

## **Gazetteers Enriched: A Conceptual Basis for Linking Gazetteers with Other Kinds of Information**

*Ryan Shaw, School of Information and Library Science, University of North Carolina at Chapel Hill*

### **Introduction**

The gazetteer is a tool that has taken on new importance in the context of networked, digital communication. Gazetteers are used to facilitate access to geographic information systems, to integrate data related to places, and to enable text-oriented search of spatial information (or spatially-oriented search of textual information). Many information systems, including the major Web search engines, rely on proprietary or publicly available gazetteers to provide place-oriented search. As tools that support the description, discovery, understanding, and processing of information, gazetteers are exemplars of a broader class of tools known as *knowledge organization systems* (Hodge 2000).

As these tools have moved onto the Web there has been a tendency to de-emphasize the differences among them, in favor of assimilation into a generic “cloud” of linked data. Certainly there is value in recognizing affinities among different knowledge organization tools and designing systems in which they can be used together. But dissolving all of these tools in a “global data space” (Heath and Bizer 2011) may not be the most effective approach. Distinctions among knowledge organization tools are grounded in their different functions. While these distinctions have never been crisp, effacing them completely leads to focusing too much on details of minting identifiers and modeling data, rather than the purposes to which the data will be put.

The purpose of this essay is to explore how one might embed gazetteers in a networked environment without losing sight of what makes them distinctive tools. First I review the general notion of a knowledge organization system. I then argue that the distinctive function of a gazetteer is to *provide descriptions identifying the referents of proper names*—typically places, but in principle any named thing that can be spatially located. Next I consider how to expand this functionality both horizontally, giving the gazetteer greater breadth, and vertically, giving it greater depth. Gazetteers expand their breadth by including names and descriptions of things other than places, such as people, material objects, or historical events. Their depth increases when the descriptions of the objects they describe are enriched with more information about how those descriptions were constructed.

## Knowledge Organization Systems

“Knowledge organization system” (KOS) is an umbrella term for a variety of tools including controlled vocabularies, authority files, synonym rings, taxonomies and classification schemes, thesauri, and ontologies (Tudhope and Koch 2004, Lei Zeng 2008). Reference works such as encyclopedias and bibliographies can also be understood as KOSs (Hjørland 2007). Essentially a KOS is any tool that brings together related concepts and their names in a meaningful way, such that users of the KOS can easily comprehend the relationships represented. Various kinds of KOS differ primarily in how they are structured. In the following paragraphs I briefly describe some of these structural differences; readers interested in learning more should consult Lambe (2007, 4-48).

The simplest structure for a KOS is just a list, for example a list of Irish civil parish names. But a long list can be cumbersome to use, so one typically wants a way to break it into pieces (e.g. parish names grouped by county), each of which can be represented by a name in a more manageable list (of Irish county names). Now we have a tree structure, which we can use to make increasingly broad generalizations up the tree and increasingly narrow distinctions down the tree. We can augment these *broader-than* and *narrower-than* relations by adding relations of *equivalence* (this name and that name actually refer to the same parish) or *association* (these two parishes were originally parts of one larger parish). A KOS that relates names in these three ways is known as a thesaurus, an example of which in the geographical domain is the Getty Thesaurus of Geographic Names (TGN).<sup>1</sup> A thesaurus like the TGN differs from a GIS in that its records generally do not contain coordinate information, because the purpose of the thesaurus (like any KOS) is to understand semantic relationships, not to display information on a map.

The most complex form of KOS is an ontology, which attempts to specify relationships among entities using a formal logic so that new relationships can be algorithmically inferred. Algorithmic inference requires that definitions of relationships be perfectly consistent and unambiguous. Unfortunately, relationships as used and understood by people rarely have these properties. But compromise is possible. The Simple Knowledge Organization System (SKOS) is an ontology specifying the components of KOSs and the relationships they represent (Isaac and Summers 2009). By representing a KOS using SKOS, one is formally modeling the structure of the KOS and not necessarily the full meaning of relationships represented. This can allow some limited algorithmic inferences to be made, without forcing the definitions to be fully consistent and unambiguous. This is useful for making KOSs (such

---

<sup>1</sup> <http://www.getty.edu/research/tools/vocabularies/tgn/>

To appear in *Placing Names: Enriching and Integrating Gazetteers*, Indiana University Press.

as the TGN) usable by software programs, for example by publishing them on the Web as linked data or web services.<sup>2</sup>

### **Gazetteer Web Services**

The Open Geospatial Consortium, in its document outlining best practices for the design of gazetteer web services, identified four typical use cases of such services (Fitzke and Atkinson 2006). One use is *geocoding*: identifying names of places in arbitrary text or recorded speech. This involves not just finding where place names are used, but disambiguating each use by identifying the specific place being referred to. So, for example, a geocoding web service could ideally associate the name “Greenville” with Greenville, North Carolina rather than Greenville, South Carolina, by examining the discursive context in which it was used.

A second use is *navigation*: enabling users to search or browse the names of places as a first step toward accessing information about a place of interest. The assumption here is that a gazetteer can include names familiar to users and thus can act as an entry vocabulary that they can use to specify their interests (Buckland et al. 1999). Including vernacular and domain-specific names is particularly important for this kind of use. For example, a user considering a move to central North Carolina may know only the names of cities in that area, yet much useful information may be labeled with region names such as “Piedmont” or “Triangle”. By linking individual city names to these non-administrative regions, a gazetteer can help users access this information. A related use is *query expansion*, the difference being that in this case the vocabularies being bridged are those used by separate web services rather than a user and a single web service.

A third use is *selection*: finding names and other information about places that meet a certain description. This is simply the inverse of the navigation use case, wherein the user knows something about the places she is interested in, but is unable to easily refer to them by name. For example, she may want to know the names of all the places that she will drive through on a road trip from Mount Pisgah to Cumberland Knob, so that she can make a mixtape of music by artists from those places (Lamere 2012). A gazetteer web service providing information about the locations of places in North Carolina could be combined with information about the route followed by the Blue Ridge Parkway to make this possible.

---

<sup>2</sup> See <http://vocab.getty.edu/> for a SKOS representation of the TGN.

## Theorizing Gazetteers

Looking at the use cases outlined in the previous section, we can discern a common theme: the gazetteer mediates between discourse and the world. In the case of geocoding, a name is used in discourse to pick out some particular place in the world, and the gazetteer provides the background knowledge necessary (the coordinates of an extent of space) for a reader or listener to know which particular place is being picked out. When a gazetteer is used for navigation, it is the user's discourse that is being mediated. The user is using a name to indicate what particular place he is interested in, and the gazetteer is providing to the system the background knowledge necessary for the system to "understand" which place he is indicating. The case of query expansion is similar, except now both the "speaker" (using a name) and the "listener" (trying to understand which place it refers to) are information services. Finally, in the case of selection a user is giving a description, and the gazetteer is providing the background knowledge necessary to know what places meet that description, and how they might be referred to more succinctly using their names.

This way of looking at a gazetteer is broad enough to encompass the kinds of historical and cultural applications discussed by Southall et al. (2011). A historical gazetteer mediates between *historical* discourse and the world. It provides the background knowledge necessary to understand what place is being referred to by a name in a historical text, even when that place has changed significantly or no longer exists, or the name is no longer in use. For example, a historical gazetteer could provide the knowledge necessary to know that the name "Joara" in a Spanish explorer's journal referred to a Native American settlement in the foothills of the Blue Ridge Mountains, near present-day Morganton, North Carolina (Hudson 2005). This background knowledge may be incomplete, and thus the gazetteer's mediation may be imperfect, but the goal is the same.

Any KOS that associates terms with concepts might be said to mediate between discourse and the world. When one encounters an unfamiliar term used in discourse, one can consult a dictionary to acquire the information necessary to understand it: the term's meaning. And depending on one's philosophical disposition, one might take this meaning to be part of the world. So what distinguishes a gazetteer as a distinct kind of KOS is not simply the fact that it mediates between discourse and the world, but the specific way in which it does this. To understand the specific way in which gazetteers mediate between discourse and the world, it is useful to turn to the work of the philosopher P. F. Strawson.

Strawson was interested in how the way we use language to *refer* differs from the way we use language to *describe*. To refer to something is to indicate what it is that one is talking about, while to describe something is to actually say something about it (Strawson 1950). Most of

what we say or write involves both reference (to a subject) and description (via some predicate). If I say, “Carrboro was a mill town,” I use the name “Carrboro” to introduce (refer to) the subject, and the phrase “was a mill town” to introduce a predicate describing that subject. Both reference and description involve using linguistic expressions to introduce terms (subjects or predicates) in this way. And in both cases, this introduction involves identification (Strawson 1959, 181). “Carrboro” identifies a particular *place*, while “mill town” identifies a *general concept or characteristic*. I describe the identified place by attributing the identified characteristic to it (using the word “was”).

Strawson observed that despite this similarity, the linguistic rules for identifying a particular thing one is talking about differ from the linguistic rules for identifying a general concept being attributed to that thing (Strawson 1959, 180–213). When I say to you that “Carrboro was a mill town,” you can understand the general concept of a “mill town” as long as you understand English and know what these words mean. But to understand what I mean by “Carrboro,” knowing English is not enough, or even relevant. To correctly understand “Carrboro,” you must know something about the world. Specifically, you must know what it is that uniquely identifies the town of Carrboro: some set of true statements that describe it and nothing else (Strawson 1959, 183). So, for example, you might understand “Carrboro” if you know that “Carrboro borders Chapel Hill to the west,” as this statement uniquely describes Carrboro. But of course to correctly understand *that* statement, you need to know what is meant by “Chapel Hill,” which requires knowing the set of true statements that uniquely identify the place to which *it* refers.

The example suggests how one might replace expressions referring to particular things (such as Carrboro) with sets of statements that refer to other particular things (such as Chapel Hill). Strawson claims that one can do this recursively, without fear of infinite regress, because eventually one can always reach a set of statements that uniquely describe the particular thing to which one is referring, yet which do not themselves refer to any other particular things. In other words, one can always eventually obtain a description that uniquely identifies the particular thing being referred to and that uses only some combination of general concepts (and, possibly, actually pointing to things) (Strawson 1959, 210). This is possible because we share a common spatiotemporal framework: we are embodied beings existing in space and time. By virtue of this shared framework, there is always the possibility of me uniquely describing some particular thing by locating it in space and time relative to *your* location in space and time (Strawson 1959, 24–26).

Of course, Strawson was not claiming that, when I use the name “Carrboro,” I actually have in mind some specific set of statements that would allow me to uniquely describe its spatiotemporal location to you without referring to any other particular things. However he

To appear in *Placing Names: Enriching and Integrating Gazetteers*, Indiana University Press.

*was* claiming that when I refer to a particular thing using a name like this, I *presuppose* that such statements could be made (Strawson 1959, 183–186). I do not assert these statements; I take them for granted. If I did not take it for granted that I could, if pressed, actually identify the referent of the name I used, then my use of the name would be meaningless. As Strawson put it, identification of a particular thing via the use of its name “rests on” empirical facts in a way that specification of a general concept does not. Understanding the meaning of words used in discourse is not enough to understand references to particular things; one must understand something about the world: the relative locations of things in space and time.

### **The Distinctive Function of a Gazetteer**

Armed with Strawson’s theory, we can now state more clearly the distinctive function of a gazetteer: it supplies identifying descriptions for proper names. Proper names are “*syntactically simple* expressions that refer, or at least purport to refer, to *particular* objects/individuals” (Reimer 2010). Proper names are needed when members of a communicative community frequently need to refer to some thing, and they intend to all be referring to the same thing, and there is no short description known to be known to all that could be used to identify the thing (Strawson 1974, 36). In such a situation there is a need for an expression that is not tied to a specific description of the thing, which can continue to refer to the thing even as it changes, and which can successfully be understood as referring to the same thing even by people who know that thing differently (Strawson 1974, 38). This expression is a proper name. A name may be given to a particular thing through an act of naming, as when parents name their children or a government names a town, or they can simply arise over time in response to the need to identifyingly refer.

When a speaker or writer uses a name to refer to a particular place, she must know something about the world that enables her to identify specifically which place she is referring to. This knowledge consists of a description of the place that uniquely identifies it. By using the name, she presupposes that her audience has such knowledge too. But of course, they may not. If this happens in conversation, the listener can ask the speaker questions in order to elicit the uniquely identifying description necessary to understand the name. But when the name appears in a written text, this is usually not an option. So this is when a gazetteer is needed: given an unfamiliar place name, a gazetteer can supply a uniquely identifying description of the place referred to by that name. This description may not and probably will not be the same one known and presupposed by the original writer. But that is not necessary. All that is necessary is that both the description presupposed by

To appear in *Placing Names: Enriching and Integrating Gazetteers*, Indiana University Press.

the writer and the one supplied by the gazetteer identify the same place (Strawson 1959, 183).

Understanding gazetteers as tools for obtaining uniquely identifying descriptions of things referred to in discourse helps explain a few things about them. First, it explains why many gazetteer standards focus on the provision of geospatial footprints (coordinates for points, lines, or polygons). This is not because gazetteers are simply adjuncts to GIS, but because a global coordinate system is particularly efficient and efficacious way to provide a uniquely identifying description of a place. Recall that the ultimate ground for successful reference to particular things is our shared spatiotemporal framework. A global coordinate system is just a tool for systematically generating descriptions of locations within the spatial component of that framework. (A calendar performs the same function for the temporal component.) These descriptions (e.g. 35°55'14"N 79°5'2"W) can thus replace the contingent, idiosyncratic and possibly cumbersome descriptions (e.g. “the town just to the west of where I went to university”) presupposed by the users of names.

The need to uniquely identify also explains the need for feature types. Two things can exist in the same place at the same time, if they are different kinds of things. Strawson gave the example of a man and his body (1974, 14). In a somewhat similar way, one can distinguish between a populated place and an administrative division that have the same location. Of course it is likely that the establishment of the populated place predated the creation of the administrative division that shares its location. So it is true that it is rare for even things of different types to occupy *precisely* the same extent of space-time. But it is not impossible. And given that it is difficult to establish precise spatiotemporal boundaries for a vague thing like a populated place, it is convenient to use type as a distinguishing factor.

Finally, this problem of vagueness also explains why gazetteers have traditionally been focused on physical landscape features and administrative boundaries. Both of these can be precisely located using geospatial coordinates, although for different reasons. Administrative boundaries can be precisely located because they are established by fiat, while physical features can be precisely located because they are “public objects of perception” (Strawson 1959, 45). Public objects of perception are material things that different people can see and touch and directly locate and agree that they are the same things even when encountered at different times. They are things like mountains and rivers and coastlines. Strawson calls these things “basic particulars” because they are the kinds of particular things we can use to establish a shared spatiotemporal frame of reference even in the absence of a global coordinate system. All of our other locating activity, including the establishment of administrative boundaries and even the establishment of coordinate systems, depends on our ability to publicly perceive and locate these basic particulars.

## **Broad Gazetteers**

Of course gazetteers should not be *limited* to covering basic or easily georeferenced things like physical features or administrative boundaries. But our understanding of where places are is ultimately grounded in our ability to locate physical features, and so it is unsurprising that these features have been a major focus of gazetteer construction. Precisely locatable places are the points in reference to which we can *relatively* locate less precisely locatable or even unlocatable places (Elliott and Gillies 2011). There may not be enough information known about a place to actually enable it to be located. What is necessary is that the description given is sufficient to uniquely identify the place, not that it precisely locates the place. What makes the place eligible for inclusion in a gazetteer is that it is the kind of thing that *could* be located given the right information, whether or not it is actually locatable in practice.

Conceived *broadly* as tools that supply identifying descriptions for proper names, gazetteers can provide information about any object of interest in the world that 1) has a name, and 2) can, at least in theory, be spatiotemporally located. Such objects are a subset of the class of things that Strawson calls particulars; they are the subset of particulars with proper names. People, groups, organizations, and institutions are particulars (they can be located in space and time) and they usually have names. Events—the things that individuals and groups do—are also particulars, although relatively few of them are given proper names. Finally, there are named historical periods such as the Renaissance or the Late Bronze Age. There are many other kinds of spatiotemporally locatable particulars, but lacking proper names most of them are not suitable for inclusion in gazetteers. Periods, events, people individually and in groups, and places: these are the primary subjects of gazetteers conceived broadly.

A broad gazetteer does not provide information about the names of non-particular things such as classes, kinds, relations, numbers, or other general concepts (except to the extent that these names are used in descriptions of particular things). The category of “named particular thing” is broad, but not so broad as the category of “concept” claimed as the domain of KOSs more generally. Typically the things listed in broad gazetteers will be individual persons, groups of people, organizations and institutions, periods and events, and places. But they also may be things that are less easily categorized, like cultures, mentalités, political processes, social conditions, or behavioral patterns. These things can all be localized in space and time, so they are particular things, but they will only be suited to inclusion in a broad gazetteer if they have been given proper names, such as the First Great Awakening or the Great Migration. Names like this refer to large-scale heterogeneous groupings of people and attitudes and processes. They are not directly observable (Strawson

To appear in *Placing Names: Enriching and Integrating Gazetteers*, Indiana University Press.

1959, 44-45) and thus their contours are open to interpretation. To deal properly with these more complex particular things, a gazetteer needs greater depth.

## **Deep Gazetteers**

Rossum and Lavin's (2000) study of definitions of the "Great Plains" region exemplifies what an entry in a "deep" gazetteer might look like. They examined 50 published maps identifying the "Great Plains" by means of a boundary line. Their goal was not to establish a consensus on the boundaries of the region, but to treat each published interpretation equally and observe the range of perspectives taken and judgments made. They observed a wide variety of criteria used to delimit the region, which they classified broadly into "physical" and "cultural" criteria. An example of the former is the eastern edge of the Rocky Mountains, widely but not universally used as criteria for defining the western boundary of the Great Plains. Examples of cultural criteria include the edges of Native American tribal territories or modern state boundaries. These different criteria resulted in widely varying boundary definitions, which nonetheless shared a (relatively small) common "core." Analyzing the changing definitions over time, Rossum and Lavin observed a trend toward more use of cultural characteristics as boundary-defining criteria, and an accompanying growth in variance of the areas enclosed by the defined boundaries.

Rossum and Lavin's example suggests a general form for deep gazetteers. Each entry in a deep gazetteer is associated with a place name that has been given different definitions over time. Each definition is a uniquely identifying description of the named place. These descriptions might come from maps (as in Rossum and Lavin's study), prose descriptions in the text of scholarly publications, or ordinary "shallow" gazetteers. Each description is associated with bibliographic data documenting when, where, and by whom it was published. Because what is being uniquely identified is not just a spatiotemporal location but also a particular definition of a spatiotemporal location, it is important that each uniquely identifying description is "semantic," reflecting the criteria used for the definition. While coordinates might be sufficient for uniquely identifying a location, they are not necessarily sufficient for uniquely identifying a definition of a location, because two different people might use different defining criteria yet settle upon the same spatial footprint. Conversely, two people might use the same criteria (e.g. the edge of the Rocky Mountains) to define a boundary, yet differently translate that boundary into coordinates (since the "edge" of a mountain range is itself subject to interpretation).

Such a gazetteer could be made yet "deeper" by recording not only the published definitions of a place, but all known citations or uses of each definition. This would serve two purposes. First, it would link each gazetteer entry to a wider expanse of literature "downstream" from

the published definitions. This would enable a scholar to move from an appearance of a name such as “Great Plains” in a text (whether or not the name is accompanied by a definition or citation) to a history of uniquely identifying descriptions associated with the term, and then from any one of those descriptions to a bibliography of works ratifying that description. Second, it would provide some basis for comparing the relative influence of various definitions and how it has changed over time. Rather than simply giving each definition equal weight, a deep gazetteer could use citation data to identify “mainstream” or “marginal” definitions.

Strawson observed that “proper names ... owe their referential utility to the complexity or variousness, or both, of their descriptive hinterlands” (1974, 42). Different people typically have different relationships to, and thus differently know the persons, places, events and other particular things about which they communicate. In order to communicate about a thing they need to be able to identifyingly refer to it at different times and have confidence that they are continuing to refer to the same thing. Yet given their different bases of knowledge, there will likely be no one commonly known description of the thing that can uniquely identify it. This is when it is useful to give the thing a proper name, which can serve as a constant point of reference among communicators who know it differently (Strawson 1974, 38). It is the varying yet overlapping sets of background knowledge presupposed by each use of the name—what Strawson called its “descriptive hinterlands”—that give it its referential utility. A deep gazetteer provides a map of these descriptive hinterlands.

### **Implications for the Design of Gazetteer Services**

Gazetteer services can be used for identifying proper names used in text or speech, browsing through or searching over lists of names, finding names related to a given name, and querying for named things by description. There is no reason in principle to treat places differently from other named things that can be located in our spatiotemporal framework, such as periods, events, individual persons and groups of people. This is not an exhaustive list; one can imagine naming and spatiotemporally locating things like specific mentalités (world-views, e.g. “American exceptionalism”) or technologies. (Many period names are defined on the basis of and named after spatiotemporal patterns of technology use, estimated on the basis of surviving artifacts.) Some named things, such as physical landscape features or administrative divisions, might be precisely locatable at a specific time using a global coordinate system. But in most cases it will be necessary to locate named things in space and time by relating them to other named things (as when we partially locate a person in time and space by relating her to her birthplace).

To appear in *Placing Names: Enriching and Integrating Gazetteers*, Indiana University Press.

There are many ways that named things can be related to one another. Different people may know different subsets of the relations that any one thing participates in, and thus have different kinds of knowledge supporting their command of the name. This will be particularly true of things that are “vague ideas” (Paasi 1986, 125) like regions. These kinds of things can be identified as having location and duration, but estimates will vary depending on the specific perspective taken. And because such things are always participating in some historically contingent process of institutionalization (Paasi 1986), over time there may be more or less consensus on the identity of the named thing. A deep gazetteer can trace the history of how a name has been used to communicate. This history can provide some basis for making distinctions between different stages in the production of the name’s symbolic significance, or among the different meanings the name has had for different communicative communities at a given time.

Gazetteers can build depth by gathering all known examples of name use, and documenting what is known about the circumstances of each use (Southall et al. 2011, 141). Of course, compilers of a deep gazetteer may intentionally restrict their scope. One gazetteer may include only uses of names in scholarly publications; another may include any mention of a name on Twitter, while another may include any record tagged with the name in a data repository. Regardless of scope, however, the documentation should include a date and place for each usage (e.g. when and where it was published), and any assertions made about the name such as identifying it with a particular range of space or time (possibly indirectly or approximately), relating it to other names, or associating it with general concepts (such as categories). By documenting attestations, deep gazetteers can provide tools for historicizing and contextualizing the processes through which their names have taken on symbolic significance, and mapping or modeling degrees of consensus and ambiguity at different points in this process (Mostern 2008, Southall et al. 2011).

A deep gazetteer, unlike an ontology, is not intended to specify an unambiguous standardized conceptualization of each thing it describes. Rather it is intended to map the usage of a proper name by different people at different times in different contexts, and the overlaps between the things picked out by these uses. A gazetteer is a tool for recording and discovering common patterns of language use. These patterns can potentially be aggregated at different levels, when it is desirable to characterize what passes for consensus on the referent of name at a particular time or in a particular discourse community. These aggregations have no special status: they reflect regularities in processes of making meaning, but there may be more than one way to characterize those regularities (Shaw 2013). However, the fact that the referents of the names documented in gazetteer are particulars—things locatable in our shared spatiotemporal framework—may mean that regularities in

To appear in *Placing Names: Enriching and Integrating Gazetteers*, Indiana University Press.

their patterns of use are easier to discern than they are for patterns of word use in general (Strawson 1974, x).

Current best practices for publishing Linked Data forces publishers to make decisions about identity, in order to assign URIs to particular things. Typically publishers are also pushed to make decisions about how to categorize these identified things under general concepts. Thinking of gazetteers as tools for understanding (the history of) name usage suggests a change of approach. Rather than trying to pin down whether “Great Plains” is one concept or fifty concepts, I can remain agnostic about the precise outlines of the conceptual divisions, and simply note that the name has been used at various times to make various assertions (e.g. about where the boundaries of the region are). Counting uses of a name is simpler than counting its referents. To answer the question of whether two different authors referred to the “same” thing by their separate uses of the name “Great Plains,” I can make a judgment based on what each use explicitly asserted, and what each use seemed to presuppose, about the referent. If combining all the assertions made or presupposed between the two uses leads to contradiction, I have some evidence to suggest that they might not be referring to the same thing. But there is still no need for me to make a strong denial of identity or to judge whether they are referring to different kinds of thing.

## **Conclusions**

When designing information services it is critical to think clearly about not only the representation of the information but also the function of the service. The distinctive function of a gazetteer service is to supply identifying descriptions for proper names. If we conceive of gazetteers broadly, these may be the proper names not only of places, but also of anything that is possible to spatiotemporally locate. Some things, like periods or regions, can be located only approximately, because spatial and temporal extents are subject to interpretation. Proper names for such things are indispensable, as they allow us to communicate about them without requiring us to agree on precise descriptions of them (Searle 1958). The precise descriptions associated with a proper name can vary over time and from person to person. A deep gazetteer is a tool for tracing these variations.

Current approaches to publishing data on the web are ill suited for providing information services that mediate between discourse and the world. These approaches proceed from the assumption that, given two instances of using a name, we can say definitively whether the two uses refer to the same thing (in which case we can replace the two names with a single identifier) or whether they refer to two different things (in which case we can replace the two names with two different identifiers). This forces us to explicitly identify a thing before we can communicate about it. But this is not how names work in natural discourse, where

To appear in *Placing Names: Enriching and Integrating Gazetteers*, Indiana University Press.

we communicate about things by implicitly presupposing that identification is possible. We ought not abandon the semantics of natural discourse simply to fit a convenient formalization. By focusing our efforts on representing published statements that use names, rather than trying to definitively identify and classify the things referred to by those names, we can build gazetteers with breadth and depth: powerful tools for navigating the complex relationships between what we say and where we are.

## References

- Ankersmit, Frank R. 1983. *Narrative logic: A semantic analysis of the historian's language*. The Hague: M. Nijhoff.
- Buckland, Michael, Aitao Chen, Hui-Min Chen, Youngin Kim, Byron Lam, Ray Larson, Barbara Norgard, Jacek Purat, and Fredric Gey. 1999. Mapping entry vocabulary to unfamiliar metadata vocabularies. *D-Lib Magazine* 5 (1). <http://www.dlib.org/dlib/january99/buckland/01buckland.html>.
- Elliott, Tom, and Sean Gillies. 2011. Pleiades: An un-GIS for ancient geography. In *Digital Humanities Conference 2011*. Stanford. <http://dh2011abstracts.stanford.edu/xtf/view?docId=tei/ab-192.xml>.
- Fitzke, Jens, and Rob Atkinson, ed. 2006. Gazetteer service - Application profile of the Web Feature Service implementation specification. [http://portal.opengeospatial.org/files/?artifact\\_id=15529](http://portal.opengeospatial.org/files/?artifact_id=15529).
- Heath, Tom, and Christian Bizer. 2011. Linked data: evolving the Web into a global data space. *Synthesis Lectures on the Semantic Web: Theory and Technology* 1 (1): 1-136. doi: 10.2200/S00334ED1V01Y201102WBE001.
- Hjørland, Birger. 2007. Semantics and knowledge organization. *Annual Review of Information Science and Technology* 41: 367-405. doi: 10.1002/aris.2007.1440410115.
- Hodge, Gail. 2000. Systems of knowledge organization for digital libraries: Beyond traditional authority files. Washington, DC: Council on Library and Information Resources. <http://www.clir.org/pubs/reports/pub91>.
- Hudson, Charles. 2005. *The Juan Pardo expeditions: Explorations of the Carolinas and Tennessee, 1566-1568*. Tuscaloosa: University of Alabama Press.
- Lambe, Patrick. 2007. *Organising knowledge: Taxonomies, knowledge and organisational effectiveness*. Oxford: Chandos.
- Lamere, Paul. 2012. Roadtrip mixtape. Music Machinery, June 17. <http://musicmachinery.com/2012/06/17/roadtrip-mixtape/>.

To appear in *Placing Names: Enriching and Integrating Gazetteers*, Indiana University Press.

- Lei Zeng, Marcia. 2008. Knowledge organization systems (KOS). *Knowledge Organization* 35 (2-3): 160-182.
- Mostern, Ruth. 2008. Historical gazetteers: An experiential perspective, with examples from Chinese history. *Historical Methods* 41 (1): 39-46. doi: 10.3200/HMTS.41.1.39-64.
- Pardo, Juan. 2003. *Account of Florida, 1566-1568*. Wisconsin Historical Society; Smithsonian Institution Press.
- Paasi, Anssi. 1986. The institutionalization of regions: a theoretical framework for understanding the emergence of regions and the constitution of regional identity. *Fennia* 164 (1): 105-146.
- Reimer, Marga. 2010. Reference. *The Stanford encyclopedia of philosophy*. <http://plato.stanford.edu/archives/spr2010/entries/reference/>.
- Rossum, Sonja, and Stephen Lavin. 2000. Where are the Great Plains? A cartographic analysis. *The Professional Geographer* 52 (3): 543-552. doi: 10.1111/0033-0124.00245.
- Searle, John R. 1958. Proper names. *Mind* LXVII (266): 166-173. doi: 10.1093/mind/LXVII.266.166.
- Shaw, Ryan. 2013. Information organization and the philosophy of history. *Journal of the American Society for Information Science and Technology* 64 (6): 1092-1103. doi: 10.1002/asi.22843.
- Southall, Humphrey, Ruth Mostern, and Merrick Lex Berman. 2011. On historical gazetteers. *International Journal of Humanities and Arts Computing* 5(2): 127-145. doi: 10.3366/ijhac.2011.0028.
- Strawson, P. F. 1950. On referring. *Mind* 59 (235): 320-344. <http://www.jstor.org/stable/2251176>.
- Strawson, P. F. 1959. *Individuals: An essay in descriptive metaphysics*. London: Methuen.
- Strawson, P. F. 1974. *Subject and predicate in logic and grammar*. London: Methuen.
- Tudhope, Douglas, and Traugott Koch. 2004. New applications of knowledge organization systems. *Journal of Digital Information* 4 (4). <http://journals.tdl.org/jodi/index.php/jodi/article/viewArticle/109/108>.