected patterns. The number and variety is both fascinating and unpredictable but the following are examples of the kinds of investigations that current and future hardware and software will allow.

- Types of names. How many of the various types of names (Biographic, Physical, Activity, etc.) are there and where are they located? These percentages can be determined for the state as a whole, for a county or for other selected areas.
- Frequency of a name or names. How often are particular names used, where and by whom? What are the relative frequencies of generics (*creek* versus *brook* versus *branch*)? Do generic variations correlate with ethnic or cultural patterns? What are the linguistic frequencies and how are these related to the interactions among ethnic groups?
- Density of names. How many names are there per physical or cultural unit? What types of names are common to arid or timbered zones? What are the relationships between name densities and population densities?

As we progress from the collection of specific facts, we can use the database to overview geographic onomastics. We will be able to see the forest as well as the trees. However, everyone does not agree with Burke's optimistic view with which I began this article. Reed Karaim (1995) raises some philosophical and intellectual doubts and concerns about the computer age and the information super highway. The different perspectives of Burke and Karaim point to both the advantages of and problems caused by the computer. I believe one must take the middle course; make full use of this wonderful tool but do not let it take control of your life.

Notes

1. For a description of the origins and evolution of GNIS and the database which has been generated, see Roger L. Payne, "Development and Implementation of the National Geographic Names Database," in this issue.

References

Clarke, Gordon W. 1977. *Polk County, Oregon, Place Names*. Monmouth, OR: Oregon College of Education.

Karaim, Reed. 1995. "Technology and its Discontent." *Civilization* 2: 47-51.

LaLande, Jeff. 1995. *From Abbott Butte to Zimmerman Burn: A Gazetteer of the Rogue River National Forest*. Medford, OR: Rogue River National Forest.

McArthur, Lewis A. 1992. *Oregon Geographic Names*. Portland, OR: Oregon Historical Society.

Smith, Grant. 1992. "Describing Types of Placename Information." *Names* 40: 299-304.

Williams, Gerald. 1986. *McKenzie River Names*. Portland, OR: Oregon Historical Society, Manuscript.