THE SOCIAL CONSTRUCTION OF FEATURE TYPES IN VOLUNTEERED GEOGRAPHIC INFORMATION

Katharine E. Currier\(^1\) (currier@geog.ucsb.edu)
Michael F. Goodchild\(^1\) (good@geog.ucsb.edu)
Linna Li\(^1\) (linna@geog.ucsb.edu)
Indy Hurt\(^1\) (indy@geog.ucsb.edu)
Peter Mooney\(^2\) (peter.mooney@nuim.ie)

1. Department of Geography, University of California, Santa Barbara,
   Santa Barbara, California, USA
2. Department of Computer Science, National University of Ireland Maynooth,
   Maynooth, Co. Kildare, Ireland
THE SOCIAL CONSTRUCTION OF FEATURE TYPES IN VOLUNTEERED GEOGRAPHIC INFORMATION

Abstract

Maps and geographic data sets represent features of the landscape that are significant both to the compilers of such information and to eventual users. Mapmakers are faced with numerous decisions including how geographic features are to be categorized. While traditional methods of geographic data acquisition have emphasized authoritative solutions, the growth of volunteered geographic information (VGI) has introduced new complexities. Cultural and linguistic differences in spelling, terminology and even the fundamental conceptualization of geographic entities, themselves present challenges to amassing a dataset that accommodates rather than obliterates this diversity. This paper examines various approaches that demonstrate the socially constructed nature of feature identification and classification, ranging from the top-down solutions of authorities to the bottom-up or folksonomic solutions of VGI. A potential improvement to the structure of point-of-interest databases is explored whereby spatially binary, rather than unary, data would be collected, illustrating the geographic relationship between a feature and its contributor. This solution is problematic in traditional cartography, but feasible in today's geographic databases, and might help to address problems that can cause friction and on occasion mushroom into international incidents.

Introduction

While map-making has traditionally been the domain of expert surveyors and cartographers, recent years have witnessed the very rapid growth of a new form of geographic data acquisition in which members of the general public act as observers and contributors, aided
by tools such as GPS, fine-resolution imagery, and online mapping software (Goodchild, 2007). The spectacular success of the OpenStreetMap project testifies to the feasibility of this approach and its appeal but raises many issues, including concern about data quality (Goodchild and Li, in press). This paper is concerned with another fundamental issue that has received scant attention to date: the basic ontologies employed and their relationship to those of traditional mapping.

Geographic data collection efforts have traditionally been driven by a central organization seeking to fulfill a broadly based need for data of a particular nature. In directing the data collection effort, the organization specifies (1) which geographic features are of interest, (2) the accuracies of the procedures to be used to measure position, (3) how these features are to be categorized, and (4) how they are to be represented in a data repository. The focus of this paper is on the third of these, the feature taxonomy.

In the United States, the Federal Geographic Data Committee (FGDC) guides the development, use, sharing, and dissemination of geographic data (http://www.fgdc.gov/). This committee has identified seven core themes of critical importance to the nation’s spatial data infrastructure: the cadaster, digital orthoimagery, elevation, geodetic control, governmental unit boundaries, hydrographic features, and transportation networks (FGDC, 2008). In published standards each theme is formalized in a data dictionary, which defines the structural and content requirements that a dataset must satisfy. The standards additionally specify a “minimal level of data content that data producers, consumers and vendors shall use for the description and interchange of those data” (FGDC, 2008:1).

While the FGDC specifies geographic data themes of interest and standards governing data accuracy and storage, naming and classification of features is implemented elsewhere. The US Board on Geographic Names (BGN) is a federal body originally established in 1890 to
standardize geographic naming (Yost, 2011). It sanctions two gazetteers, one for domestic and one for foreign features, as official repositories cataloging the name, location, classification and other attribute information of geographic features. The Geographic Names Information System, developed and maintained by the US Geological Survey, includes domestic (and Antarctic) features. Each feature is classified as one of approximately 65 feature classes (see http://geonames.usgs.gov/pls/gnispublic/?p=gnispq:8:2194733487509219). Foreign (and undersea) features are cataloged in the Geographic Names Server (GNS), maintained by the National Geospatial-Intelligence Agency. The GNS defines and includes approximately 680 feature designations further categorized into nine classes (see http://geonames.nga.mil/ggmagaz/feadesgsearchhtml.asp). These two gazetteers serve as the US federal and national standards for geographic nomenclature, and the BGN holds the authority to approve, with some exceptions, proposals for new geographic names and to arbitrate disputes over existing names (Orth and Payne, 1997).

Crowdsourcing approaches to geographic data collection

In contrast to the top-down process traditionally embraced by government mapping agencies, the crowdsourcing approach relies upon the general public—the crowd—not only to contribute data, but also to define the parameters of data collection. The volunteered geographic information (VGI; Goodchild, 2007) that results derives both its content and, to different extents, its structure from its contributors. In 2004 Thomas Vander Wal (2007) coined the term *folksonomy*, to describe a “user-created bottom-up categorical structure development with an emergent thesaurus.” This phenomenon has been recognized in social media services such as Flickr and Del.icio.us, where *ad hoc* labeling and tagging characterizes resources on the Web.
(Vander Wal, 2007). In VGI services, online data resources are not only tagged by data consumers, they are contributed by many of the same people.

OpenStreetMap (OSM) is considered here as an exemplar of the folksonomic approach to geographic data collection. As an “editable map of the whole world” (http://www.openstreetmap.org), the service depends upon data contributions by Internet users and is global in scope. According to the OSM Beginners Guide, selection of features for inclusion is driven by contributors: “If you feel it helps people find their way, then map it” (http://wiki.openstreetmap.org/wiki/Beginners_Guide_1.1). Clearly focused on navigation, a list of the most commonly mapped feature types is provided in the OSM Wiki (http://wiki.openstreetmap.org/wiki/Map_Features). The Beginners Guide qualifies this statement, however, and explains that contributors are not limited by the list and are encouraged to “be creative” and “map what is important to you” (http://wiki.openstreetmap.org/wiki/Beginners_Guide_1.1).

Feature classification in OSM occurs through a folksonomic process, allowing contributors to classify features through self-defined tags. A tag is a key=value pair used to describe a feature, for example amenity=library (http://wiki.openstreetmap.org/wiki/Tags). Contributors may define tags however they like, though they are encouraged to select from a core set of standard choices. Novel feature types, while allowed to exist in the database, must be approved by other OSM contributors before being adopted as a standard tag (http://wiki.openstreetmap.org/wiki/Map_Features). Approval is achieved through a voting process. Once part of the map, features may be modified freely by other OSM contributors.

A third approach to geographic data collection hybridizes the top-down and folksonomic approaches, relying upon some parameters defined by a central organization yet still built upon
contributions by volunteers. Google Map Maker (GMM), a proprietary service, exemplifies this approach. GMM is a near-global service, allowing contributors to edit features in most countries of the world. The description of feature editing in GMM’s Getting Started Guide is broad, stating that GMM “allows you to add and update geographic information” (http://www.google.com/support/mapmaker/bin/static.py?page=guide.cs&guide=30028). GMM solicits geographic data from volunteers, yet the GMM interface constrains the feature types available for classification. Depending on the geometric nature of the feature representation (point, line, or area), a contributor must select from a list of 8 to over 2,000 feature-type names accepted by the service as of January 2012. The service includes conventional choices such as Restaurant, Lake, and City, but in addition to these choices a contributor may classify and submit geographic data representing professions (e.g., Allergist), activities (e.g., After School Program), conceptual places (e.g., Bird Watching Spot), and general subjects (e.g., Arts and Humanity). Data are approved or denied by other GMM contributors and, in some cases, Google staff (http://www.google.com/support/mapmaker/bin/answer.py?answer=1108786). Approved features may be edited by other contributors, whose edits are subject to the same approval process.

Diverging from traditional methods of collecting geographic data, the folksonomic and hybrid methods have evolved as means to collect VGI. They are free from many of the limitations faced by traditional top-down methods: the need for professional data collectors; the comparatively long time required for a small number of professionals to gather data, and the data’s consequent lack of currency; and the imperative that the data conform to and represent a single, consistent perspective. In many cases, volunteers of geographic data are plentiful and
cheap, contributing without monetary compensation. They herald from a vast pool of Internet users, capable of generating volumes of data that cannot possibly be produced by a limited number of professionals in the same amount of time. The same breadth of volunteers guarantees that data will be contributed by people with a range of world views shaped by different social, cultural, political, linguistic, and other influences. The perspectives embedded in the resulting geographic data have the potential to be as diverse as the people who contributed them. Spelling mistakes, recognized differences of spelling within the same language (e.g., between US and UK English), and variation in common terms (e.g., freeway and motorway, off-ramp and slip road) all help to complicate matters.

VGI is clearly different in many respects from the traditional arrangements for geographic data production. It contrasts a few highly trained and expensive experts with a potentially vast number of untrained, unrewarded amateurs. It contrasts top-down taxonomies with bottom-up folksonomies. Perhaps most important, however, is VGI’s global nature, and the ability of networked citizens to transcend national boundaries. A national mapping agency can impose its standards uniformly within national borders. Where differences exist between countries, for example because of disputes about boundary locations or feature names, the national mapping agency can simply present the national view. Thus the English Channel to the Ordnance Survey of Great Britain becomes La Manche on the maps produced by the French Institut Géographique National; and the maps of boundaries in the Himalayas differ sharply if Google’s maps are accessed from India, from China, or from elsewhere in the world.

**Mapping across boundaries**

The global approach of an OSM or GMM, however, is necessarily much more homogenizing, if it is to aim for a single, global map. Other efforts to create such a geographic
dataset have solicited the participation of national mapping agencies, relying upon their cooperation to adopt standard specifications. The International Map of the World (IMW) project began in 1891 with a proposal at the Fifth International Geographical Congress to create a set of 1:1,000,000 scale maps covering the terrestrial globe (Pearson et al., 2006). A committee of representatives from ten countries was formed to eke out the specifications for this set, a slow process rife with nationalist bickering. By 1913 a set of standards was finally approved, including the decision to use Greenwich as the prime meridian and metric units of linear measurement, and a Central IMW Bureau was established to facilitate interactions between the national mapping agencies. Contributions from France, Great Britain, and the US (before its withdrawal in 1913) dominated the early days of the project, with these nations producing maps covering large stretches of North America, Europe, Africa, and parts of Asia. The project, which was taken over by the United Nations in the early 1950s, continued through much of the twentieth century until a UNESCO report in 1989 concluded that it was no longer feasible (Pearson et al., 2006).

Reasons cited for the project’s failure include insurmountable financial burden, conflicting nationalist interests, concerns over ownership of the data, and lack of a clear objective that the IMW was designed to meet (Rhind, 2000). A second project, Global Map, arose in response to a call for global environmental data issued at the UN Conference on Environment and Development in 1992 (Pearson et al., 2006). In 1996 the International Steering Committee of Global Map (ISCGM) was established and in 1998 the national mapping agencies of all countries of the world were formally invited to participate in the effort. The ISCGM adopted the specifications for Version 1.0 later that year, and by 2000 this version had been completed and was released. The product included eight data layers: vegetation, land cover, land
use, elevation, drainage, population centers, boundaries, and transportation. These layers were
derived in large part from existing digital data, including data from the Advanced Very High
Resolution Radiometer (AVHRR) and the US Defense Mapping Agency’s Digital Chart of the
World. Participating national mapping agencies were charged with using and verifying the data
Corresponding to their own territory, with a tier system established whereby wealthier nations
would assist developing nations to produce their own geographic data. This arrangement sought
to avoid the situation that resulted in nations with larger mapping resources asserting
cartographic authority over nations lacking similarly dedicated resources.

The IMW and Global Map efforts in essence relied upon two levels of standardization:
one specified by the driving multinational organizational body, and the other inherent within
each national mapping agency’s internal process of map production. Today’s crowdsourcing
approaches, including those taken by OSM and GMM, operate only under one such level of
standardization—that imposed by the interface through which data are collected. National
standards are irrelevant, as volunteers in one country can easily contribute data about places in
another country. Cultural and linguistic differences pose challenges to a global approach and
raise the specter of the kind of cultural imperialism that pervaded mapping during the colonial
era (Harley, 1988). Somewhat paradoxically, crowdsourcing may be even more imposing than
national mapping agencies.

**Terminology and semantic issues in VGI**

To what degree is VGI sensitized to cultural and linguistic issues? Accommodating and
representing this diversity lies at the heart of many challenges faced by VGI services. OSM’s
folksonomic solution is to provide free-form text-entry fields: contributors may define tags using
text of their choosing. The system, therefore, can accommodate any description submitted in any
language (within the limits of the Web browser and computer hardware). This flexibility comes at a price, however, as variations in spelling and tag construction may cause unintended extensions to the service’s ontology. Mooney and Corcoran (in press a) found that misspellings account for numerous unique entries in OSM’s database, such as tersiary, medow, and forrest. They also observed tags that had been constructed in ways that contradicted documented conventions, such as by assigning a street’s name to the highway key. Mistakes like these become novel feature types in the OSM ontology, clearly in error.

GMM constrains the initial feature type classification to a list of predefined values, thereby avoiding the problem of spelling variation though at the cost of reduced flexibility. Contributors must fit their data to GMM’s classification system, which is necessarily incomplete in the context of limitless feature-type possibilities. This shortcoming becomes obvious in certain situations; one might wonder, for example, why it is possible to submit a Bowling Alley or a Little League Field but not a Velodrome or Kilikiti Pitch.

Close inspection of the list of feature types available in GMM reveals a bias that could be interpreted as cultural imperialism: is a grocery store in Beijing properly classified as a Grocery Store or an Asian Grocery Store? Geographical convention would suggest the former; the classification of a grocery store as Asian only makes sense if the store is located outside of Asia, where grocery stores are assumed to sell goods of their own culture (i.e., not Asian). It is interesting to note, however, that American Grocery Store is absent from GMM’s list of feature types, doubtless a frustration for anyone wishing to document a store in South Africa that sells American foodstuffs.

Both GMM and OSM face challenges associated with the semantics of feature-type names and inconsistencies in their use. Many features can be reasonably described by more than
one term, especially across cultures, leading to lexical diversity that does not necessarily correspond to semantic diversity. *Pitch* and *field* may both refer to an outdoor area where sports are played, but in GMM and OSM a text search using only one of the terms will fail to return results containing the other. This issue, known as *synonymy*, could be addressed by employing automatic query expansion methods, for example using a thesaurus or controlled vocabulary to broaden the search (Manning et al., 2008). GMM would seem to have an advantage in this instance by controlling the terminology available for classification. The service does not appear to have completely achieved systematization, however, as ambiguities still exist: what distinguishes, for example, an *Airstrip* from an *Airstrip/Runway*, both of which are available choices for feature types in GMM?

The reverse semantic problem may arise, as well, when a single term is used by different contributors to mean different things. Mülligann et al. (2011) discuss this problem in the context of partonomy, where the feature type *school* might be differentially applied to represent a school complex as well as individual school buildings. Regional linguistic differences, too, can confound attempts to discern meaning behind assigned terminology. In some English-speaking regions the term *school* indicates K-12 institutions while in others the term is applied to colleges and universities, as well.

A deeper issue arises from the different ways in which individuals of different cultures conceptualize geographic features. In the continuous-field conceptualization (Longley et al., 2005), the Earth’s surface is characterized by a set of variables and classes that can be measured at any location; examples include atmospheric temperature, surface albedo, soil type, and land use. In this view discrete objects are identified through cognitive activity, and it follows therefore
that the specific objects identified by an individual will be affected by culture, language, and many other factors.

Mark and Turk (2003) discuss how geographic features are classified among the Yindjibarndi people of northwestern Australia. They note that the decisive factor in water-body classification is not size, as is typical in English-speaking cultures, but rather permanence. A *yinda* may be a large lake or small pool, but it must be permanent in order to qualify as such. Differences in geographic feature classification are apparent even within Western cultures as well. Mark (1993) points out that translations of geographic feature terms are frequently approximate, even across Western languages. In a linguistic analysis of a digital information-exchange standard, Mark (1993) found variations in the way terms would be applied to geographic entities in English, French, and Spanish. Terms in the different languages, he found, describe geographic feature categories that are delineated differently within each linguistic group. While overlap was common, a direct one-to-one correspondence between terms in the different languages rarely existed. Even more problematic is the case where a concept underlying a geographic term is specific to a cultural group, as in the concept of *ejido* found in Mexico. While possible to cite a definition, for example “village lands communally held in the traditional Indian system of land tenure that combines communal ownership with individual use,” (http://www.britannica.com/EBchecked/topic/181602/ejido) an equivalent term does not exist in English. Clearly, a list of geographic terms translated into different languages, as exists in the GMM interfaces of different countries, may fail to adequately capture the meaning behind the original terms.

While it may be convenient to discount these linguistic and conceptual issues as trivial, their significance occasionally comes to a head. In an analysis of feature-change history in
OSM, Mooney and Corcoran (in press b) noted an apparent ‘tag war’ over the classification of a highway feature, whose value flip-flopped back and forth between Trunk (a road designation) and Construction over a period of years as different contributors asserted their opinions. A more serious dispute arose over the classification and rendering of the city of Jerusalem as a feature in OSM. According to an extended dialogue in the OSM Forum (http://forum.openstreetmap.org/viewtopic.php?id=13178), this feature drew criticism for being rendered in Hebrew as a result of its name key. In a rare move to arbitrate between Israeli and Palestinian contributors, the OSM Data Working Group removed the name key and issued a statement requesting that it not be re-added (http://wiki.openstreetmap.org/wiki/Data_working_group/Disputes).

**Authority: Local vs. vocal**

Questions of authority in mapmaking have gained prominence as volunteers are increasingly relied upon as geographic information sources. Features are incorporated into the maps of OSM and GMM through mechanisms influenced by the contributor base as well as by the moderation process. As contributors drive the content of both services, regions of the world with few contributors are less likely to be mapped as comprehensively as regions with many contributors. When features are mapped in these regions, they may be more likely to come from contributors who are not local but rather have visited as tourists or otherwise gained information about the regions, such as through aerial imagery. OSM’s detailed coverage of Haiti in the immediate aftermath of the 2010 earthquake was contributed almost entirely by volunteers outside Haiti, working from fine-resolution imagery. Mapping, in this case, relied upon a ‘god’s eye’ view (Elwood 2006), requiring visibility and interpretation only from above.
In both OSM and GMM, contributors themselves are part of the moderation process that determines how features are represented in each service. In order for a new feature type to be approved as a standard type within OSM, it is voted on by members of the OSM contributor community (http://wiki.openstreetmap.org/wiki/Map_Features). Actual feature edits submitted to OSM do not require approval before appearing in the map but are subject to editing by other contributors. In GMM, feature edits must be approved by a GMM moderator before they are fully adopted and subsequently published in Google Maps. Google Map Maker moderators include both members of the GMM contributor community who wish to moderate as well as Google staff (http://www.google.com/support/mapmaker/bin/answer.py?answer=1108786). Except in special circumstances, such as the Jerusalem incident mentioned above, mapping authority in both GMM and OSM rests, in large part, with the vocal majority of each contributor community who exercise their ability to moderate.

Moderation by a community of contributors has been successful in the case of many crowdsourcing projects, and has been advocated as one means to improve data quality in VGI (Goodchild and Li, in press). While ostensibly democratic, rule by the majority may not always be the most appropriate modus operandi for evaluating VGI. When the majority is located far from a feature in question, for example, local knowledge may provide a more authentic source of information than that generated by the (foreign) majority (Dobson, in press). A Yindjbarndi person wishing to document a yinda close to her home may be frustrated when this feature type is not approved by other OSM contributors who are unfamiliar with the term. If classified as a feature type Water and further specified as yinda—a possibility in GMM—such a contribution would be slow to gain approval in GMM, too, where the number of Yindjbarndi-speaking
moderators is undoubtedly small. Unfortunately, in many cases a single policy delegating the source of authority may not satisfy all purposes for VGI.

**Collection and ownership of VGI**

Knowing something about *who* contributed a piece of information, and the circumstances surrounding the information’s contribution, is often fundamental to evaluating that information. This kind of metadata is as important as the actual VGI itself, as was made obvious through controversy surrounding two recent events: the Bowman Expeditions of the American Geographical Society (AGS), and ‘slum mapping’ efforts prompted by an agreement between Google and the Operational Satellite Applications Programme (UNOSAT) of the United Nations Institute for Training and Research (UNITAR).

AGS President Jerome E. Dobson described the Bowman Expeditions as a means to gather information about people and regions all over the world in comprehensive, free, and open geographic information systems, while establishing relationships between US and international institutions and people (Dobson, 2009). Each expedition was to include one or more geography professors and several students who would travel to a region in order to conduct place-based research with local universities, studying a topic of the leader’s choice (Herlihy et al., 2008). The México Indígena project, a prototype Bowman Expedition initiated in 2005, collected information through participatory research mapping conducted with indigenous residents of northern Mexico (Herlihy et al., 2008). The project sparked international debate when a statement issued by the community of San Miguel Tiltepec (2009) condemned the project’s leadership for allegedly failing to disclose that the US Foreign Military Studies Office had contributed funding to the project. Multiple critiques ensued from both sides of the US-Mexico border, accusing the project of “geo-piracy” (Valdivieso, 2011) and violating ethical research
practices by gathering intelligence for the military (Bryan, 2010; Melquiades, 2010). Members of the Indigenous Peoples Specialty Group of the Association of American Geographers (AAG) called for an inquiry into the project (Agnew and Harden, 2009), which prompted the AAG Executive Committee to make appropriate revisions to the society’s ethics statement in November 2009 (http://www.aag.org/cs/about_aag/governance/statement_of_professional_ethics).

Much of the controversy over the México Indígina project was not directly related to the actual information collected from indigenous volunteers but rather the intent driving the information’s collection. Similar suspicions have led to criticism of a recent partnership between Google and UNITAR/UNOSAT (UNITAR, 2010). In particular, GMM’s role in collecting and providing information about slums has been criticized (see Maron, 2011). In May 2008 an Expert Group Meeting on Slum Mapping was convened in Enschede, Netherlands as a joint effort of the International Institute for Geo-Information Science and Earth Observation (ITC), the United Nations Human Settlements Programme (UN-HABITAT), and the Center for International Earth Science Information Network (CIESIN) of Columbia University (Sliuzas et al., 2008). Their purpose was to discuss methods for identifying and delineating slum areas. A list of cities identified by the meeting’s participants as priority areas for slum mapping later appeared on a Google website (https://sites.google.com/site/mapyourworldcommunity/events/slum-mapping). The website urged readers to contribute information about these locations through GMM. In 2009 Google and UNITAR/UNOSAT entered into a special agreement that would allow sharing of GMM data to aid in humanitarian relief and disaster response efforts (UNITAR, 2010; UNITAR, 2011).
At around the same time the Humanitarian OSM Team, an initiative to use OSM for humanitarian response (http://wiki.openstreetmap.org/wiki/Humanitarian_OSM_Team), turned their attention to mapping slums. They launched an effort to engage local residents in using OSM to map Kibera, an area of Nairobi widely regarded as Africa’s largest slum (Hagen, 2009). The leader of the Map Kibera project, Mikel Maron, condemned Google for “appropriating the appearance of open data community methodologies, yet maintaining corporate control of what should rightfully be a common resource” (Maron, 2011). His attack referred to Google’s license policy on data collected through GMM, which restricts the data’s use to services that are non-commercial and are not similar to services already offered by Google (https://services.google.com/fb/forms/mapmakerdatadownload/). Maron contrasted this with OSM’s data license, which he maintained is less restrictive and more open (Maron, 2010).

As in the case of the México Indígina project, the circumstances surrounding the collection of information from volunteers was called into question, this time over data licensing policies. Ethical concerns prevailed in both instances, over failure to disclose a VGI collection effort’s funding source in one, and over ownership and use of the VGI in the other.

The idea of ‘slum mapping’ itself, as embodied by OSM and GMM, raises another issue familiar to VGI: what, in fact, is a slum? According to UN-HABITAT (2007a), a “slum household” is

…a group of individuals living under the same roof in an urban area who lack one or more of the following:
1. Durable housing of a permanent nature that protects against extreme climate conditions.
2. Sufficient living space which means not more than three people sharing the same room.
3. Easy access to safe water in sufficient amounts at an affordable price.
4. Access to adequate sanitation in the form of a private or public toilet shared by a reasonable number of people.
5. Security of tenure that prevents forced evictions.
Under this definition, any urban area inhabited by homeless people—who obviously lack the amenity described in the first point—could be considered a slum. In practice, however, ‘slum mapping’ is an activity restricted to certain regions of the world. Absent from the list of regions targeted by GMM are North America, Australia, Europe, and Russia, among others (https://sites.google.com/site/mapyourworldcommunity/events/slum-mapping). Clearly the social construction ‘slum’ holds socio-economic, linguistic, cultural, and geographic connotations.

**Linking social group identity to VGI**

Controversies such as those over the México Indígina project and slum mapping initiatives have arisen as VGI has become a more prevalent resource, and as its collection is increasingly endorsed by humanitarian organizations, government and educational institutions, and private corporations. Fundamental to these controversies is information about the data’s origin, often known in the GIS community as lineage or provenance. A potential solution lies in elaborating the structure of point-of-interest databases, such as those used by OSM and GMM, from spatially unary to spatially binary (Goodchild, 2011). Instead of storing information only about a single location, preserving spatially binary information, or information about two locations, would reveal the geographic relationship between a feature and its contributor. In addition to the traditional name, type, and location of a contributed feature, the identity and location of the social group responsible for each feature’s naming and typing could be stored as well. Technically, this would mean extending the triple <name,type,location> to include a fourth element <where is it called that?>. While GMM and OSM currently store the login name of the contributor associated with each feature edit, this means little unless the login name is attached to a profile or other ancillary information, such as an IP address.
Illustrating the geographic relationship between a feature and its contributor would serve as a first step toward addressing semantic ambiguity, terminological inconsistency, and data-collection transparency issues often associated with VGI. Unique meanings of terms and names of features are often linked to a geographic footprint. Terms and names are not always shared by spatially defined social groups, yet the association is often strong enough to be useful.

Implicit meaning in spatially binary information is only useful if it is made explicit. Understanding that the term *school* is used differently on the east and west coasts of the US is essential if one is to take advantage of knowing a VGI contributor’s location. Other types of spatially binary information, such as the geographic relationship between two features in a point-of-interest database, have been proposed to assist in improving the quality of VGI. Mülligann et al. (2011) discuss analyzing the usage and implicit meaning of feature-type classifications to this end. Common-sense rules-of-thumb, such as knowing that a bar is likely to be located in the vicinity of other places associated with drinking alcohol, eating, and meeting friends, can be used to detect probable errors and suggest likely feature scenarios in VGI. Measures of spatial-semantic similarity have been proposed by Mülligann et al. (2011) to clean VGI, recommend feature type classifications, and suggest areas where a feature might occur.

Information technology, which has the capacity to store and apply vast quantities of rules, can liberate users from the obligation of knowing and remembering such rules. A VGI service that assists a contributor by suggesting related feature-type names could avoid some of the previously described issues of misspellings, redundancy, and inconsistency. The Alexandra Digital Library project, for example, developed a Feature Type Thesaurus in order to provide a bridge across the different typing terminologies used by major national gazetteers (Hill, 2000). Such a reference could be generated (automatically or semi-automatically) for VGI services to
assist contributors as they entered their contribution. Presently, OSM and GMM attempt to address these issues by maintaining lists of conventions for users to follow when classifying features. Essentially data dictionaries, these lists are documented in wikis and online discussion boards. As these lists grow, contributors will be less and less likely to know and remember all of the conventions and would benefit from automated assistance that incorporated them. As another example, if a contributor began to digitize a road crossing another already digitized road, a dialog box might appear asking the contributor to be specific about whether the roads crossed at grade or by means of an overpass, a critical issue in using a road database for routing. The service would therefore assist the contributor in editing the map in a fashion consistent with other contributors. Rather than removing or controlling variation, the technology could be used to accommodate differences and synthesize the resulting body of VGI.

**Conclusion**

The growing popularity and acceptance of VGI as a legitimate source of global information have emphasized questions of validity and data ownership to a degree absent from traditional top-down methods of collecting geographic information. One advantage offered by VGI is the possibility to encompass and represent multiple viewpoints. To achieve this end, services such as OSM and GMM will need to collect geographic information in a way that accommodates the cultural and social diversity of volunteers globally while maintaining transparency in data collection. They will need to address issues of semantic ambiguity, terminological inconsistency, and the differences underlying conceptualizations of geographic features, and indeed to the very existence of those features. Finally, they will need to adopt data models that intimately associate culture and language with volunteered facts, rather than abstracting them to the often-overlooked level of metadata.
The diversity reflected in VGI collected to date emphasizes the fact that features in point-of-interest databases are hardly absolute entities, independent of their origin. In order to achieve a more balanced global view, services such as OSM and GMM may need to adopt systems of moderation that are not simply based on majority rule but that give greater credence to local expertise. Collecting spatially binary data—including the locations of both a feature and its contributor—would provide this context. The ability to display features based on the location of their contributors would not only add interest to the resulting maps, it would acknowledge one connection between a volunteered feature and its origin. Third-party applications exist that allow one to visualize the locations of a particular contributor’s edits to OSM (e.g. http://yosmhm.neis-one.org/). However, until documentation of a contributor’s location becomes standard, the geographic relationship between feature and contributor must often be inferred (Lieberman and Lin, 2009).

Harvesting additional metadata automatically, for example data describing a contributor’s editing history, would lend further context to the data being contributed. Statistics such as the number of edits made, locations of those edits, types of features edited, etc., would be useful not only as attributes of a contributor but directly associated with each feature, as well. While many of these types of statistics are already collected, they are largely inaccessible to the casual map browser. Metadata of this type could be called explicitly to display different perspectives in the resulting map. What parts of the world have been mapped by Spanish contributors? Where do the most prolific contributors reside? Which features are most heavily edited, and by whom? OSM provides access to data that can be used to answer many of these questions using applications such as the OSM Contributor Analysis Tool (Smith and Mooney, to appear). Integrating the
capabilities of such tools into the websites collecting and visualizing VGI, as well, would add depth to the features appearing on the map.

Central to the issues raised in this paper is feature type ontology, which defines what can be mapped and how to label those things. Cultural and linguistic differences suggest that there can be no universal ontology of feature types; it makes little sense for people in Germany to dictate how features are mapped in Australia. As VGI increasingly replaces information collected through traditional means, new ways of achieving synthesis while accommodating its inherent diversity will need to evolve. This challenge is well suited to be met by the tools of information technology.

The past few centuries saw a steady growth in the importance of national mapping agencies, each dedicated to establishing a single top-down view of the world that was intrinsically hostile to alternative views. VGI is but one manifestation of the end of this modernist era, and its replacement by a more nuanced, diverse, and human-centered approach. Ironically, global VGI projects such as OSM and GMM also appear to be moving simultaneously in the opposite direction, in their desire to create uniformity not only at the national level, but globally, cutting across all differences of culture and language. Thus we are left wondering whether the endpoint of this process will be a return to modernism and the failed International Map of the World, or progress to a post-modern diversity that engages more fully with local, human issues.

Acknowledgements

We would like to thank Stephanie Truitt, Bryan Karaffa, Tiffany Armstrong, David Avon, Lisa Berry, and Aleks Hammer-Barulich for their contributions to this work.
References


