

OpenStreetMap:

User-Generated Street Maps

The OpenStreetMap project is a knowledge collective that provides user-generated street maps.

The process of mapping the Earth accurately was, until recently, the preserve of highly skilled, well-equipped, and organized individuals and groups. For many years, it was usually the role of surveyors, cartographers, and geographers to map the world and transcribe it on paper or, since the 1960s, into the computer. Lewis and Clark's expedition to map North America's West, and Lambton and Everest's Great Arc expedition to measure India, are just two famous episodes in the history of maps and map making. Each country has an established national mapping agency charged with keeping the national maps accurate and current (for example, the US Geological Survey and the UK Ordnance Survey).

Less than a decade ago, it was common to assume that a person needed a university-level degree to be able to measure the Earth and transcribe the information on paper or into the computer as well as expensive equipment and infrastructure to support his or her work.

This, however, has changed dramatically over the past decade. On 1 May 2000, US President Bill Clinton announced the removal of selective availability of the GPS signal¹ and, by so doing, provided much improved accuracy for simple, low-cost GPS receivers. In practical terms, this made it possible to acquire the receiver's position with an accuracy of 6 to 10 meters in normal conditions, in contrast to roughly 100 meters before the "switch off." Attempts to develop location-based services predated this announcement² and were based on information from mobile

phone masts or other beacons. However, these methods hadn't gained much market share owing to their technical complexity and inability to provide a universal coverage. In contrast, GPS enabled the development of cheap receivers with good positional accuracy, and, by mid-2001, it was possible to purchase a receiver unit for roughly US\$100.³ These receivers helped more people than ever before collect information about different locations and upload it to their computers. However, until 2002, when an interchange standard (GPS eXchange format or GPX) was published, manipulating and sharing this information was a complicated task that required computing and data manipulation knowledge. Fortunately, most GPS receiver developers rapidly adopted the GPX standard, and, by 2004, it had become commonplace (www.topografix.com/gpx.asp).

The wide availability of high-quality location information has enabled mass-market mapping based on affordable GPS receivers, home computers, and the Internet. Although a range of projects based on user-generated mapping has emerged, OpenStreetMap (OSM) is probably the most extensive and effective project currently under development. In this article, we review the project and provide an overview for the techniques and methodologies used within it.

OpenStreetMap Background

Technological changes over the past 10 years, in combination with increased bandwidth and the ability to provide better tools for collaboration, have led to "crowdsourcing"⁴—a term developed from the concept of outsourcing in which business operations are transferred to remote, many times cheaper locations.⁵ Similarly,

Mordechai (Muki) Haklay
and Patrick Weber
University College London

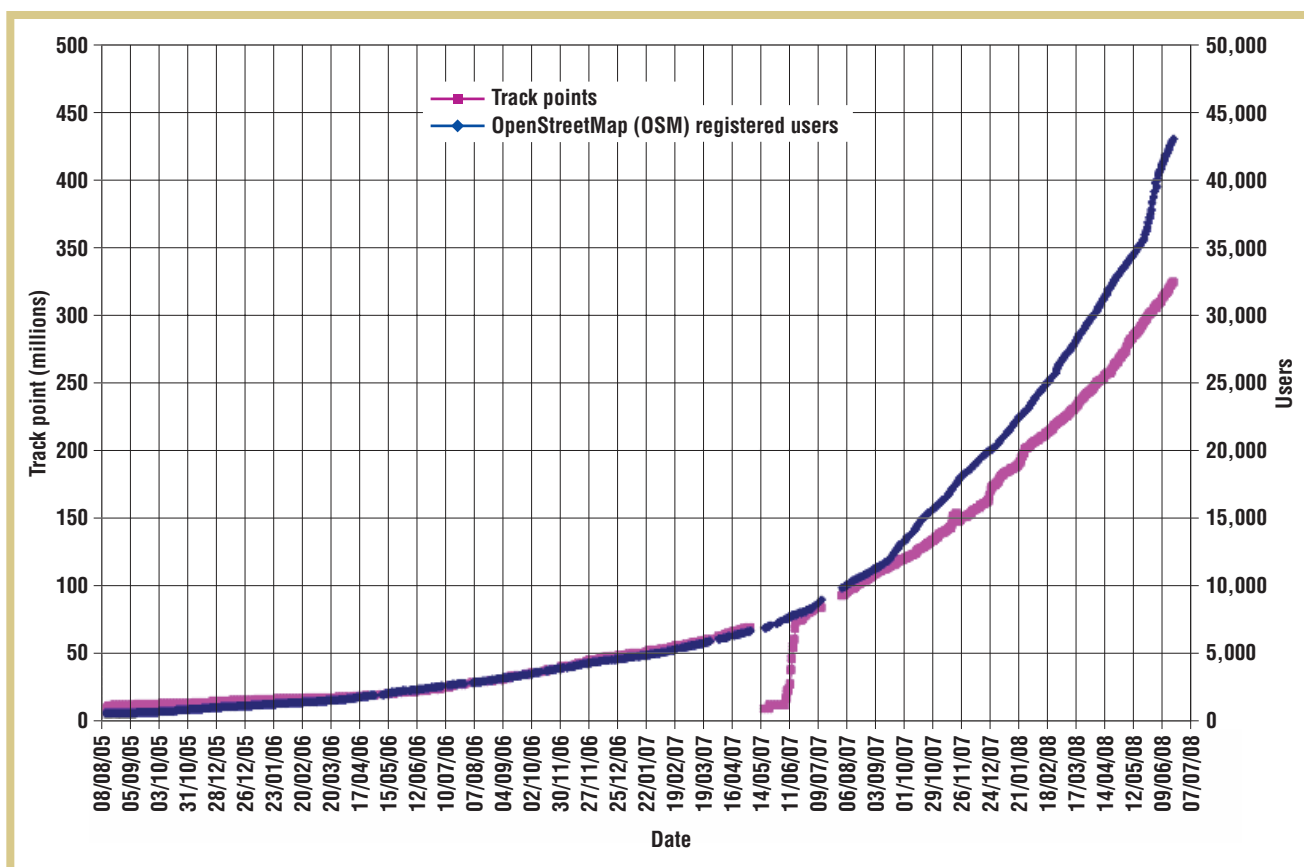


Figure 1. Graph of user and contributions growth to OSM on a monthly basis. The graph shows the accelerating growth in number of users and the rapid increase in data entry measured in track points (source: <http://wiki.openstreetmap.org>).

crowdsourcing is how large groups of users can perform functions that are either difficult to automate or expensive to implement.

OSM follows the peer production model that created Wikipedia; its aim is to create a set of map data that's free to use, editable, and licensed under new copyright schemes. The project, born at University College London (UCL) in July 2004, was founded by Steve Coast; although Coast moved on to start his own company, UCL still supports and hosts the main server infrastructure. As of May 2008, OSM had more than 33,000 registered users (with approximately 3,500 currently active contributors), and data contribution growth continues to rise quickly (see Figure 1). A considerable number of contributors edit the world map collaboratively using the OSM technical infrastructure,

and a core group, estimated at approximately 40 volunteers, dedicate their time to creating and improving OSM's infrastructure, including maintaining the server, writing the core software that handles the transactions with the server, and creating cartographical outputs. There's also a growing community of software developers who develop software tools to make OSM data available for further use across different application domains, software platforms, and hardware devices.

A key motivation for this project is to enable free access to current geographical information where, in European countries, accurate digital geographical information is considered to be expensive and out of the reach of individuals, small businesses, and community organizations. In the US, where basic road information is available through

the US Census Bureau's Tiger (Topologically Integrated Geographic Encoding and Referencing)/Line program, the details provided are limited to streets and roads only; it doesn't include green space, landmarks, and the like. In addition, owing to the high cost of mapping, the Tiger system's update cycles are infrequent and don't take into account rapid changes. Commercial geographical information products from providers such as NAVTEQ are also expensive and aren't available for individual users in an accessible format.

The OSM project's hub is the main OSM Web site (www.openstreetmap.org), which contains four parts. Visitors are first greeted with a Google Maps-style online mapping interface, which lets visitors pan, zoom, and search the OSM world map and discover which geographical areas are completed. An

export function allows users to download portions of the OSM information in different raster and vector formats for further use or processing. The editing tab allows anyone to contribute to the project by digitizing geographical features, uploading GPX traces from hand-held GPS units, or correcting errors they might have discovered in their local areas. The OSM community wiki, open for all registered users, contains information about the project and offers guidance on best practices to casual and advanced mapping contributors and an extensive documentation of the project's technical infrastructure.

Noticeably, OSM decided to follow the route of allowing only registered users to edit the map, not following Wikipedia's open-for-all approach. OSM project leads wanted the ability to trace the information source in case of copyright disputes and ensure the information was maintained.

In the following sections, we describe the OSM GeoStack — the set of tools that lets users capture, produce, communicate, aggregate, and consume the geographical information produced in the project (see other work⁶ for further

weight online Flash-based editor, Potlatch, which lets users add, update, or delete geographical features through a relatively easy-to-use interface. The interface is kept deliberately simple, with more advanced functionality provided through keyboard shortcuts; Potlatch gives extensive guidance to users by providing predefined tagging schemas for frequently occurring features (such as motorways or primary roads).

Potlatch also lets users upload and integrate GPX tracks recorded from handheld GPS units. Other important data sources from which users trace road networks and other features is satellite imagery and out-of-copyright maps, which are integrated into the mapping interface. Notably, at the end of 2006, Yahoo granted OSM the right to use its satellite imagery Web service to trace roads and other features. For example, Yahoo hosts detailed aerial imagery of Baghdad, Iraq, which let OSM contributors remotely map the city in great detail, resulting in the most detailed online map of Baghdad to date. However, this map is based solely on amateur image interpretation, and, despite the original effort, very little fur-

can be extended using several independently developed plug-ins. Examples of user-contributed plug-ins include custom Web mapping service (WMS) background imagery and Yahoo aerial imagery, live recording of external GPS data, and a data and tagging scheme validation tool, to name just a few.

Apart from individual user contributions from GPS tracks and the digitizing of aerial imagery, OSM has also taken advantage of the availability of free geographical information in certain parts of the world. For example, over the past year, US contributors have imported the public-domain Tiger information, adding to OSM's comprehensive street and highway coverage for the entire US. In July 2007, commercial navigation information provider AND (Automotive Navigation Data) donated the entire street map of the Netherlands to the project, thus completing street-level mapping of the first country in OSM. Local and national authorities have contributed information to the project as well—for example, the Isle of Man's Department for Local Government and the Environment donated geographical information to OSM, enabling the creation of an excellent map of the island.

OpenStreetMap's aim is to create a set of map data that's free to use, editable, and licensed under new copyright schemes.

discussion of the GeoStack). As already noted, the Web site provides a mechanism to consume OSM data — but in the next section we move to the starting point of the stack, where data is captured and produced.

Editing Tools

User-contributed geographical information is obviously a core part of OSM, and the OSM developer community has made a considerable effort to implement tools to facilitate user contributions to the database. For most casual contributors, the OSM Web site offers a light-

ther information (such as street names) has been added to the map.

More experienced OSM contributors also use the Java OpenStreetMap Editor (JOSM), an editing suite with an interface more akin to traditional geographical information systems (GIS) packages. The application lets users import, edit, and tag OSM data offline and allows bulk uploads of OSM updates through the OSM application programming interface (API). JOSM offers advanced functionalities such as linking OSM features to photos and audio notes, supports data conflict resolutions, and

Technical Infrastructure

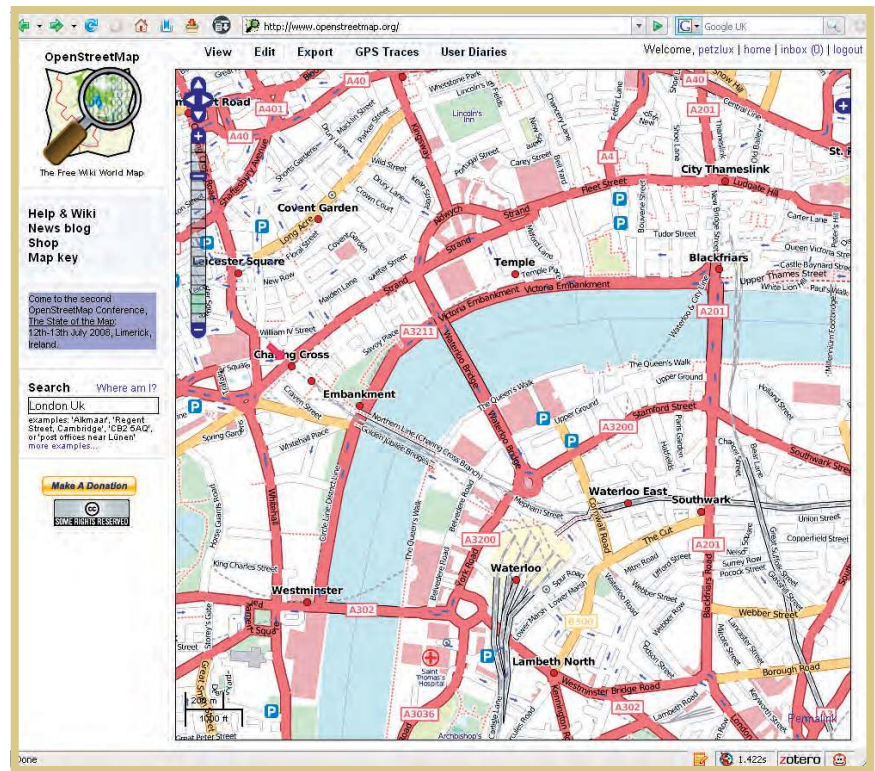
The next element of the GeoStack focuses on the way in which geographical information is stored and organized. Here, at the core of OSM data management, it's easy to see how open source philosophy permeates the project's technical infrastructure. OSM is built iteratively using the principle that the simplest approach to any problem is the best way to ensure the success of the project as a whole. OSM's developers deliberately steered away from using existing standards for geographical information from standard bodies such as the Open Geospatial Consortium (OGC)—for example, its WMS standard. They felt that most such tools and standards are hard to use and maintain, citing performance issues with, for instance, MapServer (a popular open

Figure 2. Main OpenStreetMap Web site Slippy Map with detailed coverage of central London. Notice the ability to edit, export, and upload data by using the top tabs (source: www.openstreetmap.org).

source WMS) and a lack of adaptability of OGC-compliant software packages to support wiki-style behavior.

At the heart of OSM's technical infrastructure lies the central database holding the live data, which is implemented in MySQL. The database schema is designed to support wiki behaviors, such as versioning and rollbacks, and keeps copies of modified or deleted features indefinitely. All geographical entities are recorded as points (nodes), which contain the latitude and longitude coordinates along with user name and timestamp information. Linear and area features are defined by reference to a list of ordered nodes, called *ways*. Area features aren't explicitly defined in the database schema—rather, they're defined implicitly by the condition of a way that's closed (the first node of a way is the same as the last one) and explicit tagging conventions (using the **tag area=yes**).

Along with the geographical coordinates of features in the OSM database, attributes of features are recorded for each node and way as semicolon-separated **key=value** pairs (for example, **type=pub;name=The Bull**). This tagging schema, which is increasingly being developed into a complex taxonomy of real-world feature classes and objects, is a core part of the OSM initiative and is community-driven. Any member of the community can contribute to and update the schema by proposing new **key=value** pairs. Discussion on the OSM wiki leads the community to vote on tag propositions and reach a consensus on the definition of and best practices for feature tagging. Interestingly, and sometimes controversially, the OSM database doesn't pose any limitations on the tags users can attach to features, and thus adhering to the core tagging schema the community agreed upon



is voluntary. Guidance on best practices for tagging is given at the editing software level for casual contributors. The lack of restrictions also benefits users because they can freely implement custom tagging schemas for specific applications.

Access to the core OSM database is provided by a dedicated RESTful API, which is implemented in Ruby on Rails and supports authentication, enabling users to add, update, and delete geographical features. The API accepts and outputs data in OSM XML, a dedicated data transport format developed for the project that replicates the databases' specific entity model. All editing tools use this API for accessing and updating the main database. As a result, editing and presentation tools can be developed independently from the database, with the lightweight communication protocol acting as a glue between the elements of OSM's GeoStack.

Mapping Outputs

Given the database, which contains all the geographical information and attri-

butes OSM contributors have collected, the main cartographic output from the OSM information is presented on the OSM Web site as a Google Maps-like interface, coined "Slippy Map," which uses the open source AJAX library OpenLayers to dynamically update the map display and allow interaction with users. As users drag the map, the visible extent is updated and new map tiles are requested in the background without reloading the entire HTML page. A search function (implemented as an external Web service) lets users quickly find cities, villages, or other points of interest (POIs) in the database. Recently, OSM developers added an export tab, which lets users quickly generate map images, PDF files, and raw data downloads of custom bounding boxes.

The default set of tiles on the main OSM Web site (see Figure 2) is rendered using Mapnik, an open source library for generating high-quality map images. It uses a weekly database dump as the source for the rendering of map tiles, given that live rendering of tiles on client request would be too computationally

expensive to be practical; map tiles are therefore rendered for all zoom levels and saved on the server so that they can be served rapidly as static images.

Owing to the open source nature of all the tools needed for map rendering, several OSM contributors have developed custom map tile sets that cater to specific needs and target audiences. One example is a tile set that highlights cycle-path networks and other features relevant to cyclists.

Because of the high demands on computing during rendering, a community grid element of OSM has developed. Tiles@Home (T@H) is a distributed map-tile-rendering system that comprises a central coordinating server and approximately 100 active software clients (as of May 2008), which members of the OSM community contributed. The main server distributes rendering jobs between clients, which collect the relevant data from the OSM API and render a set of map tiles that are then uploaded back to the server for distribution. The main benefit of T@H is that it distributes the computational load between multitudes of clients, enabling the system to quickly render large numbers of up-to-date map tiles. T@H is especially useful for OSM mapping contributors that want to quickly see the results of data changes, as they can request specific ar-

or Gosmore to generate maps quickly. Other members of the OSM community are also developing import and conversion tools to support OSM data in GIS packages such as Manifold GIS (www.manifold.net) and ArcGIS (www.esri.com/software/arcgis/). Significantly, vendors of commercial GIS packages, such as CadCorp SIS and Global Mapper, have recently included OSM XML data support out of the box.

Users have converted OSM information for use on a multitude of devices, including mobile phones, PDAs, and GPS units. A community-maintained software package lets users translate OSM data into the Garmin IMG GPS map format, despite this format's proprietary nature and lack of documentation.

Social Collaboration

Unlike Wikipedia, on which individuals create the majority of content at disparate locations, the OSM community organizes a series of local workshops (called "mapping parties"), which aim to create and annotate content for localized geographical areas. These events are designed to introduce new users and contributors to the community with the hands-on experience of collecting, processing, and uploading data to the OSM project. The meetings might take the

Europe spent two days driving, cycling, and wandering around the island with GPS receivers to collect a complete coverage of roads and footpaths. After collecting the individual contributions, processing them, and uploading the data, a practically complete map of the island emerged.

Although these community events positively contribute to the overall project by generating new data and street labeling, they aren't only meant as data collection exercises. Mapping parties play an essential part in creating and fostering local OSM user groups and creating a vibrant social community around the project (see Figure 3). User groups have now formed all over the UK, as well as in continental Europe and the rest of the world. A detailed discussion of a mapping party appears elsewhere.⁷

Motivations and Challenges

OSM's success should be attributed to its founder's vision of the project as a combined social and technical challenge. In many interviews and presentations, Coast emphasizes the social side's importance: "A big aspect of getting OSM off the ground was the mapping parties: getting drunk and arguing with people."⁸ Nick Black, another OSM core contributor, noted that people have a range of reasons for getting involved in the project—from certain ideological views such as a belief in the provision of free information to improve the world, to anti-national mapping agency views, to those who enjoy going out and mapping or sitting at home and writing computer code, to those who enjoy feeling like part of a community.⁹ Coast also adds that there's an "addictive" aspect of getting involved in the project, which adds to the participants' commitment.

Indeed, OSM can boast significant achievement and is currently going through a period of rapid growth in terms of the number of users who contribute to the map, its visibility as the leading Open Geodata project, and of the number of map edits. Its infrastruc-

OpenStreetMap is built iteratively using the principle that the simplest approach to any problem is the best way to ensure the success of the project as a whole.

reas to be added to the T@H rendering queue. Typically, such requests for updated map tiles are fulfilled in a matter of hours, in contrast to up to seven days for a Mapnik tile set update.

Users interested in only a small area and who don't need a Web-mapping server set-up can use local software rendering packages such as Kosmos

shape of informal and small gatherings for a few hours to complete missing features of a small defined neighborhood, up to more ambitious efforts that take several days and involve several dozens of participants. One of the first mapping parties took place on the Isle of Wight, off the south coast of England, in May 2006. More than 30 participants from

Figure 3. A mapping party. In February 2008, Schuyler Erle and Mikel Maron, contributors to OSM, held a series of multiday mapping parties in several universities to create local contributor groups (source: www.flickr.com/photos/tags/freemapindia2008/).

ture is demonstrating how a modern, open source, and light structure can provide a reliable and scalable GIS; the innovation that OSM offers can also teach many established GIS providers valuable lessons. The mainstream GIS literature has already noted and discussed OSM and similar activities.¹⁰

Along with these achievements, some open issues still must be taken into account, such as the fitness for purpose of OSM data, the influence of geography and participation on the project, the ability to continue to update the information, and licensing.

Since early 2007, with the completion of the mapping of Cambridge, OSM information and the cartographic ability of Mapnik demonstrated the usefulness of the information for several cartographic products and presentation of paper maps. However, this coverage isn't universal. Although, as Coast notes, "it's important to let go of the concept of completeness,"⁸ some idea about information quality is crucial for evaluating how fit OSM data is for various applications. Considerations that should be included are how well an area is covered by data, the data's accuracy in terms of positional accuracy and attributes, the consistency in terms of classification or data-capture procedures, and quality control. Currently, OSM doesn't provide any of these measures and doesn't have internal quality assurance procedures. Preliminary results from an evaluation of completeness at UCL suggest that OSM has covered about 29 percent of roads in England. Even in London, where the project started, large areas are undermapped.

In many applications, attribute infor-



mation, such as traffic directions and restrictions or street names, can be crucial. Although data capture based on aerial imagery sped up road mapping, the road name can't be captured without a survey at the ground level unless the surveyor breaks copyright laws and uses an existing map to copy the name. Furthermore, mistakes from misinterpreting aerial imagery also must be rectified through ground survey. Here, there's no escape from the world's physical geography, and OSM is dependent on willing contributors on the ground in the area under survey. The same is true for updates and changes to the data—for example, when a road changes from two-way to one-way owing to local traffic regulations.

Although the common argument from OSM advocates is that because of the data's free nature, people who spot a mistake are likely to be more motivated to rectify it, in practice, OSM is exhibiting the same participation inequality as many other user-generated content projects.¹¹ A few users are responsible for contributing a significant amount of the information—this phenomena is also visible in terms of places, where many areas are mapped by a single user. It's reasonable to assume that, as on Wikipedia, the number of users who don't contribute any improvement to the map can reach 99.8 percent, to

which the complexities of adding or updating data to OSM due to the usability of the data entry application and the need to understand the ontological schema need to be added. In short, this is a challenging aspect that the OSM community must solve.

Finally, the issue of licensing has bedeviled OSM from the start. The project started with the Creative Commons framework (CC-BY-SA), which has proved to be unsuitable for geographical information. For example, the "share alike" principle means that by mixing OSM information with other sources of information, the resulting map must be shared under the same terms. However, OSM creates the base map, which in most geographical analysis projects is the starting point. The map is augmented either through cartographic processing to produce products such as tourist maps or combined with other sources of information, such as the census, to create thematic maps. It can also be combined with sensitive commercial information for analyzing store locations, among other things. In all these cases, the CC-BY-SA prevents the use of the data or might complicate matters significantly. The OSM community has been discussing a new license for more than two years and hopes to publish one soon, although it's clear that some use issues will remain unresolved.

the AUTHORS



Mordechai (Muki) Haklay is a senior lecturer in geographical information science in the Department of Civil, Environmental and Geomatic Engineering at University College London. His research interests are in public access to environmental information, human-computer interaction, usability engineering for GIS, and the societal aspects of GIS use. Haklay has a PhD in geography from UCL. He's a member of the ACM, a fellow of the Royal Geographic Society, and a member of the Association for Geographical Information (AGI). Contact him at m.haklay@ucl.ac.uk.



Patrick Weber is a third-year engineering doctorate student in University College London's Department of Computer Science. His research interests lie in the application of GI science and spatial decision support systems to gain a better understanding of foreign direct investment in London as well as the wider accessibility and public awareness issues for geographical information science. Weber has an MSc in GIS from UCL and an MS in physical geography from the Université Louis Pasteur, Strasbourg. Contact him at p.weber@ucl.ac.uk.

OSM provides a good example of the social and technical aspects of user-generated content communities. Using terminology found in other work,¹² OSM is a knowledge collective that creates Open Geodata as its main objective. Simultaneously, it includes a peer production network, as different groups within the organization are developing different aspects of the project—digitizing tools, map-rendering software, server software, and running activities such as mapping parties. It's utilizing community computing grids in the process of rendering map tiles through T@H. OSM uses social mobile computing, especially during mapping parties when participants coordinate their work via mobile GPS receivers and mobile phones. The group-forming network appears on the main wiki, which contains information about the project, and also through an array of active mailing lists, Web forums, Internet Relay Chats (IRCs), and other modes of computer-mediated communication (CMC). Finally, social accounting is evident in OSM—for example, in the contributions of various members of the OSM community highlighted on a Web site about the amount of computing they've contributed or how many edits they've performed during the last week, month, and year.

OSM also demonstrates some impor-

tant aspects in terms of geographical information delivery. The data format and structures demonstrate a simple approach for developing and delivering geographical information, in contrast to current practice within the GIS industry. On the other hand, OSM data isn't complete or consistent across the world, or even across London, where the project started. The data's accuracy is unknown, given that there are no systemic and comprehensive quality assurance processes integral to the data collection. Furthermore, there's no intention of universal coverage; as Coast remarks: "Nobody wants to do council estates. But apart from those socioeconomic barriers—for places people aren't that interested in visiting anyway—nowhere else gets missed."⁹

Although the OSM project started in 2004, it's still in its early stages—the area of user-generated geographical information will surely grow in the future with applications in the private, public, and voluntary sectors. It has already shaken the world of geographical information, and the wider industry is paying attention to the emerging business models that OSM volunteers have created. ■

ACKNOWLEDGMENTS

We thank the anonymous reviewers who provided useful suggestions on an earlier version of this article. Thanks to the Royal Geographical Society's Small Research Grants program, which

provided early support for the engagement of OpenStreetMap in 2006.

REFERENCES

1. B. Clinton, *Improving the Civilian Global Positioning System (GPS)*, Office of Science and Technology Policy, Executive Office of the President, 1 May 2000; http://clinton4.nara.gov/WH/EOP/OSTP/html/0053_4.html.
2. A. Giordano, M. Chan, and H. Habal, "A Novel Location-Based Service and Architecture," *6th IEEE Int'l Symp. Personal, Indoor, and Mobile Radio Communications, Wireless: Merging onto the Information Superhighway* (PIMRC 95), IEEE Press, 1995, pp. 853–857.
3. J. Hightower and G.A. Borriello, *Survey and Taxonomy of Location-Sensing Systems for Ubiquitous Computing*, UW CSE 01-08-03, Dept. of Computer Science and Eng., Univ. of Washington, 2001.
4. J. Howe, "The Rise of Crowdsourcing," *Wired*, June 2006; www.wired.com/wired/archive/14.06/crowds.html.
5. T.L. Friedman, *The World Is Flat: A Brief History of the Twenty-First Century, Updated And Expanded Edition*, Farrar, Straus, and Giroux, 2006.
6. A.J. Turner, *Introduction to Neogeography*, O'Reilly Media, 2006.
7. C. Perkins and M. Dodge, "The Potential of User-Generated Cartography: A Case Study of the OpenStreetMap Project and Mapchester Mapping Party," *North West Geography*, vol. 8, no. 1, 2008, pp. 19–32.
8. GISPro, "The GISPro Interview with OSM Founder Steve Coast," *GIS Professional*, no. 18, Oct. 2007, pp. 20–23.
9. N. Black, "OpenStreetMap—Geodata Collection for the 21st Century," presented at the AGI annual conference, 2007; www.slideshare.net/nickb/nick-black-openstreetmap-geodata-collection-for-the-21st-century.
10. M.F. Goodchild, "Citizens as Voluntary Sensors: Spatial Data Infrastructure in the World of Web 2.0," *Int'l J. Spatial Data Infrastructures Research*, vol. 2, 2007, pp. 24–32.
11. J. Nielsen, "Participation Inequality: Encouraging More Users to Contribute," *Alertbox*, 9 Oct. 2006; www.useit.com/alertbox/participation_inequality.html.
12. A. Saveri, H. Rheingold, and K. Vian, *Technologies of Cooperation*, Inst. for the Future, 2005.