Chapter 4
Metadata Squared: Enhancing Its Usability for Volunteered Geographic Information and the GeoWeb

Barbara S. Poore and Eric B. Wolf

> *Arguably, given the rich interactivity of geographic information, usability applies not only to the systems but also to the content of those systems: the structure and portrayal of the data and metadata within them. This is where the usability industry is relatively weak, and therefore where one of the biggest research challenges lies.*
>
> (Davies et al. 2005)

> *In the earlier world dominated by paper maps the body of information described by metadata was a single map, and an intimate association existed between a map's contents and its marginalia. In the digital world, however, the concept of a data set is much more fluid.*
>
> (Goodchild 2007a)

Abstract The Internet has brought many changes to the way geographic information is created and shared. One aspect that has not changed is metadata. Static spatial data quality descriptions were standardized in the mid-1990s and cannot accommodate the current climate of data creation where nonexperts are using mobile phones and other location-based devices on a continuous basis to contribute data to Internet mapping platforms. The usability of standard geospatial metadata is being questioned by academics and neogeographers alike. This chapter analyzes current discussions of metadata to demonstrate how the media shift that is occurring has affected requirements for metadata. Two case studies of metadata use are presented—online sharing of environmental information through a regional spatial data infrastructure in the early 2000s, and new types of metadata that are being used today in OpenStreetMap, a map of the world created entirely by volunteers.

B.S. Poore (✉) • E.B. Wolf
Center of Excellence in GIScience, U.S. Geological Survey, Saint Petersburg, FL, USA
e-mail: bspoore@usgs.gov; ebwolf@usgs.gov

D. Sui et al. (eds.), *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*, DOI 10.1007/978-94-007-4587-2_4,

Changes in metadata requirements are examined for usability, the ease with which metadata supports coproduction of data by communities of users, how metadata enhances findability, and how the relationship between metadata and data has changed. We argue that traditional metadata associated with spatial data infrastructures is inadequate and suggest several research avenues to make this type of metadata more interactive and effective in the GeoWeb.

4.1 Introduction

Geospatial metadata is commonly referred to as data about data. Metadata describes the content, quality, and origins of a geospatial data set. According to the US Federal Geographic Data Committee (FGDC), which pioneered geospatial metadata standards in the 1990s, metadata was critical for the online delivery of data, allowing users to find, understand, and reuse data sets produced by others (FGDC 2000). Metadata allowed organizations to better manage their investments in geospatial data and provide information to online catalogs and clearinghouses (FGDC 2000). The metadata standards were developed when the Internet was in its infancy, but since then, the use of the Internet as a medium of communication and exchange among data producers and users has burgeoned. Metadata has taken on a more vital role in locating and managing the enormous amounts of geospatial data now available in the GeoWeb (Scharl and Tochterman 2007; Tsou 2002).

There is evidence that many who work with geospatial data sets consider metadata inconvenient, complex, and difficult to produce, creating a "metadata bottleneck" (Batcheller et al. 2009; Batcheller 2008; Tsou 2002). Although GIS professionals may acknowledge the importance of metadata, it often falls off the working agenda. The paradox of metadata is that while the costs accrue to the data-producing organization, many of the benefits accrue to the users (National Research Council 2001). Data producers have encouraged their employees to generate metadata by various means: simple fiat—thou shalt write metadata; providing specialized metadata tools; and in some cases by not allowing data to be submitted to a system-absent completed metadata. Despite these efforts, many data sets lack associated metadata (for a recent example, see Hennig et al. 2011).

If metadata is a usability issue for data producers, it is equally so for the end user. In an online discussion in 2010 about geospatial metadata in the GeoWeb (Fee 2010), a commenter points out that the usability of metadata for discovering the contents of a geospatial data file can be critically impacted by something as simple as the title of the data set:

> Improve how the metadata Title is constructed. This sounds so basic, but it's really important over time. Somehow require the user to create a human friendly Title for their data right up front, so that the metadata doesn't default to some cryptic file name. (Haddad 2010)

The producer/user disconnect has been recognized as a key reason that metadata may impact the usability of geospatial data sets (Comber et al. 2008). The Federal Geographic Data Committee was advised to improve the usability of geospatial metadata by structuring data-sharing partnerships to bring data producers and users into closer alignment (National Research Council 2001). But the usability problem does not just consist of the binary of expert producers and nonexpert users; there is a third element, technology (Moore 2010). Moreover, Internet technologies now permit users to become the producers of geospatial data (Coleman et al. 2009; Budhathoki et al. 2008). Experts understand controlled vocabularies and domains, but their solutions do not scale. Users understand local contexts and use cases, and they are more numerous than experts, but they do not necessarily understand expert vocabularies and domains. Machines can process large volumes of data and can be programmed to identify and process structured data, but they are poor at interpretation and contextual meaning.

This chapter examines how new types of metadata, spawned in the technology-mediated shift from the paper world to the online world, might lead to a more usable, interactive model for metadata. This interactive model would result from active negotiation among expert data producers, machines—defined here as structured programs or software—and data users who are empowered by software to produce data for themselves. This new model overturns the traditional view of metadata in which information about the data set is simply conveyed through a transparent communication system from expert to users (Poore and Chrisman 2006).

The impetus for examining the role of metadata in the GeoWeb resulted from a project the USGS undertook in 2010 to test whether volunteered geographic information (Goodchild 2007b) could be incorporated into *The National Map* (www.nationalmap.gov) of the USGS (Wolf et al. 2011). This ongoing project is using the database structure and the editor Potlatch 2, developed by the OpenStreetMap (OSM) community, to collect and manage geospatial data produced by volunteers. OSM is an open-source street map of the world, created and maintained entirely by volunteers (www.openstreetmap.org). To support simultaneous use by many different users and to record a complete history of all edits, the OSM database stores metadata at the level of the node, which many institutional GIS do not.

In researching the technical aspects of how elements of open and crowdsourced projects can be adapted to the needs of spatial data infrastructures (Onsrud 2007), the authors noted a recent uptick in theoretical discussions of metadata in both the formal GIScience literature and informal discussions online in what we are calling the geoblogosphere. A search of the Web of Science index shows papers on geospatial metadata increased from 1 to 2 per year to an average of 5 or more per year in 2004, perhaps reflecting an increasing interest in ontologies (e.g., Rodriguez et al. 2005). The geoblogosphere is decidedly nonacademic, but it too has been the site of prolonged discussions of metadata. A blog post by Fee (2010) "Let's Save Metadata" attracted attention from both GIS professionals and so-called neogeographers (Turner 2006) who are recasting the role of mapping on the Internet. These recent journal articles and blog discussions form the backdrop against which we evaluate how metadata is being remade in the age of the GeoWeb.

Our thesis is that there are qualitative differences between today's collaborative online mapping projects and the previous generation of multi-institutional data-sharing projects—spatial data infrastructures. We consider geographic information and metadata to be media for communication (McLuhan 1964; Sui and Goodchild 2001, 2011; Sui 2008) and explore the relationship between the new media practices being developed by the community of neogeographers and the efforts of traditional mapping endeavors such as *The National Map* to incorporate citizen contributions.

Media shifts have been profound, and yet metadata practices have not changed much since the mid-1990s. We categorize media changes that metadata might need to undergo in four areas:

- *Usability*—a quality attribute describing how easy it is for the user to interact with a program or piece of software (Nielsen and Loranger 2006).
- *Support for coproduction of data by communities of users*—the recognition that nontraditional users of geospatial data, whether called neogeographers or "citizens as sensors" (Goodchild 2007b), are producing and sharing large quantities of geospatial data online. This is related to changes that have been characterized as Web 2.0 (O'Reilly 2005). This media shift requires metadata to match the scale and dynamism of the current GeoWeb, reflecting the simultaneous edits of large volumes of data and supporting applications enabled by online sensor networks and location-based services (Pultar et al. 2010).
- Shifts in requirements for *findability*—that is, "the degree to which a particular object is easy to discover or locate" (Morville 2005: 4), applicable to both individual objects and systems.
- Altered *relationships between data and metadata*—the idea that in the disorder that is the current Internet, everything is data (Weinberger 2007); metadata and geospatial data are no longer distinguishable.

To examine changing attitudes to metadata, we draw on online sources, interviews, and case studies from information-sharing communities both before and after the so-called Web 2.0 revolution. We consider how metadata enhances the usability of geospatial data and how metadata itself may have to change to accommodate the media shift. The first case comes from a study of information sharing by a number of groups in the Pacific Northwest in the late 1990s as they constructed a regional spatial data infrastructure to help remediate environmental conditions responsible for the decline of native salmon stocks (Poore 2003). The second study is based on a content analysis of online discussions by volunteer neogeographers as they debate metadata and use an open-source mapping platform to map the streets and buildings of Haiti in the aftermath of the 2010 Haitian earthquake. Finally, we suggest alternatives to current metadata paradigms that might bring data producers, data users, and technology together into one interactive system in which users have more input into the production of metadata.

4.2 Background

The Content Standard for Geospatial Metadata (CSDGM) (FGDC 1994) was developed and promoted by the Federal Geographic Data Committee (FGDC) in the mid-1990s on the cusp of the Internet era. It was further adapted and published as a standard while retaining its essential form (International Organization for Standardization 2003). The metadata standard was primarily intended to help large organizations manage their geospatial data holdings; accommodating the end user was a lesser goal. The metadata standard codified common elements to describe geospatial data including data set identification, data quality, data set organization, spatial references, entity and attribute information, distribution constraints, and information about the metadata producer (FGDC 2000). In media transitions, new media often borrow and repurpose the forms of old media. Because the CSDGM straddled the paper and the digital eras, it cobbled together elements of two earlier media forms, the library card and the map collar.

4.2.1 The Library Model of Metadata

The card in the card catalog of a traditional library contained metadata about a physical object—the book—but the card was just a pointer to the book. It said little about the content of the book. Once the reader had located the book on the shelf, she or he had no further need for the metadata—the book *was* the content. The CSDGM and its descendants were built using a structured language based on the Standard Generalized Markup Language (International Organization for Standardization 1986), adapted from the library community (FGDC 2006; Goodchild et al. 2007). In the metadata standard, information about how the data set should be described was rigidly specified by logically constrained production rules that identified each permitted element (or field such as “keyword”), how the elements fit together, which elements were compound, which could repeat, which were required, and what expressions were permitted within each element.

This structure allowed metadata records to be parsed by machine. Metadata describing data sets housed on distributed servers were indexed and stored in a centralized registry or digital card catalog (FGDC 2006) called a clearinghouse or portal. The contents of these spatial data portals were frequently not exposed to the Internet at large to be passively crawled by spiders and indexed for full-text searching by text engines. Rather, spatial data portals became specialized for geographic information and relied on the logical structure of the metadata to facilitate precise searches. For example, a user could specify a land-use data set in Florida from 2009 using the keywords from the metadata record and receive just the data set she or he required without having to wade through a million documents from a Google search on the open Web.

A big problem for the library model is the constant need for updating and maintaining the catalog (Li et al. 2010). Furthermore, producing metadata that will meet the requirements of a complex and highly structured standard with 334 elements is hard. The “added rigor” of adhering to the production rules exacts a huge price in human labor (Shirky 2005a). Due to the persistent legacy of the library model and the complexity of the standard, metadata has typically been managed quite separately from geospatial databases. This has led to an expanding role for metadata managers or curators who extract metadata from the actual data producers and resolve data integration issues (Millerand and Bowker 2009; Schuurman 2009). This tends to distance the metadata from both the data producer and the user.

4.2.2 The Map Model of Metadata

Having found the record of a potential data set, the user will want further information about the quality of the data it describes as well as information on how to obtain it. This is where the CSGDM departs from the library model. The metadata points to the data, but once one has the metadata, one still does not have the data. Maps and the geospatial data derived from them do not yield everything a user needs to know, the former due to its status as an image, the latter because it is expressed in machine language; thus, metadata takes on an explanatory role in addition to its pointing function. Like the map collar or legend, metadata contains information about the contents and quality of the data set.

In the paper world, the map collar provides additional information on the author, location, map scale, the subject, the symbols used, etc. But even the most detailed legend cannot adequately "explain" a map (Wood et al. 2010). Most map legends obliterate the traces of the work practices that went into making the map, compressing the map information into numbers and other symbols (Latour 1999). Strictly in terms of processing data, metadata is a narrative form (first we did this, then that). In theory it can be expansive, describing work processes in detail, but in practice, like the map collar, it often fails to adequately explain the genesis of the data to an outside user. In standard FGDC metadata, narrative explanations of work practices are chopped into data elements and separated from the geospatial database as surely as though the map collar had been cut away. The severing of metadata from the data can lead to user confusion in the GeoWeb.

The library catalog as a finding aid depends on the separation of metadata from the data, but this separation induces new usability problems for the user as well as rendering the metadata incapable of reflecting the rapid change to database transactions in real time. To suit the media changes that have accompanied the GeoWeb, metadata must become interactive and embedded directly with the data, reflecting changes from data producers and users alike.

4.2.3 Interactive, Embedded Metadata in the Digital Age

Amazon.com provides a model for how metadata operates in the online world and demonstrates the three-cornered relationship of producer, user-producer, and technology. A search for a book on Amazon will result in a virtual page that contains descriptive metadata similar to that in a traditional library card catalog—typically title, author, publisher, and publication date. This descriptive metadata points not to the physical copy of the book but to information about how to purchase a physical or digital copy of the book, much as in the FGDC model. In addition, Amazon supplies professional reviews of the book, suggests additional books the user might like based on their past purchasing behavior and the behavior of other users, and allows the user to easily save a link to information about the book to a "wish list." By linking user behavior—search terms and purchasing decisions—directly to the book page, Amazon is using transactional user-centered metadata to enhance usability of the site. In addition, Amazon has supplied a means for users to interact with each other around books by having user reviews and user lists linked to the page. This ever-expanding universe of explanatory and transactional information, much of it generated by the site's users, is metadata, even though it is not formalized or authoritative. This miscellaneous explanatory information is networked on top of a unified platform that supports multiple simultaneous edits.

Amazon behaves somewhat like a physical library in that the page about the book (metadata) points to the location where the book can be found. But the book has, in many cases, also been digitized. As the instance of a physical book, the digital book can also be called metadata because it is not the "real" book in much the same way that the map is not the territory (Korzybski 1933: 58). There is no longer a distinction between data and metadata. As Goodchild (2007a, above) noted, not only is the concept of a data set fluid, so is the concept of metadata. In fact, metadata *is* spatial data (Chrisman 1994). Search provides a way into the information about the book, but the other metadata is accessed through links. Thus in the user's experience, Amazon avails itself of the model of the Internet. The way Amazon deploys metadata as a rich context of explanatory information follows the new media model described above. Metadata about the book enhances usability, is produced by communities of users, and enhances findability of related information.

4.3 Formal and Informal Discussions of Metadata

Two seemingly opposed themes—that geospatial metadata is not simple enough but at the same time not complex enough—appear in present-day metadata discussions. The former is most apparent in the neogeography community; the latter among academics. These are essentially usability issues, though they are different in kind. As Goodchild notes (2007a, quoted above), the digital age has rendered distinctions between data and metadata more slippery. In distributed online mapping systems, usability is a more complex problem than just making a map interface easier and more intuitive to use. Usability must start with the data and their metadata (see Davies et al. 2005, quoted above).

4.3.1 "Let's Save Metadata": Neogeographers

A challenge "Let's save metadata" was recently posted on a geoblog (Fee 2010). Fee's main complaint is the lack of usability in the FGDC standard, and he cites the human/machine dichotomy. Producing metadata following the standard and reading the typical metadata record are hard. Servers can use XML to talk to each other. "But servers rarely read and write metadata on their own without human interaction. Thus the reality of the situation is we poor humans have to ingest and parse metadata regularly<XML>YIKES<?XML>" (Fee 2010). An example of this is the XML

```
<?xml version="1.0" encoding="ISO - 8859 - 1"?>
< ! DOCTYPE metadata SYSTEM "http : // www . fgdc. gov/ metadata/ fgdc - std -
001 - 1998 . dtd">
<metadata>
    <idinfo>
        <citation>
            <citeinfo>
                <origin>University of Florida GeoPlan
Center</origin>
                <pubdate>20101220</pubdate>
                <title>GENERALIZED LAND USE DERIVED
FROM 2010 PARCELS - FLORIDA DOT DISTRICT 7</title>
                <geoform>vector digital data</geoform>
                <pubinfo>
                    <pubplace>Gainesville,
FL</pubplace>
                    <publish>University of Florida
GeoPlan Center</publish>
                </pubinfo>
                <othercit>FDOT District 7</othercit>
                <onlink>http ://www.fgdl.org</ onlink>
                <lworkcit>
                    <citeinfo>
                        <othercit>Source -
2010 Automated - 2010</othercit>
                    </ citeinfo>
                </ lworkcit>
                < ftname
Sync="TRUE">ETAT.D7_LU_GEN_2010</ftname></ citeinfo>
        </ citation>
```

Fig. 4.1 Snippet of XML code of the metadata for a land-use map of Florida downloaded from the University of Florida GeoPlan Center

expression of a land-use data set downloaded from the Florida GeoPlan Center (Fig. 4.1). In practice, this format is rarely encountered—more frequently the metadata is rendered by the machine in a more common indented format (Fig. 4.2), but even this more approachable format demands that the reader do most of the work of decoding the questions she or he wants to ask about the data set.

What matters to users are answers to the who, what, when, where, how and why questions, but “those questions are hard to parse out of metadata” (Fee 2010). This discussion about the usability of metadata and what users really want essentially rehashes those of the mid-1990s when the metadata standard was proposed, indicating that there are unresolved usability issues that have persisted for nearly two decades (Schweitzer 1998).

Professional GIS software such as ArcGIS automates the production of metadata to some extent, although Fee rightly observes that current GIS software could do a better job at this. Furthermore, many of the people commenting on this blog online do not use professional GIS software. Many of the 72 responses can be identified with the tenets of neogeography, although there were also some responses from

GENERALIZED LAND USE DERIVED FROM 2010 PARCELS -FLORIDA DOT DISTRICT 7

Metadata also available as

Metadata:

- Identification_Information
- Data_Quality_Information
- Spatial_Data_Organization_Information
- Spatial_Reference_Information
- Entity_and_Attribute_Information
- Distribution_Information
- Metadata_Reference_Information

Identification_Information:

Citation:
Citation_Information:
Originator: University of Florida GeoPlan Center
Publication_Date: 20101220
Title:
GENERALIZED LAND USE DERIVED FROM 2010 PARCELS -FLORIDA DOT DISTRICT 7
Geospatial_Data_Presentation_Form: vector digital data
Publication_Information:

Publication_Place: Gainesville, FL
Publisher: University of Florida GeoPlan Center

Other_Citation_Details: FDOT District 7
Online_Linkage:<http://www.fgdl.org>
Larger_Work_Citation:

Citation_Information:

Other_Citation_Details: Source - 2010 Automated - 2010

Fig. 4.2 Metadata shown in Fig. 4.2 in indented format

individuals who could be called professional GIS users from the academia, the industry, and the government.

There was a general agreement that metadata needed to be simpler to produce and make. One remark summed up many of the comments. “Unless a caveman can do it, users won’t read or write meaningful metadata. And relevant metadata must be stored and travel with the data” (Entchev 2010).

This last point—the necessity to store and transmit metadata with the data—elicited a good deal of discussion on the geoblog. In our comparison of two case studies, we show the difference between the standard approach to metadata that has been developed by the professional GIS community in spatial data infrastructures, and the approach of the open-source mapping community which is experimenting with data structures that store metadata at the level of the individual data object.

In an era of too much information, findability becomes preeminent (Morville 2005). Google has accustomed us to the idea that simple keyword searching (land use, Florida) should be all that is necessary for the user to search for data, and yet these fields are buried in an overly complex metadata structure (Gould 2006a).

In fact, if one does a Google search for "land use, Florida", one comes up with several FGDC metadata records near the top of the search, based most likely on the "keyword" field. So the original theory of structured metadata was correct. It can be useful for findability. The problem in this case lies in the uneven adoption of the standard and the issue of not exposing more metadata records to the Internet at large. The need for compliance with a top-down mandate may also have doomed metadata's potential. Alternative bottom-up, user-generated taxonomies or folksonomies (Vander Wal 2007) may work better (Gould 2006b). Capturing and exploiting user tags, which are the locally generated equivalent of the metadata standard's keyword fields, could potentially produce an emergent ontology that would aid findability. This issue is discussed further below using the example of OpenStreetMap.

4.3.2 Metadata and Meaning: GIScience

There has also been an increase in academic papers on metadata since 2005. In his review of the adoption and spread of metadata standards since the mid-1990s, Goodchild (2007a) calls for user-centric rather than producer-centric metadata, emphasizing easy-to-understand measures of data quality and tools to assess fitness for users' unique purposes. This is in concert with the discussions in the neogeography community.

On the other hand, several researchers have moved in a different direction, calling for more complexity, either a different kind of metadata or further metadata extensions. These arguments largely center on the idea that metadata, as currently structured, does a poor job of capturing differing meanings, or semantics embodied in a database (Comber et al. 2008). Schuurman and Leszczynski (2006) have proposed additional formal metadata categories for data semantics to assist data interoperability, achieved through database ethnographies (Schuurman 2008). Although machine understanding of semantics, as embedded in metadata, might be desirable, extending the metadata standard will complicate an already complex structure. Gahegan et al. (2009), in their work on community-based knowledge in cyberinfrastructures, caution that ontologies alone cannot capture meaning because they ignore "use-cases, provenance data, social networks and workflows."

4.4 Metadata Top Down

Traditional metadata associated with spatial data infrastructures can be examined according to the four criteria set out at the beginning of this chapter: Is it easy to use? Does it reflect coproduction by a community of users? Does it enhance findability? What is the relationship between data and metadata?

In the late 1990s, federal and state agencies in the Pacific Northwest of the USA were building a shared, multi-organization regional data set of rivers and streams to assist recovery planning for the 22 species of salmon that had been listed as endangered or threatened in 1999 (US Department of Commerce 1999). Over 40 organizations participated in the development of a common data model for hydrography (water) data and built an online clearinghouse fashioned after the FGDC model (Poore 2003).

4.4.1 Usability

The usability of traditional metadata is affected by the compression of work practices, the modular structure of the metadata, and the separation of the metadata from the data. Metadata aims to describe the products (data) of work practices (data analysis and data production) that are mediated by technologies. These practices emerge from situated learning when communities work together on particular problems (Lave and Wenger 1991). Situated knowledge, being primarily tacit knowledge developed over the course of a project, is difficult to translate for other communities. An example of what is lost when working procedures are compressed into metadata can be seen in the records of stream databases that were kept in the Pacific Northwest Hydrography project.

A metadata record of the stream layer from the Six Rivers National Forest (1999), which covers a large area of federal land in Northern California, demonstrates the problem faced by these regional integrators. The process of producing the stream layer for the forest is described in the metadata, but the work details are of necessity condensed. The early history of this data is omitted. The ultimate source of much of the digital data on streams in the Pacific Northwest was digital data derived from the 1:24,000-scale USGS topographic maps in the early 1990s. These data, known as digital line graphs (DLGs), were generalized to 1:100,000 scale and shared with the US Forest Service and other agencies. But the data were inadequate for watershed level work; the generalizing process omitted much detail on intermittent and smaller streams, and the maps from which the data were derived were out of date.

To be useful at a local scale within the Six Rivers National Forest, the stream network had to be densified—adding back in the stream information that had been removed when the DLGs were created. The Forest relied on a then-current densification process called crenulation to delineate streams that were not included in the USGS DLGs. The metadata refers to this process and includes a reference to existing practice (Maxwell et al. 1995). Crenulation is a process of inferring the course of stream channels by tracing the folds or crenulations down a slope on a contour map. This process can be traced back to the scientific literature of the 1930s on geomorphology of stream channels. Thus, a long history of scientific discovery and insight is compressed and translated into the one word, "crenulation," that appears in the metadata record. This process of compression and translation is characteristic of the circulation of scientific knowledge (Latour 1999), but in order to understand the how the data were created, the user must dig into various scattered sources.

Eventually, the original Six Rivers water data set was integrated into a larger data set composed of all the stream layers in the Forest Service's Pacific Southwest Region (US Department of Commerce 2004). This metadata record shows further compression of data techniques and origins. A newer software-based modeling technique, based on flow accumulation, was used to densify the streams in several of the watersheds. The older ephemeral streams from the Six Rivers National Forest that had been densified by the hand crenulation method were discarded. The link backward to the Maxwell method was severed. Severed as well are the work practices of a previous community that drew on a long tradition of local observation and scientific knowledge. Does the end user of the integrated data set need to know the back story of crenulation? Perhaps not, depending on the use to which the data set is put, but the new metadata emphatically warns that there is not necessarily a link between what the data set portrays and what was directly observed in the watershed:

> IT MUST BE CLEARLY UNDERSTOOD THAT THIS DATA SET, AT THIS TIME, IS NOT INTENDED BY, NOR IS CAPABLE OF, DISPLAYING WHERE WATER IS ACTUALLY FLOWING ON THE LANDSCAPE.

4.4.2 Community

It is not so easy to reconstruct the community that built these data sets of the streams. This community is not reflected in the metadata. Nor is there any way to recover the specific history of this data set. The "Time Period of Content"—a field in the metadata—only reflects the time at which the data set was produced. It is static and does not reflect this long history of where the data or the production techniques originated.

4.4.3 Findability and the Separation of Metadata from Data

As to findability, the results are mixed. We could not locate the original Six Rivers National Forest data and the Maxwell reference through the National Forest website, but we were able to find the integrated Southwest Division metadata record discussed above. After much searching through Google and other sites, we found an Esri geodatabase through the National Hydrography Dataset (www.nhd.gov). The reference to the Maxwell process was preserved several layers deep in the geodatabase and without citation of the relevant literature. Findability and usability suffer when the history of the data is hard to reconstruct through the reorganization of the software packages and processes by which the data have been conveyed to the user.

4.4.4 Metadata Bottom Up or Metadata Squared

Digital media have spawned new practices for categorizing data. Classification practices that were adequate for the physical world, in which each unique object had a unique place, have of necessity changed (Shirky 2005a). In the digital world, an object can be in many places at once and can be "about" many different things at the same time, leading to a proliferation of information. This proliferation demands networked user-generated classifications, a bottom-up ontology, a "new order of order" (Weinberger 2007).

4.4.5 OpenStreetMap

OpenStreetMap (OSM) (http://www.openstreetmap.org) was started by Steve Coast in 2004 (Wikipedia 2010). The goal of OSM is to make an all-volunteer online map of the world that will be free of use restrictions and open to all (http://wiki.openstreetmap.org/wiki/FAQ). Anyone can edit the map, discuss the map, create tutorials and other explanatory material, freely access the data, and influence the future direction of the map. OSM volunteers are attracted to the project through engagement with the online data and mailing lists. Frequent mapping parties, held throughout the world, solidify the community by adding the face-to-face experience. A distinctive feature of the wiki software that underlies OSM is a history of all changes to the map over time. OSM can be evaluated according to the same criteria used for evaluating traditional metadata: usability, coproduction by community, findability, and an altered relationship between data and metadata.

In OSM, there is no longer any separation between map and collar, that is, the data and the metadata. This is a new type of mapping medium. The map becomes a platform or canvas on which the user is invited to draw, that is, to edit the map. Users are motivated to contribute to the map for various reasons (Budhathoki 2010), including the strong user community, but a desire to assert creativity is important (Budhathoki 2010). The system supports almost instantaneous updates, validating the mapping platform as a creative endeavor.

The data structure also gives the OSM community the ability to respond quickly to emergency situations in which better maps are needed. Volunteers from across the globe began mapping the street network in Haiti within hours of the earthquake in January 2010. Announcements on the OSM wiki, mailing lists, and social networking spread the word about the need for mappers. Face-to-face crisis camps brought together at least 700 mappers in cities around the world to map (Waters 2010). High-quality satellite images released for public use by GeoEye and DigitalGlobe were the primary vehicles, along with old CIA maps from the 1940s, that volunteers used to trace streets and buildings in the damaged areas (Silver 2010; Maron 2010). The resulting maps were used by many organizations in the response and recovery (Ball 2010). The community organized and coordinated itself entirely using the ancillary metadata that surrounds the map.

4.4.6 Metadata Types

There are two different types of metadata preserved by the OSM platform: object-level metadata and ancillary metadata; there is no overarching metadata document as in traditional geospatial data. Object-level metadata is incorporated directly into the data structure, making no distinction between data and metadata (Weinberger 2007.) The data is very simple and expressed in XML. Data elements or data primitives are nodes (a point expressed in latitude and longitude), ways (ordered interconnection of nodes), and relations (sets of nodes or ways) (http://wiki.openstreetmap.org/wiki/Data_Primitives). The metadata identifies the node, its coordinates, the user who created the node, the editing session (changeset) of which it is a part, the version number of the edit, and the date and time it was edited. Elements can have any number of user-generated tags, consisting of a key and a value. Below is an example of OSM metadata drawn form a user-created map of Port-au-Prince, Haiti, in the aftermath of the earthquake.

<node id="613826766" lat="18.5450619" lon="-72.3305089" user="samlarsen1" uid="5974" visible="true" version="2" changeset="3636891" timestamp="2010-01-16 T23:34:34Z">

<tag k="building" v="collapsed"/>

<tag k="source" v="GeoEye"/>

</node>

In this case, the node is identified as a building that is collapsed (e.g., tag k="building", v="collapsed"). The source is from a particular user, samlarsen1, who digitized the node based on the GeoEye imagery of Haiti.

Because OSM changes continually over time, sometimes quite rapidly as in the Haitian earthquake crisis, object-level metadata is necessary. Object-level metadata facilitates the communal character of map production. Changes to the map can be tracked, and the history can be "rolled back" to a previous state easily if an error is detected by another editor.

In addition to the metadata that resides directly in the data structure, there is a vast, swirling universe of ancillary data describing the map, explaining how to use it, and facilitating community discussion and debate. These data are quite diverse, consisting of computer programs, tiling schemes, IRC chat rooms, YouTube video demonstrations, tweets, e-mail discussion groups, and wiki pages. All of these are socially mediated, produced by the community, and accessible to any user. They are not unlike the Maxwell manual described in the hydrography example above—not formal metadata but associated documents that can amplify the meaning of the data by describing mapping practices. In the format of traditional metadata, these linkages are often lost. But in the GeoWeb, there is no extra expense in linking to them. These ancillary data are much like the cloud of information surrounding digital objects on Amazon.com. We refer to this as metadata squared due to the possibility of endless proliferation.

The community of users makes final decisions in OSM, unlike Wikipedia, where edits are semi-anonymous and controversial topics are supervised by a group of editors. In OSM, one must have an account to edit the map, making the user not only identifiable but also accountable to the community. Trust is placed in the adage, common among open-source computer programmers, that "given enough eyeballs, all bugs are shallow" (Raymond 1999). Questions about the map can be posted on the wiki help pages. Users vote for their favorite question and badges are awarded for participation in answering. This gamelike feature serves to build community.

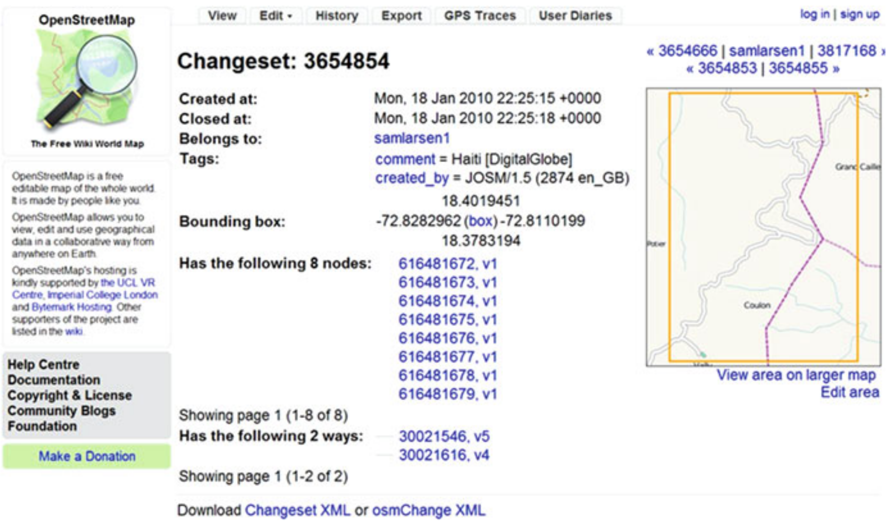

Fig. 4.3 Port OpenStreetMap changeset 3654854, screenshot from www.openstreetmap.org

One can search for a particular user in the wiki. This redirects one back to the user's page in the map interface. The page for the user "samlarsen1," who edited the node discussed above, lists the areas that he has mapped and links to a history of his edits. One editing session (or changeset) took place on January 18, 2010 (Fig. 4.3), when Larsen mapped a road southwest of Port-au-Prince near Grand Goâve. This page is undeniably metadata. It gives the geographic location of the nodes the user has contributed, the editing software used (JOSM), the imagery from which the mapping was derived (Digital Eye), a list of the nodes, and the roads (ways) that the nodes contributed to.

Each of these nodes and ways has its own dynamically generated page with a graphic designed to facilitate human exploration of the data. For example, on the main page for changeset 3654854, the user can click on the graphic, bringing him or her back to the full OSM map so he or she can see the geographic context in which the nodes and ways fit.

This close coupling between wiki, map, and user information and the deployment of different media makes for a rich understanding of the data set. In addition, the user has the ability to download this changeset or any number of other changesets that might have been produced by "samlarsen1" or other users. Users can manipulate this changeset data and its associated metadata in many different ways, by user, by tag, or by geographic area. Examples of relevant code are given in the wiki. In addition, programmers who work on OSM have begun to construct various tools to manipulate user tags to provide interesting ways to visualize OSM data. For example, Tagwatch aggregates user tags three times a week and provides statistical information on which tags are being used by the community (http://tagwatch.stoecker.eu/).

A tag is a key = value pair that can be attached to the nodes, ways, relations and even changesets (http://wiki.openstreetmap.org/wiki/Tags). Tags serve a role similar to attributes in a more traditional, relational database model except that they are not constrained by a top-down schema. The content of tags is up to the user. Any tag can be used, as long as it is verifiable, although this is not strictly enforced. If a user cannot find a relevant tag, he or she can propose a new tag which is voted on by the community. This can lead to confusion, as demonstrated by the tags that emerged during the Haiti mapping. Collapsed buildings were identified in a number of different ways, as "earthquake:damage=collapsed_building", "earthquake:damage=collapsed", "building=collapsed", and several misspelled variants. Tags recommended by the community for use in future disasters reflect the most frequent usage.

As OSM matures, tag analysis with such programs as Tagwatch could be used to generate a user-centered, bottom-up ontology (Shirky 2005b). Work in the library community advances the notion that social semantics—relationships between tags generated through a social process such as OSM participation—can capture local meaning best and can be disambiguated and systematized using controlled vocabularies (Qin 2008). In the case of geospatial metadata, OSM tags on structures could be compared to a controlled database such as the USGS's Geographic Names Information System (http://geonames.usgs.gov/domestic/index.html) to generate a gazetteer that would include semantic relations as well as unofficial names for structures that mirrored local customs.

4.4.7 Evaluation

In terms of the criteria for evaluation—usability, coproduction by community, findability, and the altered relationship between data and metadata—the OSM approach to metadata seems to do well at accommodating a community of mappers who are not necessarily professionals and allowing for simultaneous, distributed updating.

The Humanitarian OpenStreetMap Team (HOT) (http://wiki.openstreetmap.org/wiki/Humanitarian_OSM_Team) in the Haitian crisis demonstrated the rapidity with which maps could be made of an area that had not been not previously well mapped, and these maps were widely used by first responders (Osborne 2010). The simplicity of the underlying data structures and the built-in support for distributed communications in OSM provided a platform which could rapidly scale for multiple simultaneous edit sessions during crisis events.

Usability: Fitness for use is one aspect of usability most familiar to professional mapping endeavors. However, there are places, such as Haiti, that are poorly mapped, if they are mapped at all. Any given map of such a place is useful when compared to having no map at all. The OSM data have the added benefit of being free from governmental or corporate license restrictions. Furthermore, crowdsourced maps tend to improve over time. In Europe, OSM data are nearing the positional accuracy of data from the national mapping agencies, many of which restrict the use of their data (Haklay 2010a). OSM data has also proven useful for commercial interests such as MapQuest.

But just because these maps can be useful, especially in areas where there are no maps, does not mean that all aspects of usability for the end user have been carefully considered. Usability refers to the interactive affordances of the system as a whole. How easy is it to find, access, edit, or understand the data? In his analysis of the completeness of two volunteered data sets—the OSM map of Haiti produced by the volunteers of the HOT and a map of Haiti produced by volunteers working within Google Map Maker (GMM)—versus a more official data set produced by the United Nations Stabilization Mission in Haiti (MINUSTAH), Haklay (2010b) found that the official data set was the most complete. Each data set contained features that the others lacked; however, the lack of metadata to explain data semantics, which differ among the maps, would make it difficult for the end user to integrate the three data sets. Even though the OSM and GMM data were produced by volunteers aligned with the neogeography community and the MINUSTAH data were produced by professionals, the OSM and Map Maker data were delivered to the user in a manner similar to traditional GIS products. That is, knowledge of the semantics of the data was left to the end user to interpret. The MINUSTAH data set contained operational geographic information such as road conditions that reflected its use for humanitarian workers in the field. In an online discussion that followed this analysis, discussants generally agreed that in situations where there are specific first-responder needs that are not being met by the metadata that accompanies generic data sets, some intermediary is needed to translate or adapt the data to these needs. Thus, metadata can directly impact the usability of the data sets and the tools.

Some users of the two mapping platforms in Haiti argued for simplicity, claiming that the proliferation of metadata squared made things more difficult:

> I believe that GMM can be a serious competition to OSM if it is simpler to use, easier to learn and thus more inviting to the casual newcomer. With GMM you have one way of mapping a simple item e.g. a bicycle track. Everybody can do it in ten minutes, no questions arise. With OSM you have two major tools, a huge load of tags, a wiki, a forum, several mailing lists, three different answers to the question, pages of contradictory documentation, plenty of old discussions and after working through all this, you realize that the question has not been resolved yet. I can see how many people would prefer the simple way offered by Google.—"Nop" on January 1, 2010 [OSM-talk] Countering Google's propaganda. (http://lists.openstreetmap.org/pipermail/talk/2010-January/046358.html)

Findability: In a paradoxical way, the proliferation of explanatory material, dueling tags, and community discussions that makes OSM useful and usable is the essence of findability in the media shift that has taken place in the GeoWeb. Morville (2005) describes the shift from theories of information retrieval, on which catalog-centered traditional metadata is based, to information browsing or foraging, in which a number of different strategies are used and the information seeker does not proceed by systematic logical steps but by pursuing leads as they emerge (Bates 2002). Information seeking relies on context, the frame of reference, environment, or setting within which information seeking is performed (Courtright 2008). So do usability and usefulness.

Community: The environment created by OSM with its links back and forth between the graphic, the textual, and the interactive differs from the spatial data infrastructure notion of neatly contained and formalized data—the catalog, the data element, and the interface. Context is community and community has a close affinity with gossip and recommendation culture of the current Internet. For people active in OSM, being able to help decide what to map may be of equal importance as the act of mapping. For the neogeographers, being able to shape the direction of the project and the map itself augurs well for a profound shift in media from the old geospatial world.

4.5 Conclusion

It has been suggested by Goodchild (2008) and Schuurman and Leszczynski (2006) that formal metadata standards need to be rethought to become more user-centric. They also propose adding new data elements to metadata for nonspatial attributes. This would afford the user better access to the context beyond technical and geometric elements and convey the tacit information that went into the making of a data set. Goodchild (2008) would augment current metadata about data geometries and lineages to better express data quality in data sets of mixed origins. These are both good suggestions, but we argue that the practices of neogeographers in the GeoWeb have shown that the metadata genie may be already out of the bottle.

The Internet has supported greater and greater interactivity, user collaboration, and the co-creation of geographic data. Metadata generated both automatically and by direct user contribution from descriptive and transactional work practices is closely coupled with the geographic data itself in these new GeoWeb systems. This proliferation of metadata, or metadata squared, facilitates finding, assessing, using, and making geospatial data.

What is the balance between centralized formalized metadata and the freewheeling metadata of an open-source community such as OSM? We argue that the neogeographers and the academics are both right. Standard metadata is not simple enough and at the same time not complex enough to give the context of data creation. These problems largely result from the media shift we have described. Many people find traditional metadata hard to manage, hard to produce, hard to use, and based on an outmoded static model of the way the Internet works. In theory, traditional metadata could be considered superior to the messy pseudo-metadata of OSM in its precise descriptions and consistent terminology; however, the difficulty lies in getting people to conform to the rigorous standards and keeping them updated. Formalized cataloging systems cannot scale at rates that match the growth of information in the GeoWeb.

Systems like OSM provide a lower cost of entry for producing and using data. The metadata is simplified and object-based, allowing for flexibility and rapid development. The ability to freely download both data and metadata supports emergent use. The proliferating ancillary material, the metadata squared, may be difficult to navigate at times, but the internal linking and the ability to switch back and forth rapidly between text and graphic modes make for an environment that can be richly rewarding for discovering meaning. The map grows and gains validity from the instantaneous feedback among users. And this growth can occur at exponential rates not realized by mapping agencies.

The OSM community might become more organized over time. Weber (2004) observes that open-source software projects can become powerful magnets that attract standards. As discussed above, there is talk on the OSM lists of developing a formal ontology of disaster-related terms. Van Exel and Dias (2011) are exploring how analysis of user behavior in OSM can serve as a proxy for trust and authority. Spatial data infrastructures could benefit by looking at the close coupling of the map and its explanatory context and could develop better systems for encouraging user feedback, as the USGS research on volunteered geographic information referred to above may show (Wolf et al. 2011; van Oort et al. 2009).

As one geoblogger put it:

> Traditional concepts like error bounds will fundamentally change because data collection is no longer happening on an annual basis, but will occur persistently from millions of globally distributed sensors. Error will be a fluid concept and not a static measure. Metadata needs to also change to be a fluid concept. The requirement for dedicated GIS metadata librarians with hundreds of metadata elements will not scale. Most importantly I think we need to stop thinking of the crowd as volunteers and amateurs. We should think of them as data collection points. This new reality is going to require innovative concepts around not only leveraging the crowd for data, but also using the crowd to ascertain the veracity of data. The crowd needs to be leveraged to verify and update metadata. (Gorman 2011)

The suggestion that people should be thought of as “collection points” along with Goodchild’s (2007b) idea of “citizens as sensors” dehumanizes the relationship between people and the places they inhabit. The suggestion that the “crowd” should be treated as a field of automata dismisses the immense value of possibly capturing the individual’s (or localized community’s) unique perspectives on place. The key here is to allow interaction and feedback (Grira et al. 2010). A system like Tagwatch might be leveraged into an ontology based on user tags. Van Exel’s work on trust and authority might provide the basis for a system of metadata verification produced by users. The ongoing work on qualitative GIS and metadata (Schuurman 2009) could lead to new avenues for enriching metadata. In short, the research frontier on how user-generated data might contribute additional meaning to the GeoWeb starts with metadata.

Acknowledgments The authors are grateful to Daniel Sui, Michael Goodchild, and Sarah Elwood for the invitation to submit this chapter to the volume on volunteered geographic information. We thank Peter Schweitzer, Martin van Exel, and two anonymous reviewers for helping us improve the structure and concepts of the paper.

References

Ball, M. (2010). What can be learned from the volunteer mapping efforts for Haiti? *Spatial Sustain*, January 31, 2010. http://vector1media.com/spatialsustain/what-can-be-learned-from-the-volunteer-mapping-efforts-for-haiti.html. Accessed July 28, 2011.

Batcheller, J. (2008). Automating geospatial metadata generation—An integrated data management and documentation approach. *Computers & Geosciences, 34*, 387–398. doi:10.1016/j.cageo.2007.04.001.

Batcheller, J., Gittings, B., & Dunfey, R. (2009). A method for automating geospatial dataset metadata. *Future Internet, 1*, 28–46.

Bates, M. (2002, September 11). Toward an integrated model of information seeking and searching. *Fourth international conference on information needs, seeking, and use in different contexts*, Lisbon, Portugal.

Budhathoki, N. (2010). Participants' motivations to contribute geographic information in an online community. Dissertation, University of Illinois Urbanna-Champaign.

Budhathoki, N., Bruce, B., & Nedovic-Budic, Z. (2008). Reconceptualizing the role of the users of spatial data infrastructure. *GeoJournal, 72*(3–4), 149–160. doi:10.1007/s10708-008-9189-x.

Chrisman, N. (1994). Metadata required to determine the fitness of spatial data for use in environmental analysis. In W. Michener, J. Brunt, & S. Stafford (Eds.), *Environmental information management and analysis: Ecosystem to global scales* (pp. 177–190). London: Taylor and Francis.

Coleman, D. J., Georgiadou, Y., & Labonte, J. (2009). Volunteered geographic information: The nature and motivation of produsers. *International Journal of Spatial Data Infrastructures, 4*, 332–358.

Comber, A., Fisher, P., & Wadsworth, R. (2008). Semantics, metadata, geographical information and users. *Transactions in GIS, 12*, 287–291. doi:10.1111/j.1467-9671.2008.01102.x.

Courtright, C. (2008). Context in information behavior research. In B. Cronin (Ed.), *Annual review of information science and technology* (pp. 273–306). Medford: Information Today, Inc.

Davies, C., Wood, L., & Fountain, L. (2005, Nov 8–10). User-centred GI: Hearing the voice of the customer. *Annual Conference of the Association for Geographic Information: AGI 05: People Places and Partnerships*, London.

Entchev, A. (2010). Comment to "Let's save metadata", February 16, 2010. http://www.spatially-adjusted.com/2010/02/15/lets-save-metadata/#comment-13108. Accessed July 27, 2011.

Federal Geographic Data Committee. (1994). *Content standards for digital geospatial metadata*. Washington, DC: Federal Geographic Data Committee.

Federal Geographic Data Committee. (2000). *Content standard for digital geospatial metadata workbook, Version 2.0*. Reston: Federal Geographic Data Committee.

Federal Geographic Data Committee. (2006). Clearinghouse concepts q&a. Federal Geographic Data Committee. http://www.fgdc.gov/dataandservices/clearinghouse_qanda. Accessed November 10, 2010.

Fee, J. (2010). Let's save metadata. http://www.spatiallyadjusted.com/2010/02/15/lets-save-metadata/. Accessed November 12, 2010.

Gahegan, M., Luo, J., Weaver, S. D., Pike, W., & Banchuen, T. (2009). Connecting GEON: Making sense of the myriad resources, researchers and concepts that comprise a geoscience infrastructure. *Computers & Geosciences, 35*, 836–854.

Goodchild, M. (2007a). Beyond metadata: Towards user-centric description of data quality. Spatial Data Quality 2007: ISSDQ. 13–15 June at Enschede, the Netherlands.

Goodchild, M. (2007b). Citizens as sensors: The world of volunteered geography. *GeoJournal, 69*(4), 211–221.

Goodchild, M. (2008, June 25–27). Spatial accuracy 2.0. In *8th international symposium on spatial accuracy assessment in natural resources and environmental sciences*, Shanghai.

Goodchild, M., Fu, P., & Rich, P. (2007). Sharing geographic information: An assessment of the geospatial One-stop. *Annals of the Association of American Geographers, 97*(2), 250–266.

Gorman, S. (2011). Statistical challenges of data at scale: Bringing back the science. http://blog.geoiq.com/2011/06/20/statistical-challenges-of-data-at-scale-bringing-back-the-science/. Accessed July 27, 2011.

Gould, M. (2006a). Meta-findability: Part 1. *GeoConnexion International Magazine, 5*(7), 36–38.

Gould, M. (2006b). Meta-findability: Part 2. *GEOconnexion International Magazine, 5*(8), 28–29.

Grira, J., Bédard, Y., & Roche, S. (2010). Spatial data uncertainty in the VGI world: Going from consumer to producer. *Geomatica, 64*(1), 61–71.

Haddad, T. C. (2010). Comment to let's save metadata. http://www.spatiallyadjusted.com/2010/02/15/lets-save-metadata/#comment-13113. Accessed July 20, 2011.

Haklay, M. (2010a). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and Planning B, 37*(4), 682–703.
Haklay, M. (2010b). Haiti—Further comparisons and the usability of geographic information in emergency situations. http://povesham.wordpress.com/2010/01/29/haiti-—-further-comparisons-and-the-usability-of-geographic-information-in-emergency-situations/. Accessed July 11, 2011.
Hennig, S., Belgiu, G., Wallentin, K., & Hormanseder, K. (2011). User-centric SDI: Addressing users in a third-generation SDI. In *Inspire Conference 2011*, Edinburgh.
International Organization for Standardization. (1986). *ISO 8879:1986 Information processing – Text and office systems – Standard Generalized Markup Language (SGML)*. Geneva: International Organization for Standardization.
International Organization for Standardization. (2003). *ISO 19115: 2003, Geographic information–metadata*. Geneva: International Organization for Standardization.
Korzybski, A. (1933). *A non-Aristotelian system and its necessity for rigour in mathematics and physics. Science and sanity*. Laxeville: International Non-Aristotelian Library.
Latour, B. (1999). *Pandora's hope: Essays on the reality of science studies*. Cambridge, MA: Harvard University Press.
Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.
Li, W., Yang, C., & Yang, C. (2010). An active crawler for discovering geospatial web services and their distribution pattern – A case study of OGC web map service. *International Journal of Geographical Information Science, 24*(8), 1127–1147. doi:10.1080/13658810903514172.
Maron, M. (2010). Haiti OpenStreetMap response. http://brainoff.com/weblog/2010/01/14/1518. Accessed November 11, 2010.
Maxwell, J., Edwards, C., Jensen, M., Paustian, S., Parrott, H., & Hill, D. (1995). *A hierarchical framework of aquatic ecological units in North American (Nearctic Zone)*. St. Paul: U.S. Department of Agriculture, Forest Service.
McLuhan, M. (1964). *Understanding media: The extension of man*. London: Sphere Books.
Millerand, F., & Bowker, G. (2009). Metadata standards: Trajectories and enactment in the life of an ontology. In M. Lampland & S. Star (Eds.), *Standards and their stories* (pp. 149–166). Ithaca: Cornell University Press.
Moore, M. (2010). Cyborg metadata: Humans and machines working together to manage information – Part 1: Text. *Online Currents, 24*(3), 131–138.
Morville, P. (2005). *Ambient findability*. Sebastopol: O'Reilly.
National Research Council. (2001). *National spatial data infrastructure partnership programs: Rethinking the focus*. Washington, DC: National Academies Press.
Nielsen, J., & Loranger, H. (2006). *Prioritizing web usability*. Berkeley: New Riders Press.
Onsrud, H. (Ed.). (2007). *Research and theory in advancing spatial data infrastructure concepts*. Redlands: ESRI Press.
O'Reilly, T. (2005). What is web 2.0?: Design patterns and business models for the next generation of software. http://www.oreillynet.com/lpt/a/6228. Accessed November 15, 2010.
Osborne, C. (2010). Mapping a crisis. *Guardian Online* http://www.guardian.co.uk/open-platform/blog/mapping-a-crisis. Accessed July 27, 2011.
Poore, B. (2003). Blue lines: Water, information, and salmon in the Pacific Northwest. Dissertation, University of Washington.
Poore, B., & Chrisman, N. (2006). Order from noise: Toward a social theory of geographic information. *Annals of the Association of American Geographers, 96*(3), 508–523.
Pultar, E., Cova, M., Yuan, M., & Goodchild, M. (2010). EDGIS: A dynamic GIS based on space time points. *International Journal of Geographical Information Science, 24*(3), 329–346. doi:10.1080/13658810802644567.
Qin, J. (2008). Controlled semantics vs. social semantics: An epistemological analysis. In *Proceedings of the Tenth International ISKO Conference: Culture and identity in knowledge organization* (pp. 5–8), Montreal, August 5–8, 2008.
Raymond, E. (1999). *The Cathedral and the Bazaar*. Sebastopol: O'Reilly.
Rodriguez, M., Cruz, I., Egenhofer, M., & Levashkin, S. (Eds.). (2005). *GeoSpatial semantics. Lecture notes in computer science* (Vol. 3799). Berlin: Springer.

Scharl, A., & Tochterman, K. (2007). *The geospatial web*. London: Springer.
Schuurman, N. (2008). Database ethnographies using social science methodologies to enhance data analysis and interpretation. *Geography Compass, 2*(5), 1529–1548.
Schuurman, N. (2009). Metadata as a site for imbuing GIS with qualitative information. In M. Cope & S. Elwood (Eds.), *Qualitative GIS: A mixed media approach* (pp. 41–56). Los Angeles: Sage.
Schuurman, N., & Leszczynski, A. (2006). Ontology based metadata. *Transactions in GIS, 10*(5), 709–726.
Schweitzer, P. (1998). Easy as ABC – Putting metadata in plain language. *GIS World, 11*(9), 56–59.
Shirky, C. (2005a). Ontology is overrated: Categories, links, and tags. http://www.shirky.com/writings/ontology_overrated.html. Accessed January 28, 2009.
Shirky, C. (2005b). Folksonomies + controlled vocabularies. http://many.corante.com/archives/2005/01/07/folksonomies_controlled_vocabularies.php. Accessed October 28, 2010.
Silver, J. (2010). Data information: How visual tools can transform lives. http://www.wired.co.uk/wired-magazine/archive/2010/09/features/data-information?page=all. Accessed November 11, 2010.
Six Rivers National Forest. (1999). *Metadata for stream*. Eureka, CA: U.S. Forest Service. http://www.ncgic.gov/GIS_Data/smf/hydro/stream.metadata.html. Accessed September 10, 2001.
Sui, D. (2008). The wikification of GIS and its consequences: Or Angelina Jolie's new tattoo and the future of GIS. *Computers, Environment and Urban Systems, 32*, 1–5.
Sui, D., & Goodchild, M. (2001). GIS as media? *International Journal of Geographical Information Science, 15*(5), 387–390. doi:10.1080/13658810110038924.
Sui, D., & Goodchild, M. (2011). The convergence of GIS and social media: Challenges for GIScience. *International Journal of Geographical Information Science, 25*(11), 1737–1748.
Tsou, M. (2002). An operational metadata framework for searching, indexing, and retrieving geographic information services on the Internet. In M. Egenhofer & D. Mark (Eds.), *GIScience 2002: Lecture notes in computer science* (Vol. 2478, pp. 313–332). Berlin: Springer.
Turner, A. (2006). *Introduction to neogeography*. Sebastopol: O'Reilly Media.
U.S. Department of Commerce, National Oceanic and Atmospheric Administration. (1999). 50 CFR Part 223: Endangered and threatened species; proposed rule governing take of threatened Snake River, Central California Coast, South/Central California Coast, Lower Columbia River, Central Valley California, Middle Columbia River, and Upper Willamette River evolutionarily significant units (ESUs) of West Coast steelhead. *Federal Register, 64*(250), 73479–73506.
U.S. Department of Commerce, U.S. Forest Service, Remote Sensing Lab, Pacific Southwest Region. (2004). Metadata for NWCSTRM03_2 2004. http://www.fs.fed.us/r5/rsl/projects/gis/data/calcovs/nwcstrm03_2.html. Accessed July 27, 2011.
Van Exel, M., & Dias, E. (2011). Towards a methodology for trust stratification in VGI. *VGI Pre-Conference at AAG*, Seattle. http://vgi.spatial.ucsb.edu/sites/vgi.spatial.ucsb.edu/files/file/aag/van_Exel_abstract.pdf. Accessed July 8, 2010.
van Oort, P., Hazeu, G., Kramer, H., Bregt, A., & Rip, F. (2009). Social networks in spatial data infrastructures. *GeoJournal, 75*(1), 105–118.
Vander Wal, T. (2007). Folksonomy: Coinage and definition. http://www.vanderwal.net/folksonomy.htm. Accessed November 15, 2010.
Waters, T. (2010). The OpenStreetMap project and Haiti earthquake case study. http://www.slideshare.net/chippy/openstreetmap-case-study-haiti-crisis-response. Accessed November 10, 2010.
Weber, S. (2004). *The Success of open source*. Cambridge: Harvard University Press.
Weinberger, D. (2007). *Everything is miscellaneous: The power of the new digital disorder*. New York: Times Books.
Wikipedia. (2010). OpenStreetMap. http://en.wikipedia.org/wiki/OpenStreetMap. Accessed November 15, 2010.
Wolf, E., Matthews, G., McNinch, K., & Poore, B. (2011). *OpenStreetMap collaborative prototype, phase one* (Open-file report of 2011–1136). Reston: U.S. Geological Survey. http://pubs.usgs.gov/of/2011/1136/. Accessed December 12, 2011.
Wood, D., Fels, J., & Krygier, J. (2010). *Rethinking the power of maps*. New York: Guilford.