

# Describing Types of Placename Information

Grant Smith

## Abstract

Any attempt to standardize data fields for the computerized study of placenames should strive to define more types of data fields rather than fewer. Also, standard data fields should be described and organized in a way that builds upon general ideas about language common among philosophers and linguistic theorists. This does not mean that any particular theory or philosophy of language should be adopted but that the context of common philosophic distinctions would likely make the data fields more intelligible and useful both to ourselves and those in other disciplines.

\*\*\*\*\*

In the December 1992 issue of *Names* (39.4: 393–94) I reported that The Placename Survey of the United States, commonly known as PLANSUS or “the Commission,” had voted to identify four types of information as basic to placename research and as “required for placename studies” recognized by The Survey. These four “data elements” are 1) the **name**, exactly as found in some source; 2) the **type of feature**, preferably using the categories of “Feature Class” of the Geographic Names Information System; 3) an indication of **location**, using coordinates or one of several other methods; and 4) the **source** of the information in the first three fields, using any standard citation procedure. Three other types of information have been discussed by the Commission as highly desirable: 5) a **unique identity number**; 6) a **pronunciation guide**; and 7) a listing of **alternates and variants**. A year earlier PLANSUS had adopted Lewis McArthur’s classification of **meaning**: i.e., **biographic, physical, biologic, activity, coinage, miscellaneous, and unknown**—which is included as 5.00 in the list of data fields in the appendix.

The purpose of this paper is not to report on the subsequent deliberations of PLANSUS, but to propose a set of principles that should guide the development and description of placename data fields in general. The Commission is pursuing its task in a practical way, attempting to find common ground among different researchers on a series of specific issues. However, common ground is by nature limited in its parameters. What I hope to outline here are some of the

philosophical parameters of such an undertaking, which I believe should be more inclusive. These may be useful to anthropologists, sociologists, linguists, local historians, governmental agencies, even the philosophers, perhaps – if we organize our data elements in a way that makes sense in the context of common philosophic distinctions.

Let me begin by addressing the first and simpler of my two basic issues. Should we want to describe and standardize *more* or *fewer* data fields? In discussions within the Commission it has become clear that a distinction needs to be maintained between essential information and desirable information. Essential information is simply what we need to know first for the purposes of identification, i.e., to know what we are talking about and to judge the reliability of our information. What we need to know first, obviously, are those four data fields which PLANSUS designated last year as “required fields”: *name*, *feature class*, *location*, and *source*. Geographic names are the subject matter of our studies, and the features and their locations are needed to individuate them. Of course, names may be individuated in other ways, as I shall discuss at the end of this paper, but geographic names must, by definition, be individuated by geographic information. That is to say, the essential and obvious distinctions must be between different features that have the same coordinates (*Bear Lake* on top of *Bear Mountain*), between similar features that have different locations (*Bear Lake 1* and *Bear Lake 2*), and between two names for a single feature. The first three “required fields” designated by PLANSUS will make these distinctions, thereby individuating all geographic names and allowing researchers to compile an accurate and comprehensive data base nationwide.

However, identifying the essential types of information does not resolve the question of how many other data fields might be useful or important. Virtually everyone who compiles information about geographic names includes in his or her study many types in addition to the four fields required by PLANSUS, but agreement on a list of types is difficult because different researchers have different interests and purposes. As Frederic Cassidy noted in his study of Dane County, “Place-names may be classified in a variety of ways, depending on the degree of detail desired, and the direction of the investigation” (29). The local historian will want to know the circumstances of the naming event, while the linguist will want to know the local pronunciation and variant spellings.

At the last two meetings of PLANSUS we have struggled to prioritize the additional types of information beyond the “required” four. The fact that we agreed on three as “desirable” probably illustrates the

similarity of our research interests rather than the logical necessity or absolute value of the three types. Beyond the first four types of information, which are necessary to distinguish one geographical name from another, there is really no reason to prioritize additional fields—*except* as one operates as an individual researcher. That is to say, the mission or role of PLANSUS should be seen as general in nature and thereby different from the interests of individual researchers. If that mission entails the description of data fields which are deemed useful and desirable in placename studies—and I think it does—then that description ought to be as inclusive as possible rather than merely reflective of the shared interests of those serving on a special commission. We need to think about what can be known about placenames rather than about what is desirable or what ought to be known.

The description of data fields requires at least two basic operating assumptions. First, we need to abandon the idea that agreement or consensus is needed to describe important types of data. All types of data are good types of data, so long as they are clearly delineated. We should shun our yearnings for consensus because shared research interests will tend to exclude specific interests. Many of us, for example, may be strongly interested in *pronunciations* and very little in *elevations*, but it is clearly the latter of these two data elements that has more immediate use for the public.

Second, we need to assume that all records will be incomplete records. If we describe many types of information, we must assume that most researchers will not be interested in all types and will not record all types. We must also assume that all types will not be recoverable and/or relevant in this or that particular locality.

What then is the value of describing standard data fields? Or why describe types of placename information beyond the first four “required” fields if we cannot assume complete records or even, perhaps, cooperative researchers? I can think of at least three good reasons: First, the process is educational. As the data fields are defined and articulated, those who do it understand what they are doing better, and once a guide is published those who are beginning to work with placenames will learn a little faster.

Second, the more fields which are described by an official organization such as PLANSUS, the more will individual researchers be inclined to find and record those types of information. There are no guarantees, of course, but a standardized description of possible information will help bring it to the attention of researchers who would otherwise overlook it. Also, students will fill in data fields as part of assignments, and some researchers

will fill in data fields simply because of the mind's impulse toward closure. For a variety of reasons, the enumeration and delineation of data fields will result in more information being recorded faster.

Third, knowledge about placenames, as it is about anything, is found in generalizations and comparisons, which depend on standard categories. Thus, the more categories which are described and filled, the more knowledge we may say we have about placenames. And the more knowledge we have the better.

These are a few reasons why we should try to describe as many data fields as possible and why we should avoid trying to prioritize them beyond the three required for identification. However, just because we might not prioritize our data fields does not mean that they should lack organization. Data fields are of different sorts, and grouping them can be instructive and help us all understand and talk about our subject matter better.

Data fields can be rather easily organized by looking at the types of data actually included in a variety of placename studies and at the descriptions of language and meaning by linguistic philosophers. In introductory semantics classes, meaning is described as the result of an *agent* using a *symbol* to designate a *referent*. In terms of placename studies, the *agent* is a *namer* who uses a *symbol*, which is a *name*, to designate a *geographic feature*.

Thus, the most general description of how language works gives us three categories of data fields: information about the name as an artifact of language, information about the feature as a geographical entity, and information about the namer as a social entity. These three categories are listed in the appendix as 2.00 through 4.08, with a brief description for each data field within the categories.

Much has been written by the philosophers of language about the relationship of the *symbol* to the *referent* and how meaning is conveyed to an audience, and particular attention has been given to proper names as a special, even crucial example of language in action. At issue, usually, is whether a name is a unique, specific identifier rather than a general term, and whether or not the name predicates anything about the referent. Willard Quine has argued that names, contrary to our intuition, are general terms because the same name is often applied to many different referents. In his system of logic he can therefore treat names as carrying some kind of meaning, but he does not describe the nature of that meaning or how it is acquired (181–83). Similarly, John Searle says that names do not “describe or specify characteristics of objects” but are “connected with characteristics of the object ... in a loose sort of way” (170).

Philosophers of language get overly complex because they try to convert all linguistic references to some form of predication. Searle himself points out the conflict in the last two sentences of his chapter on names: "The existence of these expressions derives from our need to separate the referring from the predicating functions of language. But we never get referring completely isolated from predication for to do so would be to violate the principle of identification, without conformity to which we cannot refer at all" (174). Searle's analysis would be greatly simplified if he focused on the associative nature of language and then explained predication as the process whereby a few of the attributes previously suggested by the symbol are carried over to the new referent in some way. This is the manner in which Max Black explains the interaction of tenor and vehicle in a metaphor. However, when Black mentions proper names, he reverts to the usual philosophical view that names are univocal identifiers: "There seems to be no other way of 'picking out' or identifying an admiral— if we exclude nonverbal means, such as pointing or showing a photograph—than by using a suitable designation" (96).

Those of us who do placename studies have a big advantage over the philosophers. We focus on data rather than theories. A very casual glance at placename studies reveals that the dominant interest of most researchers— excluding the Phase II contractors— is not in the actual characteristics of the thing named but in the prior references and associations of the symbol before it is used as a name in the instance at hand. Willard Quine was correct when he said that names are general rather than special or univocal terms and that this fact runs counter to our grammatical intuition. He was wrong only in trying to justify his observation by focusing on immediate referents rather than on prior referents.

If we focus on prior reference as a fourth category of data fields (this category is listed as 5.00 through 5.07 below), we shall do ourselves and the philosophers a favor. Not only do many placename studies focus on this type of data, but the data in this category illustrate the associative and non-discursive nature of most language uses. Names, as we all know, can be quite descriptive, such as *Mt. Baldy*, *Clear Lake*, or *Walla Walla* (meaning "much water"); or they can associate an incident with the site, such as *Jump-Off-Joe Creek* or *Hard Scrabble Falls*. But most often names focus our attention on things that have nothing to do with the feature. Most placenames are commemorations, idealizations, coinages, shifts, and transfers, which predicate very little about the feature but tell us a great deal about what the namers feel is important. Language can be used for explaining things, for expressing feelings, or

for any number of purposes, but in all its uses it is a social product reflecting social values. Placenames, as artifacts of language, tell us more about local history than about the physical features they designate. What is predicated about the features is how humans relate to them, the meanings brought to them. By grouping together information about prior references, we focus on a type of information that is of primary interest to a majority of researchers, that does not really fit into the other categories, and that clarifies the social and associative function of names.

The result is four general categories of data fields, shown as 2.00 through 5.07 below: name, feature, namer, and meaning. 1.00 has nothing to do with the name until the experts at the U.S. Geological Survey get their hands on it; therefore, this is not information about the name per se. 6.00, by contrast, is a data field where information may be recorded that does not fit neatly into any other data field but would be classifiable in one or more of the four general categories. The four general categories are, in fact, inclusive and comprehensive; they describe all types of information.

As promised at the beginning of this paper, I should like to close with a couple of observations about the logical processes of individuation and identification. As we scan the types of data that are compiled about placenames, it may seem plausible to assert that names could in some sense be distinguished from one another, that is to say individuated, on the basis of many types of data. But let us try, for example to distinguish names on the basis of the individual namer. It is immediately obvious that some people, such as Captain Vancouver, named many places; so in order to distinguish the particular name we are interested in we would have to draw on additional fields of information. As additional fields are used from the general categories of namer and meaning we would still not be able to say for sure that a particular placename is unambiguously individuated. Some namers use the same name more than once. This same logical exercise can be repeated for all fields of data with the same inconclusive results, except when the combination of feature class and location are used with the name. Thus, by the process of elimination we find that combined data about the name, feature class, and location are unique identifiers, and that these fields should be the only "required" data fields. Any data may be useful data, but it is these three types that define our subject matter and field of study.

## Works Cited

- Black, Max. *Models and Metaphors*. Ithaca, NY: Cornell UP, 1962.
- Cassidy, Frederic. *The Place-Names of Dane County, Wisconsin*. Greensboro, NC: The American Dialect Society, 1947.
- Quine, Willard Van Orman. *Word and Object*. Cambridge, MA: MIT P, 1960.
- Searle, John R. *Speech Acts*. London: Cambridge UP, 1969.
- Smith, Grant. "A Progress Report for the Placename Survey of the United States." *Names* 39.4 (1991): 393-94.

## Appendix

The list of data fields which follows should be viewed as illustrative and dynamic. It is what I have designed for use by my research assistants in GNIS Phase II for collecting placenames in Washington and has a few peculiarities. For example, 3.06, "Biblio" is the equivalent of the PLANSUS field "source." My assistants are instructed to use this field for documenting the source of information for the *name*, *feature class*, and first indication of *location*. If the source of information differs with additional research, bibliographic codes are to be inserted within the data fields.

### Data Base Field Descriptions for Washington State Placenames

1.00 IDNO — Use GNIS code. If new, assign an identity number beginning with WA, followed by a three letter code for the county, plus a four digit number. If it is a variant, use the GNIS code, followed by a period, followed by a number.

2.00 Name — The exact spelling of any name used at any time.

2.01 Locus or Direction — That part of the name, if any which designates a direction or spatial location.

2.02 Specific — That part of the name which labels the individual feature.

2.03 Generic — That part of the name, if any, which designates the type of feature named.

2.04 Sound — A phonetic transcription of the local pronunciation.

2.05 Language — the language from which the name is derived; if from a series, give immediate derivation first. Use three letter abbreviations.

2.06 Morph — The spelling divided into its morphological parts by hyphens.

2.07 Gloss — Morphological gloss—i.e., the literal meaning of each morphological part.

2.08 Phrasal — A phrasal translation of the meaning if other than English.

3.00 Feature — Identify the feature class as indicated in the GNIS Phase II instructions.

## 306 Grant Smith

- 3.01 Coordinates — International coordinates (longitude and latitude).
- 3.02 Heads — The source or secondary coordinates.
- 3.03 Map/Chart — Code for Phase II recording purposes.
- 3.04 County — County or counties, largest part first.
- 3.05 Elevation — In feet, if known.
- 3.06 Biblio — Code for bibliographical sources using Phase II procedures. Be sure that a full description of the bibliographical sources exists in the "Lists of Works Cited."
- 3.07 Date 1 — The date of first use, if known, in eight digits, e.g., 11/02/1867.
- 3.08 Date 2 — The date displaced, if applicable.
- 3.09 Section — Section, Range, and Township from the Public Land Survey System.
- 3.10 Nearby — Location in relation to a significant feature nearby.
- 3.11 Adm Res — Administrative responsibility: F = Federal; S = State; C = County; M = Municipal; P = Private; SD = school district; UI = unincorporated.
- 3.12 Size — In English units.

**4.00 Namer** — The name of the agent or whoever was primarily responsible for the naming.

- 4.01 C/I — C = corporate, group, or institutional namer; I = individual namer.
- 4.02 NSta — Namer's status: GovL = local government; GovR = state or regional government; GovN = national government; GovI = international or foreign official; ownr = owner; expl = explorer; vstr = visitor; cmltd = other community leader; cntw = contest winner; tsnt = transient; comc = commercial or industrial enterprise.
- 4.03 Empl — Primary employment, profession, or commercial interest of namer.
- 4.04 Title — Title of namer, if appropriate.
- 4.05 Nat Lang — Namer's native language abbreviated in three letters.
- 4.06 Gender — Namer's gender: M = male, F = female.
- 4.07 Age — Namer's age.
- 4.08 Cit — Namer's citizenship abbreviated in three letters.

**5.00 Meaning** — Prior reference of the record name: FG = feminine given name; FS = feminine surname; MG = male given name; MS = male surname; PH = physical description; BIO = flora or fauna; ACT = activity; COI = coinage; MSC = miscellaneous; UK = unknown.

- 5.01 Name — the full name of this reference if biographical
- 5.02 Empl — Primary employment, profession, or commercial interest of the biographical reference
- 5.03 Title — title of biographical reference, if appropriate
- 5.04 Age — age of biographical reference
- 5.05 Nat — citizenship of reference abbreviated in three letters
- 5.06 Sense — Additional categories of prior reference: TEX = texture; TST = taste; SHP = shape; SIZ = size; CLR = color; SND = sound; DST = distance; DIR = direction; EVL = evaluation.
- 5.07 Assn — S = shift, the name is a borrowing from another nearby feature; T = transfer, the name is a borrowing from a similar feature far away; P = prime, the name is neither a shift nor a transfer but an original application.

**6.00 Comments** — Brief comments to add essential data or clarify given data.