International Journal of Human–Computer Interaction

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hihc20

The Crowd is the Territory: Assessing Quality in Peer-Produced Spatial Data During Disasters

Jennings Anderson, Robert Soden, Brian Keegan, Leysia Palen & Kenneth M. Anderson

To cite this article: Jennings Anderson, Robert Soden, Brian Keegan, Leysia Palen & Kenneth M. Anderson (2018) The Crowd is the Territory: Assessing Quality in Peer-Produced Spatial Data During Disasters, International Journal of Human–Computer Interaction, 34:4, 295-310, DOI: 10.1080/10447318.2018.1427828

To link to this article: https://doi.org/10.1080/10447318.2018.1427828

Published online: 30 Jan 2018.

Submit your article to this journal

Article views: 484

View related articles

View Crossmark data

Citing articles: 11 View citing articles

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=hihc20

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
2018, VOL. 34, NO. 4, 295–310
https://doi.org/10.1080/10447318.2018.1427828

Taylor & Francis
Taylor & Francis Group

Check for updates

The Crowd is the Territory: Assessing Quality in Peer-Produced Spatial Data During Disasters

Jennings Anderson[a], Robert Soden[a], Brian Keegan[b], Leysia Palen[a,b], and Kenneth M. Anderson[a]

[a]Department of Computer Science, University of Colorado Boulder, Boulder, CO, USA; [b]Department of Information Science, University of Colorado Boulder, Boulder, CO, USA

ABSTRACT
Today, disaster events are mobilizing digital volunteers to meet the data needs of those on the ground. One form of this crowd work is Volunteered Geographic Information. This peer-produced spatial data creates the most up-to-date map of the affected region; maintaining the accuracy of these data is therefore a critical task. Accuracy is one aspect of data quality, a relative concept requiring standards to measure against. The field of Geographic Information Sciences has developed standards for this comparison, achieving widespread acceptance. However, the peer production model of spatial data presents new opportunities—and challenges—to traditional methods of quality assessment. Through analysis of the OpenStreetMap database, we show that temporal editing patterns and contributor characteristics can provide additional means of understanding spatial data quality. Drawing upon experiences from Wikipedia, we offer and evaluate three intrinsic quality metrics of peer-produced spatial data to assess the quality of contributions to OpenStreetMap for crisis response.

1. Introduction

This article examines methods for assessing the quality of peer-produced spatial data, or Volunteered Geographic Information (VGI), for use in crisis response. For many parts of the world, VGI is the primary geospatial data source because it is the most accessible and complete source of data for the area (Palen, Soden, Anderson, & Barrenechea, 2015; Towne, Kittur, Kinnaird, & Herbsleb, 2013). As such, crisis responders often use these datasets during disasters. For example, the 2010 Haiti Earthquake destroyed much of the country's government buildings, and with them, access to official mapping resources (Palen et al., 2015; Soden & Palen, 2014). In just a few days, organizing online, hundreds of contributors to OpenStreetMap (OSM) created the most complete map of Haiti in existence. This map became the de-facto basemap for subsequent rescue and relief operations (Soden & Palen, 2014). This early instance of "disaster mapping" was a catalyst in creating a new form of volunteer disaster response work (Palen et al., 2015). Today, thousands of online volunteers mobilize before, during, and through the recovery phase of a disaster to answer the data needs of the community and responding organizations.

Crisis informatics research seeks to understand how new technologies enable volunteers to mobilize, create, and process information about a disaster event. Therefore, of specific importance to both the OSM and the crisis response communities is disaster mapping (or crisis mapping). Disaster mapping is the practice of volunteer contributors converging online to improve the map for a region experiencing a disaster or crisis (Eckle & Albuquerque, 2015; Poiani, Rocha, Degrossi, & Albuquerque, 2016). In the case of OSM, the Humanitarian OSM Team (HOT) is an active community organizer in coordinating these tasks all over the world (Palen et al., 2015; Poiani et al., 2016). These activities leave a very distinct contribution pattern on the map: specific regions with considerably more coverage of certain objects (typically roads and buildings) than the surrounding area. These improved maps are used for emergency response, planning, routing, and more (Soden & Palen, 2014, 2016).[1] With the widespread use of OSM data in disaster response, developing and validating measures of information quality for it is essential.

Studies of online peer production systems like Wikipedia have demonstrated that high-quality, open source content can be generated by integrating contributions from non-experts (Benkler, 2006). VGI systems like OSM emulate many of the features of these peer production systems, but the spatial—rather than textual—knowledge they encode requires alternative methods for measuring and validating the quality of their user-generated content. Developing methods for assessing the quality of spatial information is a fundamental issue within the study of geographic information science (GIScience). Any representation of spatial information necessarily involves some loss of detail and thus quality. The challenge that this presents is illustrated by Borges' famous parable that imagines a civilization so obsessed with precision that they constructed a 1:1 scale map of their territory. The result of their labor was

CONTACT Jennings Anderson jennings.anderson@colorado.edu 1111 Engineering Drive, ECOT 717, 430 UCB Boulder, CO 80309-0430, USA.

Color versions of one or more of the figures in the article can be found online at www.tandfonline.com/hihc.

[1]For more information on the process of disaster mapping, see (Eckle & Albuquerque, 2015).

© 2018 Taylor & Francis Group, LLC

perfectly accurate but totally unusable (Borges, as quoted in Eco, 1995). This problem is faced not just by the designers of maps, but of all information systems: maps are abstract, incomplete, and imperfect portrayals of the phenomena they are created to represent. This condition, illustrated by Korzybski's famous maxim that "the map is not the territory," renders questions of information quality more complex than they might initially appear (Korzybski, 1958).

A central motivation for this work is the lack of authoritative reference geographic data in many parts of the world, making more traditional quality analysis by comparing to reference data sources—referred to here as *extrinsic* methods—impossible. Our research expands upon existing methods within GIScience for assessing map quality, which rely on attributes, such as completeness, consistency, and accuracy, to take advantage of the behavioral meta-data of VGI contributors' activities that are unavailable to traditional data sources. We draw on previous research measuring the information quality of Wikipedia articles based on the *intrinsic* processes generating them, such as the number of editors or how recently the data has been updated. We identify analogous generative features in OSM data and evaluate three metrics drawing on contributors' histories and temporal contexts to examine their relationship with alternative intrinsic information quality metrics. Both intrinsic and extrinsic quality assessments of VGI have been explored in GIScience. Our metrics are distinct in that they rely primarily on the metadata of the individual contributions and contributors: the details and context of how the digital volunteers converged, not just the geographic features that were contributed. This distinction connects this work from the more traditional approaches of GIScience to the fields of social computing and human computer interaction.

Using a quantitative case study method, we identify four different areas of the global map that have been the geographic focus of disaster mapping activities in the past. For each of these areas, we apply our three proposed intrinsic quality metrics, which expose varying histories of contributions, each telling a different story, consistent with its associated crisis event. We then apply these metrics to areas of the map known and agreed to be of very high-quality for comparison. The differences—exposed by these metrics—suggest we are capturing substantively different mechanisms by which VGI information is contributed, which, in turn, has implications for quality assessment. The following sections will, first, discuss the background of information quality and quality assessment in peer production; second, discuss the OSM project and describe our dataset; then next introduce three approaches to intrinsic quality analysis based on contribution metadata; and then, finally, evaluate our methods applied to various parts of the map that have been the sites of disaster mapping in the past. We conclude with a discussion of how these metrics fit within the larger domain of geospatial data quality assessment and offer suggestions for future work.

2. Background

How to measure information quality has been the subject of a substantial body of research across information science, management-related fields, and geography. We begin by reviewing work on measuring information quality using *extrinsic* and *intrinsic* data sources in the context of peer production and spatial information. The majority of prior literature assessing the quality of OSM—with a few notable exceptions (Barron, Neis, & Zipf, 2014)—has typically focused on assessing quality relative to authoritative data sources; it, therefore, overlooked the potential offered by specific intrinsic features unique to VGI to measure the quality of peer-produced data. This gap between the value of intrinsic features for measuring information quality and the underutilization of these features unique to VGI for assessing quality in OSM motivates our subsequent analysis to employ contributors' histories and temporal contexts as *intrinsic* sources of VGI quality.

2.1. Data and information quality

In this section, we (1) identify commonly used dimensions for measuring information quality through *extrinsic* and *intrinsic* dimensions; (2) examine how the quality of spatial information is traditionally assessed; and (3) discuss the importance of spatial information quality for safety-critical operations, such as disaster response.

Information quality frameworks

There are many sources of variance in information quality. Information quality problems arise because of incomplete, ambiguous, inaccurate, inconsistent, or redundant mappings between real world properties and their representation in an information system (Lee, Strong, Kahn, & Wang, 2002; Wand & Wang, 1996). We employ a taxonomy that differentiates information quality based on their use of *extrinsic* or *intrinsic* metrics.

Extrinsic information quality metrics focus on the accuracy, completeness, or consistency of the object-based measures by referencing external data sources. Questions about the syntactics (conformity to other collected data; *e.g.*, consistency) or semantics (correspondence to external or authoritative phenomena; *e.g.*, accuracy) are paramount. In contrast, *intrinsic information quality metrics* use features of the target dataset itself to assess quality by examining contexts, reputations, and processes for generating information. Questions about pragmatics (use and interpretation of information; *e.g.*, timeliness or authority) are paramount. This dichotomy, while simplistic, is useful for identifying gaps in existing approaches for measuring information quality, especially in the context of online peer production communities like Wikipedia and OSM.

Quality assessment of spatial information

Though there are many approaches for assessing map quality, those offered by the International Standards Organization (ISO), codified as ISO Standard 19113, are widely accepted. The standard has five primary approaches to assessing quality (Barron et al., 2014), summarized here:

(1) *Completeness*: Is the dataset complete?
(2) *Consistency*: Are the spatial and thematic attributes of the data in a uniform fashion?

(3) *Positional Accuracy*: How accurate are the coordinates of the map objects?
(4) *Temporal Accuracy*: If the data has a temporal element, is it accurate?
(5) *Thematic Accuracy*: Are the quantitative/qualitative attributes of the data accurate?

As we discuss later, each of these dimensions, apart from *consistency*, implements quality assessment as an extrinsic information quality metric by referencing similarity to an authoritative dataset. In some cases, extrinsic measures have used proxy datasets, such as kilometers of road in relation to population density (Mashhadi, Quattrone, & Capra, 2013), in an attempt to assess *completeness* when a suitable source of reference data is not available. *Consistency*, on the other hand, is the sole example of a quality metric in this ISO Standard that uses features intrinsic to the target dataset to assess quality.

2.2. Information quality metrics for peer production

Wikipedia's radical "anyone can edit" model integrating user-generated contributions into an authoritative encyclopedia justifiably raised concerns about the quality of the resulting information. Evaluations of Wikipedia quality emphasize that features such as the quantity of information or the number of links in an article are the most important determinants of end users' trust in Wikipedia content (Kittur, Bongwon, & Chi, 2008; Towne et al., 2013; Yaari, Baruchson-Arbib, & Bar-Ilan, 2011). Despite the major differences in the substantive content of contributors' edits, the technical designs of both the Wikipedia and OSM systems implement analogous methods for merging user contributions into a single canonical version as well as capturing similar kinds of meta-data in revision event logs about user IDs, timestamps, and content versions. This opens the possibility for translating information quality metrics from a well-validated domain like Wikipedia to a less studied domain like OSM. We compare the extrinsic and intrinsic information quality metrics used in prior research on both Wikipedia and OSM below.

Extrinsic information quality metrics

We define extrinsic information quality metrics as object-based measures focusing on syntactic or semantic "correctness" that reference external authoritative data sources. Online peer production systems like Wikipedia and OSM were created to replace authoritative incumbent products like *Encyclopedia Britannica* and government land surveys (respectively) created by expert organizations. Thus, assessing the quality of user-generated information by comparing it to expert-generated counterparts is a natural validation step. Extrinsic metrics for assessing the accuracy of Wikipedia articles have used experts to compare the number of errors in Wikipedia against other works of reference, finding that error rates were similar to or lower than authoritative sources (Giles, 2005; Holman, 2008). Other studies have explored the completeness of Wikipedia's coverage by measuring the representation or overlaps in topical coverage across sources (Brown, 2011; Halavais & Lackaff, 2008; Royal & Kapila, 2009; Samoilenko & Yasseri, 2014).

The first scholarly investigation of OSM's extrinsic information quality assessed the *completeness* and *positional accuracy* of the OSM road network for the United Kingdom as compared to the authoritative Ordnance Survey (Haklay, 2010). Although it was inconsistent, the OSM data compared favorably to the government's dataset and judged to be of good quality. Such findings are consistent with work that examined other geographic locations and employed a wider range of quality measures (Zielstra & Zipf, 2010).

Intrinsic information quality metrics

We define intrinsic information quality metrics as process-based measures focusing on pragmatic or contextual "authority" by examining the processes generating information. Most Wikipedia studies employ intrinsic measures to assess information quality and validate against community-generated labels of article quality (Kane, 2011; Stvilia, Twidale, Smith, & Gasser, 2008; Warncke-Wang, Ranjan, Terveen, & Hecht, 2015). Behavioral features like the number of revisions, the number of revisions from administrative, registered, or anonymous editors, the number of unique editors, number of reverts, and time since last revision are intrinsic characteristics that are easily computed from revision event logs (Stvilia, Gasser, Twidale, & Smith, 2007). Content features, such as word count (Blumenstock, 2008), number of references (Luyt & Tan, 2010), images, and tables (Anderka & Stein, 2012) also provide popular metrics.

Intrinsic measures of data quality are growing increasingly important to assess the quality of OSM data due to the lack of authoritative reference datasets. For many parts of the world, OSM is the most complete geographic dataset. This situation can arise because of a lack of good, official data—as is the case in some developing countries—or simply because contributions from an active local mapping community outpace official survey work. Whatever the reason, the lack of high-quality reference data limits the utility of extrinsic measures of quality in these situations. Barron, et al. acknowledge that "the quality of OSM data also depends on the project's contributors" (2014). Preliminary frameworks exist for evaluating intrinsic quality (Barron et al., 2014), evaluating the consistency of tagging schemes (Anderson, Soden, Anderson, Kogan, & Palen, 2016; Vandecasteele & Devillers, 2015), and investigating "user-centric" quality metrics based on contribution meta-data (Haklay, Basiouka, Antoniou, & Aamar, 2010; Kogan, Anderson, Palen, Anderson, & Soden, 2016; Mooney & Corcoran, 2012). GIScience has recently seen many new intrinsic quality metrics introduced with respect to OSM. Barron et al. introduce a framework discussing 25 measures requiring no external reference sets (2014) with a comprehensive review of current approaches and explanations of intrinsic quality metrics as applied to VGI. More recently, Sehra et al. created an extension for the QGIS open-source software project to allow for easier OSM data analysis, analogous to tools in commercial GIS software (2017).

Barron et al.'s intrinsic quality assessment framework highlights the importance of "fitness for purpose" in quality analysis, and separates the 25 metrics into six distinct categories to connect metrics and indicators (assessments) to specific purposes. For example, road network completeness is an assessment relevant to the use case of routing and navigation (2014). One of the areas Barron et al. consider is "user and information behavior" to include contributor activity as an indicator for quality assessment. They find that the distributions of edits per user are heavily skewed—with a few contributors doing most of the work—a finding that is common among all peer production systems. Importantly, they note while it may be expected that contributors with high edit counts create higher quality data, this thesis remains untested (Barron et al., 2014). Their call motivates the work presented here; we further explore user-based metrics with respect to the number of contributors and their respective expertise. Furthermore, as Eckle and Albuquerque point out, OSM data contributed during disaster mapping events often contains just the raw geometry (that is, an outline of a building or the path of a road) without the contextual information of attributes describing it, making existing intrinsic quality assessment techniques which rely on the object's attributes alone (such as name or type) difficult or impossible (2015).

Given this, intrinsic quality assessment based on contributor metadata becomes the most feasible type of quality assessment available for many areas of the map and this observation motivated our work in developing the metrics presented below; our metrics can be applied to any part of the map, independent of reference datasets or detailed object attributes.

At a high-level, our three metrics are straightforward to understand. Our first metric is a variation on a simple contributor-based metric: the absolute number of contributors that have been active in a region of the map. Our second metric looks at the types of objects that different contributors prefer to edit and the amount of that object type they have edited before. Our third metric looks at the overall editing evolution of a region in terms of what objects are being collectively edited by the contributors. Each of these metrics relies solely on the basic object type and the contribution metadata. This provides the *who, what, when*, and *where* attributes of each edit, and enables investigation of how the map developed in any given region. Our metrics are not replacements for other quality metrics, but rather provide richer context from which to understand the resulting OSM data.

3. Dataset and methods

3.1. Openstreetmap

Started in 2004, OSM is an open geospatial database released under the Open Database License. The main rendering of the database can be viewed as an interactive map on www.openstreetmap.org. The map is also available as a set of tiles through a web service. As a result, OSM is used as a basemap for many interactive web-based maps. The OSM website currently has over 4 million registered users; though less than 1 million users have ever edited the map data. The database has over 4 billion unique geographic points that make up the objects on the map. To illustrate the degree of completeness of the global map, Figure 1 shows just the road network in OSM.

Objects in OSM are defined by a set of *tags*: key-value pairs that identify a country boundary from a park or a bike path from a major street. The objects we focus on are roads and buildings. These are the most common objects in OSM, as well as the most edited objects during disaster mapping. They are tagged as *highways* and *buildings.*

Roads (highways)

In OSM, a road is a geometry known internally as a "*way,*" which is semantically tagged with the key "*highway*" and an associated value describing its relative prominence, such as *primary, tertiary, walkway*, etc. When roads are traced from remote imagery—as is common in disaster mapping—they are rarely tagged with a "*name*" attribute. Indeed, a road with a "*name*" attribute can be considered to contain some level of local, ground-truth knowledge, likely implying higher quality.

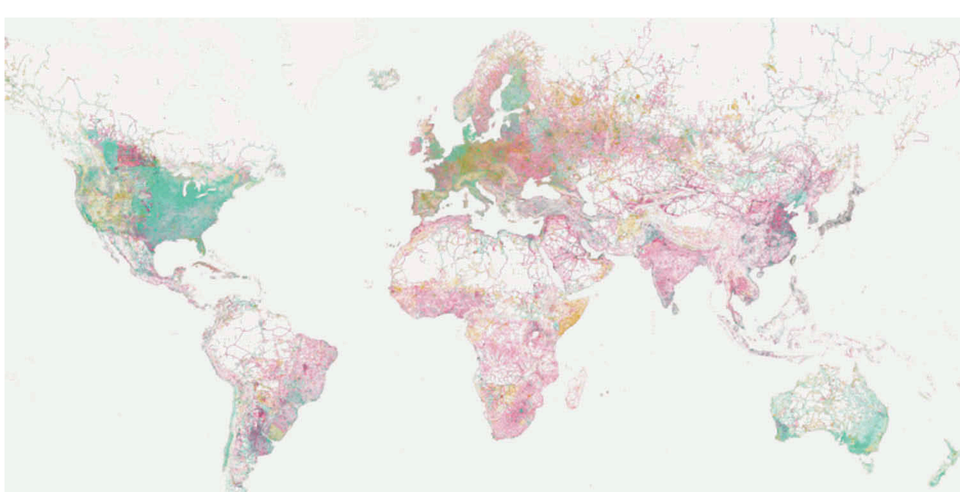
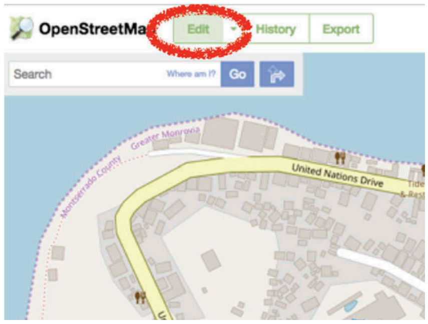
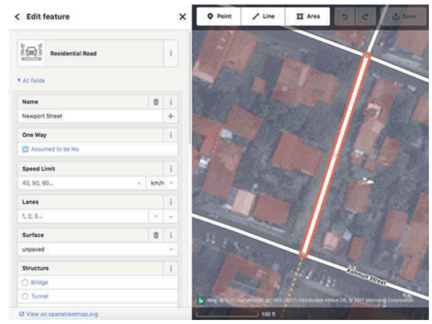


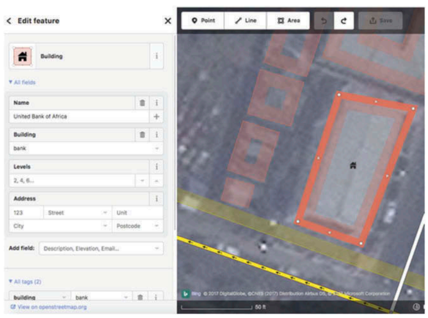
Figure 1. The road network in openstreetmap, showing global coverage and colorized by existence of the name attribute. Cyan roads include a name; magenta or orange roads and paths do not. map data © OPENSTREETMAP contributors.



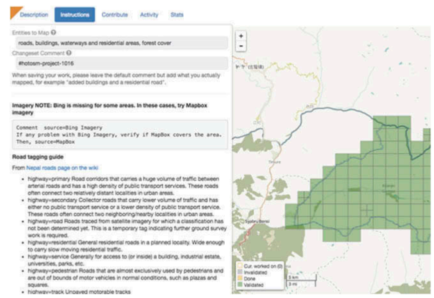
(a) The edit button on openstreetmap.org above the map



(b) Editing a road object with the in-browser editing interface, the iD editor.



(c) Editing a building feature with iD. Suggestions of attributes to add such as number of levels or the specific address are presented on the left.



(d) The Tasking Manager from Humanitarian OpenStreetMap Team (HOT) gives instructions on what to map for a specific disaster event and helps mappers coordinate their efforts. (tasks.hotosm.org)

Figure 2. Editing OSM on openstreetmap.org.

Buildings

A building is denoted by a tag describing its purpose, such as {*'building': 'residential'*} or, in many cases, simply {*'building': 'yes'*}. In OSM, buildings are typically represented by closed ways, i.e., line geometries that have the same start and end points. As of October 2017, *building* is the most common tag in OSM, with over 5.5% of all objects having this tag.[2]

Contributors can edit OSM through an in-browser editor on openstreetmap.org or through stand-alone map-editing tools that communicate directly with the database through the API. Editing OSM is depicted in Figure 2.

3.2. Obtaining OSM data

Obtaining and manipulating OSM editing data is possible through a set of public APIs and downloadable database files. A number of open source tools are available for converting the data between popular geospatial data formats. A format made popular by the web for efficient storage and serving of map data is the *vector tile*. A vector tile stores the geometry, attributes, and metadata for every map feature organized by geographic location. Tiles can be created at various zoom levels, each with a different resolution of data. The tiles used for our analysis are generated at zoom level 12. At this level, the inhabited part of the earth is comprised of about 2.5 million tiles, and these tiles have an area of roughly 100 square-kilometers at the equator.

For each of our analyses below, snapshots of the map at annual intervals from 1st January 2006 to 1st January 2017 are used to achieve annual granularity of the history of the database. We also note that in some cases where the same object was edited multiple times in one year, only the last edit of that year is counted. Some of our reported numbers are therefore an under-representation of the total editing activity in OSM. We use an open-source Javascript framework called tile-reduce to process these vector tiles in parallel.[3] We perform all spatial analysis with open-source GIS tools, and our full data processing pipeline includes a combination of javascript, postgresql, and python.

3.3. Our dataset

To evaluate our three intrinsic quality metrics with respect to the insight they can provide into the practice of disaster mapping, we selected four distinct tiles on the map that have been the geographic focus of disaster mapping events following different kinds of events (see Figure 3). These areas are: (1) Port Au Prince, Haiti, the scene of one of the first instances of major coordinated disaster mapping following the 2010 Earthquake; (2) Tacloban, Philippines, where disaster mappers digitally converged before, during, and after Typhoon Yolanda in 2013; (3) Monrovia, Liberia, a region that was part of a year-long humanitarian-focused

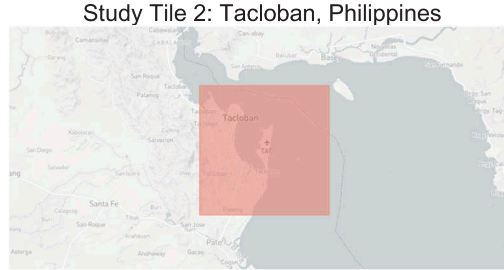
[2]http://taginfo.openstreetmap.org.
[3]http://github.com/mapbox/tile-reduce.

Study Tile 1: Port Au Prince, Haiti

Kilometers of road	1,006 km (54% with names)
Number of buildings	12,141 (7% labeled)
Contributors (all time)	494

In response to the January 2010 Earthquake, hundreds of users contributed tens of thousands of features to the map to aid disaster relief, creating the most comprehensive map of Haiti to date (Soden and Palen, 2014; Zook, et al. 2010).

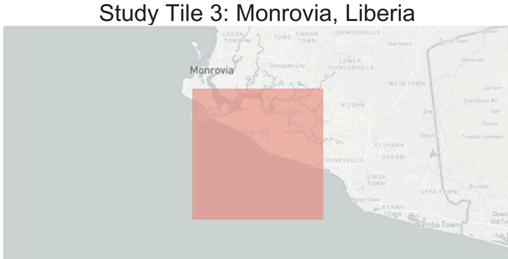
Study Tile 2: Tacloban, Philippines



Kilometers of road	257 km (35% with names)
Number of buildings	29,573 (71% labeled)
Contributors (all time)	371

Typhoon Haiyan (Yolanda) struck the Philippines in 2013, prompting hundreds of contributors to improve the map in the Tacloban region, specifically updating buildings on the map for later damage assessment (Palen et al., 2015).[4]

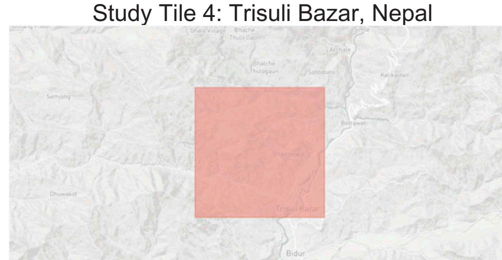
Study Tile 3: Monrovia, Liberia



Kilometers of road	174 km (32% with names)
Number of buildings	19,193 (6% labeled)
Contributors (all time)	202

In 2014, The Humanitarian OpenStreetMap Team helped coordinate disaster mapping efforts in West Africa in response to the Ebola Outbreak.[5] *This activation lasted many months, making it distinctly different from the rapid convergence of contributors on the other three tiles.*

Study Tile 4: Trisuli Bazar, Nepal



Kilometers of road	324 km (3% with names)
Number of buildings	7,596 (16% labeled)
Contributors (all time)	257

The largest disaster mapping event to date, thousands of contributors were active in Nepal in response to the April 2015 Nepal Earthquake (Poiani et al., 2016).

Figure 3. Details of the four study tiles selected for contribution-based intrinsic quality analysis. Data retrieved at the beginning of 2017.

mapping project to help relief and prevention efforts during the 2014 Ebola outbreak; and (4) Trisuli Bazar, Nepal, a region heavily impacted by the 2015 Earthquake.

At the time of these events, each of the associated disaster-mapping activations was the largest to date in terms of number of contributors. Study Tile 3 (Monrovia, Liberia) is different from the rest because the activation in response to the ebola outbreak was not a single, rapid convergence of contributors, but rather a long-term project that saw thousands of volunteers over a period of months. In comparison, the other events saw a period of rapid mobilization as contributors converged on OSM in immediate response to the natural hazard. We expect to see distinct differences then in our results between these regions. The mapping tasks are similar across all the events: for disaster mapping, tasks focus on performing detailed mapping of buildings and roads in specific regions.

For quality comparison, we have chosen two well-validated areas of the map: London, UK and Heidelberg, Germany. Previous extrinsic quality research found that these tiles are of high quality when compared to external reference datasets (Arsanjani, Barron, Bakillah, & Helbich, 2013; Haklay et al., 2010). We compare the study tiles with these high-quality tiles for each metric. The differences suggest that our metrics are capturing contribution patterns unique to disaster mapping.

Though today these tiles may appear complete, our metrics aim to expose the differences in the histories of how the data were contributed. For each metric, we discuss the specific implications the findings may have for measuring intrinsic information quality in VGI.

4. Contributor-based intrinsic quality metrics

We extend one existing intrinsic quality metric and propose two new intrinsic information quality metrics for VGI. These metrics apply to vector tiles of OSM data. Our metrics explore attributes of the data beyond geometries and visible properties; instead, they

[4]http://wiki.openstreetmap.org/wiki/Typhoon_Haiyan
[5]https://wiki.openstreetmap.org/wiki/2014_West_Africa_Ebola_Response.

examine features specific to *peer-produced* spatial data. This includes information about a contributor's previous experience with the platform for each individual contributor and the time when an object was last edited. Specifically, our metrics are:

1. Contributor Density Over Time

How many users have been active on a given part of the map? Denser maps should have higher quality as more people have been active in the area. This is a straightforward measure that was first explored by Haklay et al. (2010). Our extension to this measure focuses on temporality, looking at the cumulative density over time and marking when the bulk of contributors were active.

2. Contributor Experience

How long has a contributor been active in the OSM community? What types of objects have they mapped before? We expect that areas with experienced contributors should have higher quality. This metric works to supplement the straightforward measure of contributor density by further inspecting who the contributors are. The need for such a metric becomes especially important when considering mapping events that attract newcomers. This measure asks: "Who does a majority of the work: many new contributors, or fewer experienced power users?" Depending on this distribution, the cumulative number of contributors per square kilometer may not be as important.

3. Tile Maturity

How is the composition of objects changing over time? Areas where contributions are focused on maintaining existing features instead of adding new features may have achieved some level of completeness, itself a quality measure. Instead of examining qualities of individual contributors, this measure instead considers *collective editing activity* by looking at the bulk of types of edits in a region over time.

4.1. Metric 1: Contributor density

In one of the first intrinsic quality studies of data quality in OSM, Haklay et al. found that after 15 mappers have been active in a given square-kilometer, the positional accuracy below 6-meter resolution is "very good" in comparison to government data (2010). This study also revealed that the first five mappers to an area make the greatest impact to the positional accuracy of the data. This contributor-density method draws inspiration from Linus's Law of open source software development: "given enough eyeballs, all bugs (in software), are shallow." For OSM, the contributor-density method assumes that more mappers contributing to an area provides a greater chance that some level of data validation and quality assurance has been achieved (Haklay et al., 2010).

Globally, less than 1% of zoom-12 tiles reach Haklay's threshold of 15 contributors per square kilometer. When we examined our tiles, we found that both Port Au Prince and Trisuli Bazar reached this threshold during their respective disaster mapping events (see Figure 4). This initially suggests that the quality of these tiles became "very good" as contributors mobilized in response to the event. The spikes in contributor activity at the time of the event for Tacloban and Monrovia are significant and represent the most activity ever to occur on these tiles, but still do not reach this particular threshold of 15 contributors per square kilometer. Figure 4 also shows the density of contributors in London and Heidelberg, which surpassed 15 users/km^2 in 2008 with steady growth of an active OSM community since.

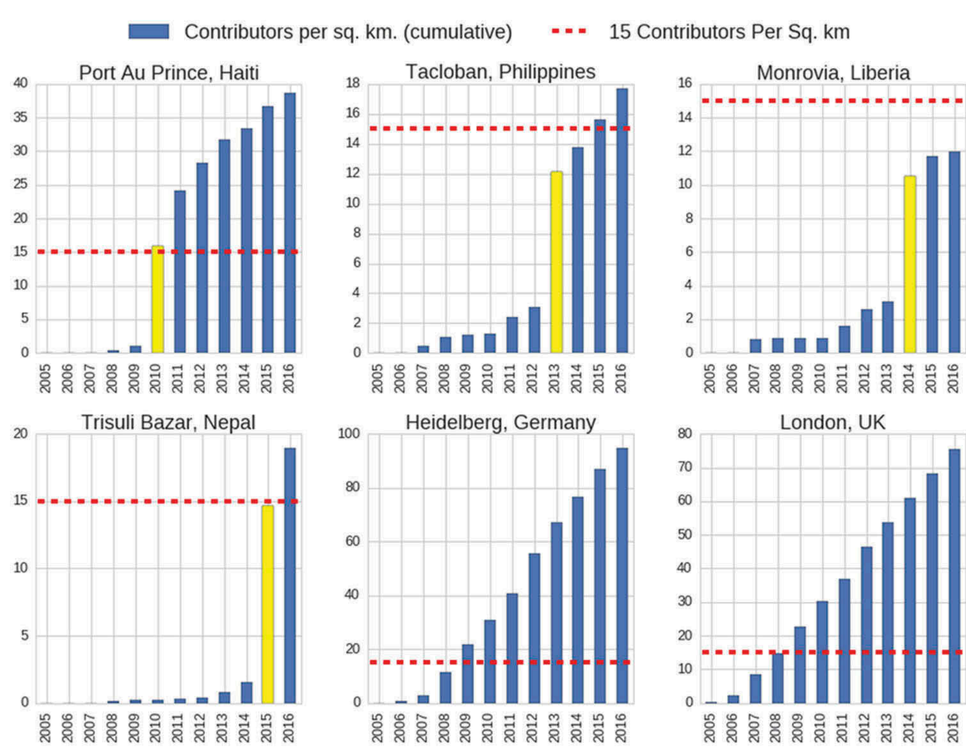


Figure 4. Density of unique contributors by tile over time (cumulative—in users/km^2). Event year is highlighted in yellow.

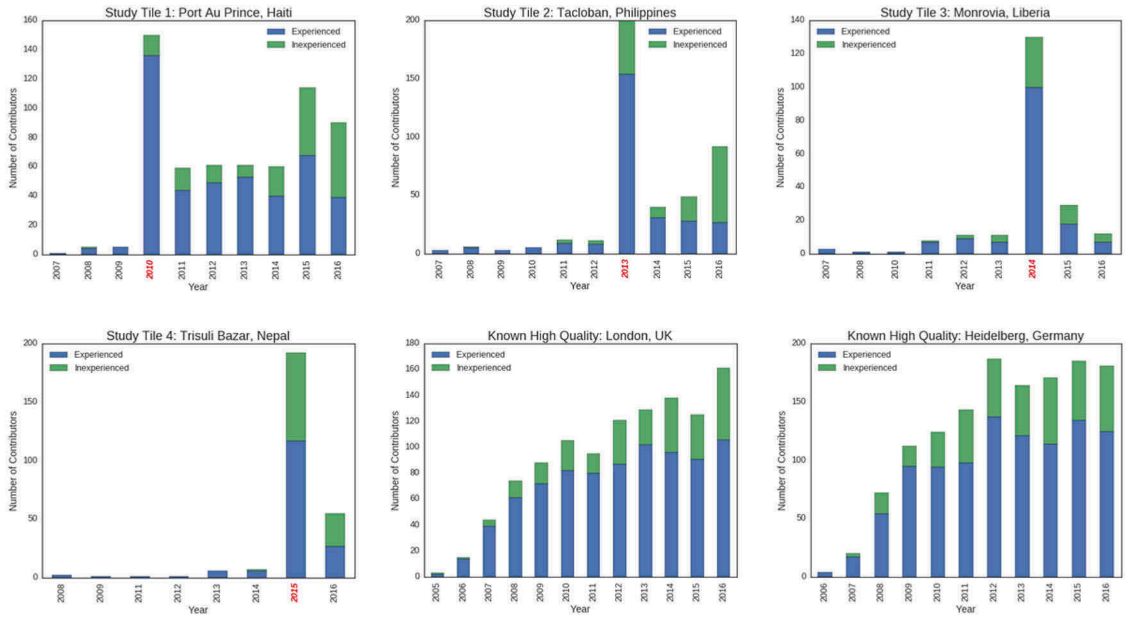


Figure 5. Users active each year on the study tiles and two known high-quality tiles for comparison. The years of disaster events for our four study tiles are labeled in red. Inexperienced and experienced users are denoted by color. Metric 2 explores these differences.

Figure 5 looks beyond cumulative density to contributor count over time to reveal the rate of growth for the number of distinct contributors. As expected, the time of the disaster-mapping event creates the most significant spike in contributors for each tile. This spike shows many contributors active during a relatively short amount of time and then never returning to edit in this area. This may lead to the *staleness of the map data* (discussed later).

In contrast, London and Heidelberg show the sustained growth of a contributor community. These communities grow steadily from the beginning and seem to level off in recent years, perhaps suggesting a level of saturation of contributors in the region. Knowing these tiles are of high-quality suggests that a sustained, growing community of contributors is a positive quality indicator for the map overall.

Port Au Prince continues to maintain an active community in the years following the earthquake, significantly larger than the contributor activities of the other tiles; this is likely the result of the work of a local mapping community group, Comunite OpenStreetMap de Haiti (COSMHA), which formalized and incorporated as part of the response to the earthquake (Soden & Palen, 2014). This sustained community of contributors has positive quality implications for the resulting map.

In contrast, an indicator of potential lower quality as a long-term result of these rapid, single mobilizations of contributors is *staleness* of the data. Figure 6 shows that six years after the event, the Port Au Prince tile has many features still tagged as *building* = *collapsed* which have not been edited since the earthquake. While these buildings may have not been rebuilt and are indeed represented accurately in the database, we cannot know for sure without more recent timestamps in these edits.

Implications for assessing information quality

This metric shows that areas that have experienced the rapid mobilization of contributors during disaster mapping events may superficially satisfy quality measures based on density of contributors with one-time contribution activity. Quality evaluations need to take into account the previous editing context and consider the amount of sustained editing activity, which requires new contributor-density measurements over time. In this vein, Haklay (2010) and Barron et al. (2014) warn that OSM quality evaluations should be localized and performed with "fitness for purpose." In the cases under study here, the purpose was to create roads and buildings data where there previously was none, and for immediate use. This is a different type of mapping activity than a local community performing sustained, detail-oriented mapping. Quality evaluations of these data need to then be aware of these generative differences in the map so as to evaluate the data within context.

Furthermore, the timing of these contributions raises the question of staleness as well. Our first metric expands on previous work by considering the age of the contribution (Barron et al., 2014). Overall, this metric is simple, yet powerful, because the results seem intuitive and can locate areas of the map where high numbers of contributors (relative to others) have been active, and moreover, how long they were active.

4.2. Metric 2: Contributor experience

Our second metric expands quality investigation to the amount of editing experience a contributor has with the objects they are editing. Note: our use of the term "experience" refers to a user's familiarity and expertise with the OSM platform. The relationship of contribution experience to map quality has been explored by a variety of methods, but most commonly it is defined by the *number of edits that a mapper has made* (Neis & Zipf, 2012). We explore a new notion of experience in terms of *days active on the platform*. Barron et al. remind us that while it seems plausible that editors with more contributions create

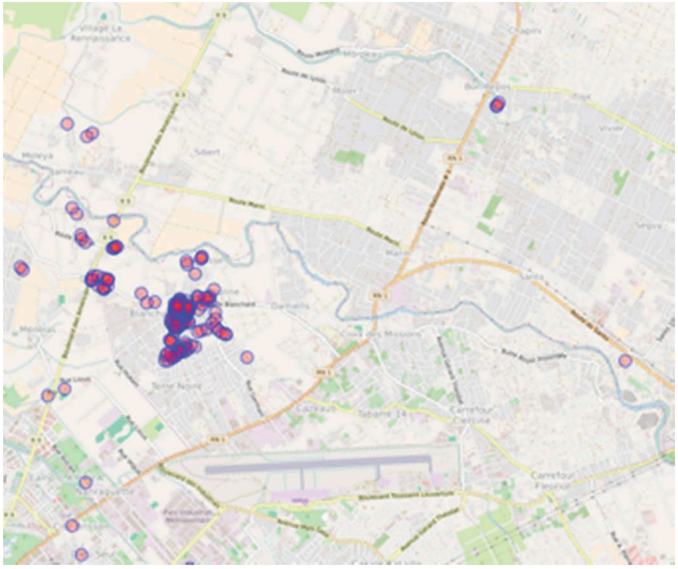


Figure 6. Features tagged as building = collapsed in port Au Prince.

higher quality data, this has not been formally evaluated (2014). For the purposes of this metric, we take a slightly modified approach to the notion of experience by classifying users with seven or more days of activity as "experienced," and users with less than seven days as "inexperienced." Because over half of all contributors have only made one edit while other contributors have made millions, the distribution of editing days per user is highly unequal and non-uniform. We empirically selected seven days because it retains an approximate log-normal distribution of edits per user, consistent with other online communities. This threshold retains 97.7% of the total edits but only 13% of the users for global OSM editing. We find this definition of experience more illuminating than previous definitions because it takes into account sustained interest and activity in OSM. A contributor active for only a weekend mapping event may create a lot of data, but has less overall experience with the platform and community norms than a contributor who has been active for more days. For this research, then, we take the equivalent of a week-long experience with the platform to be a useful minimum for understanding a range of basics about the platform and the OSM community, based on our experience with training others on OSM. However, this is a flexible variable that can be chosen at different thresholds for other purposes; we chose seven days for the models here. How definitive a line between new and experienced is drawn at this threshold is an area of active research.

Referring back to Figure 5, we see a difference between experienced and inexperienced contributors per year. For each study tile, activity spikes consistently have more experienced contributors than inexperienced. This suggests that more experienced contributors participate in disaster mapping activations than inexperienced. This has important quality implications for the data contributed during these events: specifically that these data are likely of good quality because the contributors have previous editing experience. However, the ratio of inexperienced contributors increases with each event from nearly 10 experienced users for every inexperienced user active in 2010 in Port Au Prince, to 1.6 experienced users for every inexperienced user active in Nepal in 2015. This suggests that more new mappers are becoming involved in the disaster mapping community. While this is encouraging for the overall growth of the larger OSM community (Dittus, Quattrone, & Capra, 2016), it comes with the potential that recent and future events may include more and more data from first time contributors not yet aware of specific editing or community norms. Observations of the OSM mailing list during the Nepal earthquake confirm that experienced mappers were frustrated that new mappers were not following community norms and creating square buildings.[6] To combat this, OSM editing tutorials are constantly being developed, updated, and customized for different disaster events, such as learnosm.org.[7]

We next look at what types of objects contributors have edited to explore a richer notion of experience with the OSM project. This measure assumes that, with time, a contributor's proficiency in editing specific object types improves. We look specifically at contributor preferences for mapping buildings and roads. Figure 7 shows editing habits of all OSM editors by object type. The number of buildings and road kilometers edited is calculated for all contributors and then plotted against one another. The color represents the number of contributors having edited <x> kilometers of road and <y> number of buildings. The legend on the right matches color to number of users.

The majority of the activity lies along the x- and y-axes near the origin, indicating that most users edit (1) very little, and (2) only one type of object or the other, not both. The lighter trend down the diagonal indicates that, as contributors edit more (and therefore become more experienced), their preferences for one object over the other may fade and they map both types of objects, though the majority of contributors do not exhibit this behavior. This distribution is consistent with *power contributors* in peer production systems like Wikipedia (Laniado & Tasso, 2011). This prompts the question for the quality of our study tiles: are the ratios of buildings and roads edited by power contributors versus others higher or lower than other regions of known high-quality?

Figure 8 shows the differences in object editing experience among contributors and their respective number of edits, an indicator of their experience with this object type. For each study tile, we also show the distribution for London and Heidelberg for comparison (the faint red and green dotted lines). The similarities between the distributions for London and Heidelberg suggest that this shape of distribution may yield good quality. On both tiles, we see that contributors with experience editing over 1,000 buildings map over half of all the buildings for each region. There are both positive and negative quality implications here. Fewer more-experienced users doing the bulk of the editing suggests specific expertise, but limits the amount of crowd validation that may occur (referring back to our first metric).

[6]May 2015 HOT mailing list archive https://lists.openstreetmap.org/pipermail/hot/2015-May.txt.gz.
[7]learnosm.org is an open source project maintained by the HOT and OpenStreetMap communities.

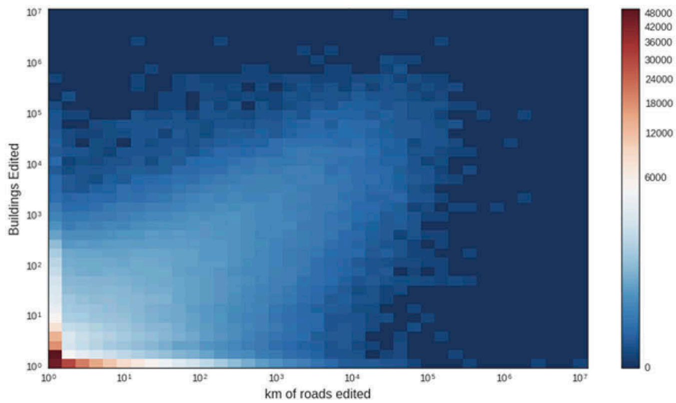


Figure 7. Editing preferences among OSM contributors.

In both Port Au Prince and Trisuli Bazar (Tiles 1 and 4), the distribution for buildings differs significantly from the known high-quality tiles. In these cases, less-experienced contributors edit a higher percentage of the total buildings. In Tacloban, however, this trend is the opposite, with more-experienced building mappers performing the bulk of the building edits.

Study tile 3, Monrovia, has the most similar distributions to the known high-quality tiles. This is fitting because the particular disaster mapping event consisted of a sustained mapping activity by an engaged mapping community over a longer period of time. This mirrors the engagement of an active local mapping community, as seen in Heidelberg and London. Across all of the study tiles, there is no notable difference in the distribution of road mapping experience and the amount of roads mapped. Further analysis is required to identify the differences here.

Implications for assessing information quality

If most of the buildings or roads in an area were created by contributors without any prior experience creating those kinds of objects, then one may be suspicious of the quality of that section of the map compared to other areas where the majority of an object type is edited by contributors with prior experience working with that object type. On the other hand, if just a few power contributors have edited most of the objects, fewer eyes have seen this part of the map, lowering the potential for more validation opportunities.

Ultimately, the differences in these distributions cannot definitively say that one tile is of higher quality than another. However, the similarities in the distributions for our two high-quality tiles may suggest a target distribution of experience versus amount of objects mapped that yields a good quality map. Departure from this distribution would then have implications for the quality of the final map, though comparing to only two high-quality tiles is not sufficiently representative to make this claim definitively. Future research should expand this study of high-quality regions to achieve a statistically significant target distribution from a larger sample of known high-quality tiles. For now, however, there is no denying that the distributions of experience with mapping buildings to the amount of buildings mapped during a disaster is distinctly different in regions that have been the subject of disaster mapping activities with a rapid convergence of contributors, for better or worse.

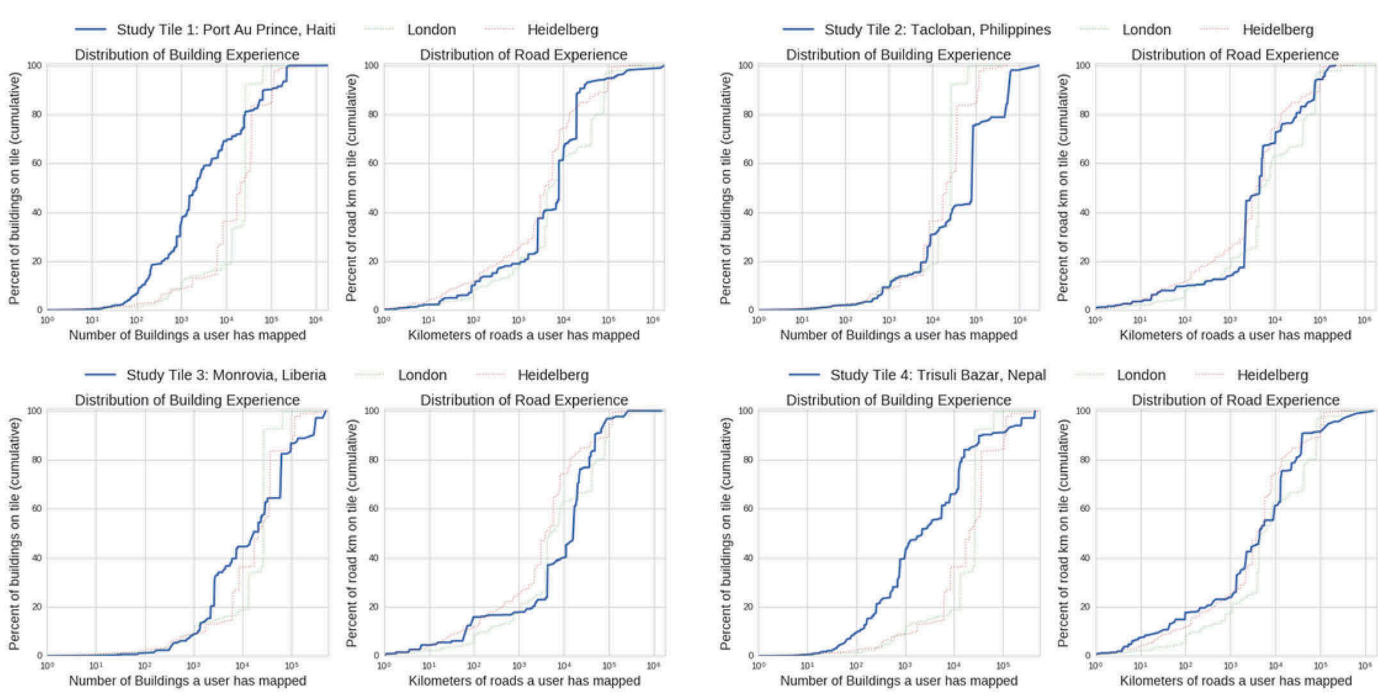


Figure 8. Percent of buildings and roads edited on each tile versus the number of buildings or kilometers of roads a user has mapped (experience). Thick blue lines represent object-level experience (*X*-axis) per cumulative amount of total edits to that object on the study tiles (*Y*-axis). The faint lines represent the same values for our high quality comparison tiles. Differences between these distributions highlight the differences in the amount of editing experience among contributors and their contributions.

4.3. Metric 3: Tile maturity (stages of growth)

The current version of the OSM database is the aggregate product of hundreds of millions edits from hundreds of thousands of users. Our third metric, what we call a tile maturity measure, breaks down the types of edits that occur in an area over time to identify distinct *stages of growth*. By looking at both the object type and the timestamp, we can identify several distinct stages of editing behavior that the map progresses through. These stages include the creation of new roads, the addition of new buildings, and, finally, a maintenance phase, where less new data is added and the bulk of contributions are edits to existing objects. In general, we know that the map grows from the road network outward (Ciepłuch, Mooney, & Winstanley, 2011). The maintenance period has been called "map gardening" (McConchie, 2013), in which continued editing of existing map objects, versus the creation of new ones, becomes the characteristic pattern of editing.

For comparison at a macro scale, we computed these stages of growth for the United States in OSM: While the number of edits continues to grow, the map does not fill in proportionally by object. In the US, 40% of the total road editing activity done to date was completed by 2009 (largely the product of a massive import of road data conducted in 2007/2008). However, it was not until five years later that buildings caught up and 40% of the total building editing activity was complete. In the last two years, only 10% of the total road editing activity has occurred, but more than 50% of the edits to buildings have taken place. There is a clear trend of roads being added first, and while these roads continue to be maintained, contributors in the US are currently in a *building* phase. We, among others, find this pattern to hold in general for OSM globally (Ciepłuch et al., 2011).

Figure 9 shows the breakdown of new roads and buildings in comparison to editing of existing objects for each of our study tiles through the years. Across every tile, we see agreement that the first stage is the creation of roads. As new road activity subsides, there is a rise in the amount of new buildings. Port Au Prince (study tile 1) appears to currently be in a *building phase*, where the majority of edits in the past couple years have been the creation of new buildings. However, the years after the earthquake show a majority of maintenance activity, likely editing and maintaining data produced during the event. This creates a false sense of completeness where one may expect the building phase to be over. As evidenced by the new building activity occurring in the last two years, however, the region is not actually in a maintenance phase, but instead back in a building phase.

Similarly, Study Tiles 3 and 4 (Monrovia and Trisuli Bazar) both appear to be in a maintenance phase. With their respective disaster mapping activities occurring more recently, it is unknown whether this current maintenance phase is the product of editing the features created during the event (similar to Port Au Prince), or if the region indeed has reached some level of building completeness and has naturally entered a maintenance phase. In both cases, the types of edits occurring during the event nicely match the type of tasks outlined by HOT, which was to add buildings to the map.[8] And in the case of Trisuli Bazar, also perform "detailed mapping" of the area.[9] These maintenance phases likely represent the annotation of descriptive tags to features created by remote mappers during the event. It is still unclear, however, whether the tiles will enter another building phase in the future, as we have seen with Port Au Prince.

Study Tile 2, Tacloban, on the other hand, saw many new buildings, but mostly editing of existing features during the year of the event, prompting further questions about the exact disaster-mapping activity. Furthermore, the tile parallels Study Tile 1 by appearing to enter a maintenance phase after the event (though with a surprising number of new roads), and is currently going through another building phase.

This potentially premature maintenance phase is common across all these regions, making the tiles appear more complete than they are, relative to other parts of the map that appear to progress through the phases of growth in an orderly fashion. However, these regions are still significantly better mapped now than they were, having been the target of disaster mapping. For comparison, the stages of growth are shown for Heidelberg and London. Heidelberg clearly follows the standard trend with maintenance behavior increasing in recent years as both building and road creation slows down. London has seen increased building activity in recent years, but still follows the general trend of maintenance behavior being more common in recent years than new roads.

Implications for assessing information quality

Given the specific order in which the map grows and matures, knowing which phase of growth a given part of the map is in gives an indication of its level of completeness (a standard quality measure). Determining these phases strictly on percentages of edit types requires neither external reference data nor specific object attributes, merely the geometry type and version number. This makes analysis of any region possible as these are basic attributes present in every map object.

Disaster mapping activity, however, interrupts this natural sequence, making the map appear to be in a different stage than it likely is. Our metric is good for showing relative tile maturity between different regions, but the context of a region is important to consider. Comparing the apparent stage of growth with the specific tasks outlined for a disaster mapping activity can provide this context. Ultimately, these stages of tile maturity are relatively easy to compute for any region of the map and offer a measure of object-level completeness, a metric that is typically only possible with extrinsic quality analysis relying on an external reference dataset.

5. Discussion

Information quality is an important concern for online peer production systems like Wikipedia and OSM, especially in safety-critical situations. Despite the similarities in the

[8] https://wiki.openstreetmap.org/wiki/2014_West_Africa_Ebola_Response.
[9] http://wiki.openstreetmap.org/wiki/2015_Nepal_earthquake.

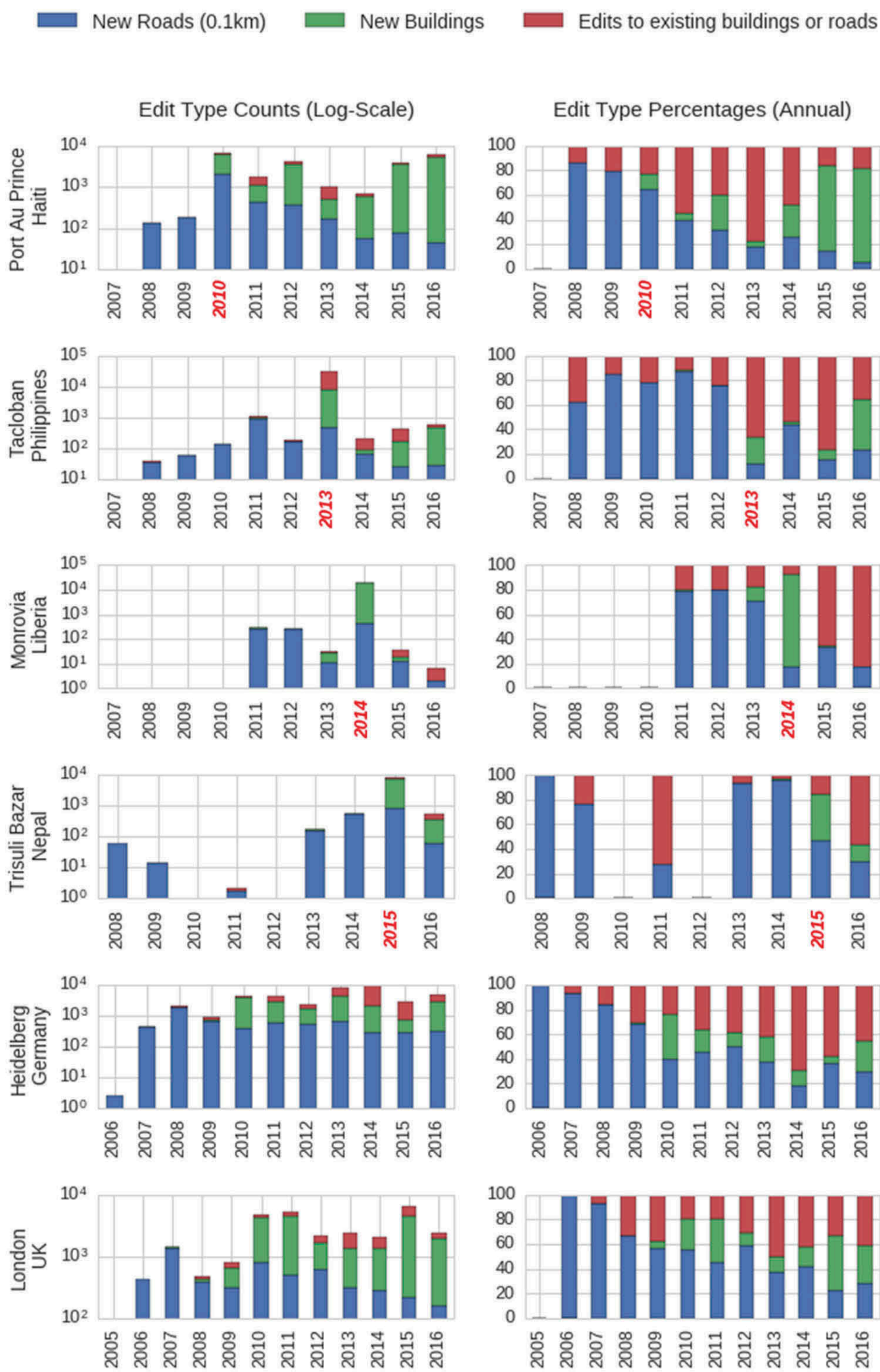


Figure 9. Stages of growth as shown by edits of each type each year. The *Y*-axis representing edit-counts are log-scaled to allow non-disaster event years to show. The percentages are shown on the right to better express the relative amount of activity. For study tiles, event year is denoted with red, italic label.

systems' affordances, the well-validated contributor-based intrinsic metrics for assessing information quality in Wikipedia have not been translated into OSM. While other intrinsic quality assessment techniques relying mainly on the spatial attributes of the target dataset have been explored for OSM within the field of GIScience, we presented three metrics using VGI meta-data about *who* made spatial contributions and *when* to develop alternative perspectives for intrinsic information quality than what is found in related work (Barron et al., 2014). We find this shift in emphasis from the spatial attributes of VGI data to contributor information in turn establishes a bridge from the GISciences into the fields of social computing and human computer interaction.

These new metrics are especially important in understanding the quality of map data produced from a large mobilization of contributors during disaster mapping. Because these data are created for use for disaster preparedness, response and recovery, having ways to assess map quality becomes a safety-critical task. The intrinsic quality metrics offered here rely on metadata about contributor activity that, as opposed to other approaches, are likely to be available in disaster mapping scenarios. They have been tested against four distinct areas of the map that were the sites of large mobilizations of volunteer mappers. These metrics exposed differences in the contributor activity between each of these areas and areas of the map known to be of very high-quality and not impacted by disaster events. The variation in the results suggests these metrics capture distinct generative processes that have implications for assessing the quality of the final product.

Metric 1 showed that while the number of contributors active in a region may indicate the size of the OSM community with direct correlation to the quality, events that draw many remote contributors to the area artificially inflate this density with one-time activity. While contributor density has been shown to be a useful intrinsic measure of quality (Haklay et al., 2010), we show that it is important to also include the temporalities of these contributions in quality assessment. Metric 2 reveals that mapping done by power contributors looks different in areas with sustained and active OSM communities than in areas experiencing the rapid convergence of digital volunteers. In terms of buildings, power contributors had less influence over the total edits in Port Au Prince and Trisuli Bazar than they have in regions with more continually active contributors. It should be noted that both these events were earthquakes—that is, sudden onset events—prompting a rapid convergence of contributors. Metric 3 reveals that disaster mapping activity may disrupt the natural evolution of the map away from the distinct phases of editing, creation, and maintenance.

Given the fundamental difficulty of extrinsic quality assessments of spatial information, intrinsic quality metrics used with other features help identify nuances in the different processes for generating peer-produced spatial information. Ultimately, each of the regions we investigated become better mapped than they were before as a result of the volunteer contributions, but as discussed above, this process played out in unique ways across each site. By combining these metrics, users of the map data can develop a richer understanding of exactly how the map came to be, such as understanding how stale the data may be due to a one-time very active community or learning about the specific expertise breakdown of the contributors. As we have shown, and as with traditional metrics of data quality, none of these metrics convey uncontestable assessments of data quality. Rather, they are intended to be used in combination with other measures to provide historical context of the editing in the region to help better understand the evolution of the map. Additionally, these analyses must be performed with a consideration of how the data will be used (Barron et al., 2014). This is further complicated when considering time- and safety-critical applications of the data such as emergency response. Ludwig et al. suggest that in emergency situations, notions of general information quality assessment are less important than the specific fit and purpose (emergency use) of the information itself (2015). Referring back to Figure 6, the "staleness" of the data today and therefore its potential to lower overall information quality for the area seems a worthy tradeoff for the value that data held during the specific emergency task for which it was contributed in 2010.

5.1. Implications for practice and design in disaster response and beyond

Authoritative data sources that can support extrinsic approaches to assessing VGI quality are often difficult to obtain outside of advanced industrialized countries. In the absence of objective ground truth, examining how user behavior and temporal context interact to generate data can identify gaps. Because these metrics only rely on the OSM database and not external sources, they can be used immediately to help disaster mapping efforts better understand the contribution patterns. Who is editing the buildings? How much experience do they have? These represent real concerns; discussion occurring on the HOT's mailing list during the Nepal earthquake response highlighted frustrations of experienced mappers over the non-square buildings being mapped by new users that cost valuable volunteer time. Our metrics could help organizers of disaster mapping activities more quickly inform their volunteers as to what is happening and/or prompt intervention where it may be most helpful.

Intrinsic methods also allow for identification of stale data in the map, requiring only the date of the most recent edit. This type of analysis could inform contributors where they should focus validation efforts. As we have shown, even in places where the map appears relatively complete, there may be stale artifacts that degrade map quality. The scale and complexity of these data coupled with the fundamental difficulty of establishing extrinsic quality for spatial information also suggests that developing and validating intrinsic quality metrics will also be essential for filtering out vandalism and attacks. Consider a map tile rapidly accumulating edits from novice or non-local contributors: Is this an instance of coordinated vandalism or disaster response? Automatic, algorithmic approaches to vandalism detection have yet to be perfected and similar approaches on Wikipedia have distorted behavior in the community and discouraged new contributors (Geiger & Halfaker, 2011; Halfaker, Geiger, Morgan, & Riedl, 2013).

5.2. Limitations and future work

Our methods are currently limited to the resolution of the specific OSM vector tiles as they are generated, both in temporality (annual snapshots only count the latest edit to an object per year) and in size (zoom level 12 may be too big to identify more spatially nuanced editing activities). Computationally, however, this approach utilizes advanced methods for parallel processing of the massive OSM database, making analysis faster and more scalable than previous methods. Because these techniques use a contributor's editing history, having entire histories instead of annual snapshots will be more accurate in the future, though this is currently an unsolved problem at scale for this domain.

Furthermore, there are currently no scalable methods of tracking over-written geometry changes. For example, if an editor squares up all the buildings in a region or slightly moves the path of a road to better match updated satellite imagery without changing other attributes of the building or road—a common type of edit—the database remains unaware of the change at the object level. That is, if only the spatial geometry of a complex feature like a road or building are changed, the change does not propagate to the object itself. Due to the data structure, identifying and tracking these activities is non-trivial and no solution exists yet for performing this at scale. These types of edits represent validation and correction and their existence has major implications for the quality of the map in that region. Incorporating such features in future research is paramount to better intrinsic quality assessments.

As indicated by a growing number of contributors with each subsequent event, data contributed to OSM in disaster mapping situations will become more prevalent. In general, this will improve the overall completeness of the map. These mapping activities help attract new members to the OSM community, create large amounts of open geographic data, and most importantly, help to satisfy the informational needs of emergency responders. As data contributed in these events become more common in the OSM database, future work could explore more longitudinal questions of community engagement and maintenance of the affected regions. For example, Dittus et al. present a study of 26 disaster mapping campaigns that sheds light on contributor engagement (and retention) across different types of disaster mapping events; of specific relevance to this work, they propose quality metrics based on data persistence and quantify user expertise and engagement using methods proposed in (Geiger & Halfaker, 2013) around the concept of an editing session, not simply number of edits or editing days (2017). Knowing that the percentage of newcomers is increasing with each disaster mapping event, more and more of the map will be the product of novice editing. Future map data quality research could further examine the correlation between new mappers and data quality across more events. Dittus et al. find that the success of these events is not dependent on the large number of novice mappers because novice mappers work slower and produce less data on average (2017). At the same time, a novice mapper that joins for a disaster event and remains part of the community inevitably becomes a more experienced contributor. While the actual amount of data contributed per mapper will vary, future work could investigate if the level of experience (and volume of contributions) per returning contributor is increasing at a rate greater than novice contributors are producing data. This would lead to a population of disaster mappers with community mapping characteristics of a non-disaster contexts like those of London or Heidelberg discussed here.

Thus far, this work is rooted in exploring metadata of VGI contributions to expand more traditional VGI quality assessment methods. Another direction is to build from quality assessment techniques in other forms of user-generated content independent of geospatial data such as social media posts. Reuter et al. discuss the implementation and usefulness of a social media API that incorporates post-specific metadata to perform quality assessment of the data based on a variety of data-use cases (2017). Future work along this vein could incorporate more social media research techniques: network analysis, content analysis, sentiment analysis, etc. that are independent of the geospatial information. Moreover, new technological solutions to improve coordination of these disaster-related crowd-sourced and peer-production activities were not discussed in depth here, current work in this domain, such as (Ludwig, Kotthaus, Reuter, Van Dongen, & Pipek, 2017) present novel methods to ensure coordination among volunteer responders to disaster events, even in the presence of network outages. Such systems are invaluable to communities of disaster volunteers with many quality implications for the data produced.

Future work may also provide valuable insight to the fields of Crisis Informatics and VGI by exploring potential theoretical and methodological consequences of these types of comparisons to community behaviors (and the metrics) to peer-production in non-disaster contexts.

6. Conclusion

The openness and availability of VGI presents new opportunities to use spatial data for applications including in essential humanitarian and safety-critical situations where rapid availability of high-quality data is paramount. We draw from the peer-produced OSM database to propose and evaluate three intrinsic quality metrics for spatial data based on the provenance of these data that build upon user behavior and temporal context. These metrics are not introduced in opposition to or replacement of existing quality assessment methods that respond to traditional concepts of quality, such as positional accuracy, map completeness, or the other ISO 19113 standards. The intrinsic measures presented here can instead *expose specific aspects of the map's history that can provide context*—especially useful when assessment by comparison is not possible. Moreover, these metrics are especially suited for identifying small and sometimes hard-to-detect changes to the map in regions that are affected by rapid disaster mapping. For safety-critical situations of disaster, where humanitarian decisions are based on maps being read by outsiders converging on an area to help, a suite of intrinsic measures that strive to communicate peer-produced map quality from the inside out, perhaps in real-time, is essential. If we anticipate that peer production platforms will continue to populate our future information environments, and certainly in times and places like disaster when convergence of information is a natural and age-old socio-behavioral phenomenon, then attention to developing rapid metrics of quality for digital data generated under socially distributed conditions will ascertain how much risk is assumed when life-and-limb decisions must be made upon them.

Acknowledgements

We thank our reviewers as well as our colleagues and collaborators at the University of Colorado Boulder and Mapbox for their time and valuable feedback.

Funding

This work is made possible with funding from US NSF Grant IIS-1524806 and a research fellowship between the first author and Mapbox.

References

Anderka, M., & Stein, B. (2012) A breakdown of quality flaws in wikipedia. In *Proceedings of the 2nd Joint WICOW/AIRWeb Workshop on Web Quality* (WebQuality '12). ACM, New York, NY, USA, 11–18.

Anderson, J., Soden, R., Anderson, K. M., Kogan, M., & Palen, L. (2016) EPIC-OSM : A software framework for openstreetmap data analytics. In *Proceedings of the 49th Hawaii International Conference on System Sciences (HICSS 2016*), 5468–5477.

Arsanjani, J. J., Barron, C., Bakillah, M., & Helbich, M. (2013) Assessing the quality of openstreetmap contributors together with their contributions. In *16th AGILE International Conference on Geographic Information Science*. (p. 14.-17).

Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic openstreetmap quality analysis. *Transactions in GIS*, *18*(6), 877–895. doi:10.1111/tgis.12073

Benkler, B. (2006). *The wealth of networks: How social production transforms markets and freedom*. New Haven, Connecticut: Yale University Press.

Blumenstock, J. E. (2008) Size matters: Word count as a measure of quality on wikipedia. In *Proceedings of the 17th international conference on World Wide Web* (WWW '08). ACM, New York, NY, USA, 1095–1096.

Brown, A. R. (2011). Wikipedia as a data source for political scientists: Accuracy and completeness of coverage. *PS: Political Science and Politics*, *44*(2), 339–343.

Ciepłuch, B., Mooney, P., & Winstanley, A. C. (2011). Building generic quality indicators for openstreetmap. In Proceedings of the Nineteenth Annual GISRUK Conference, Portsmouth, England.

Dittus, M., Quattrone, G., & Capra, L. (2016). Analysing volunteer engagement in humanitarian mapping: Building contributor communities at large scale. In *Proceedings of Computer Supported Cooperative Work (CSCW '16*), 108–118.

Dittus, M., Quattrone, G., & Capra, L., 2017, February. Mass participation during emergency response: Event-centric crowdsourcing in humanitarian mapping. In *Proceedings of Computer Supported Cooperative Work* (*CSCW '17*), 1290–1303.

Eckle, M., & Albuquerque, J. P. (2015) Quality assessment of remote mapping in openstreetmap for disaster management purposes. *Proceedings of the ISCRAM 2015 Conference - Kristiansand*, May 24–27.

Eco, U. (1995). *How to travel with a Salmon & other essays*. San Diego, California: Houghton Mifflin Harcourt. 95–106. ISBN 978-0-547-54043-6.

Geiger, R. S., & Halfaker, A. (2011) When the Levee breaks: Without Bots, what happens to wikipedia's quality control processes? In *Proceedings of the 9th International Symposium on Open Collaboration* (p. 6: 1–6:6). New York, NY, USA: ACM.

Geiger, R. S., & Halfaker, A. (2013). Using edit sessions to measure participation in wikipedia. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work (CSCW '13*), 861–869.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature*, *438*, 900–901. doi:10.1038/438900a

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of openstreetmap and ordnance survey datasets. *Environ Plann B*, *37*(4), 682–703. doi:10.1068/b35097

Haklay, M., Basiouka, S., Antoniou, V., & Aamar, A. (2010). How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. *The Cartographic Journal*, *47* (4), 315–322. doi:10.1179/000870410X12911304958827

Halavais, A., & Lackaff, D. (2008). An analysis of topical coverage of wikipedia. *Journal of Computer-Mediated Communication*, *13*(2), 429–440. doi:10.1111/jcmc.2008.13.issue-2

Halfaker, A. R., Geiger, S., Morgan, J., & Riedl, J. (2013). The rise and decline of an open collaboration system how wikipedia's reaction to popularity is causing its decline. *American Behavioral Scientist*, *57*(5), 664–688. doi:10.1177/0002764212469365

Holman, R. L. (2008). Comparison of wikipedia and other encyclopedias for accuracy, breadth, and depth in historical articles. *Reference Services Review*, *36*(1), 7–22. doi:10.1108/00907320810851998

Kane, G. C. (2011). A multimethod study of information quality in wiki collaboration. *ACM Transactions Manage Information Systems*, *2*(1) Article 4 (March 2011), 16. doi:10.1145/1929916.1929920

Kittur, A., Bongwon, S., & Chi, E. H. (2008) Can you ever trust a wiki?: Impacting perceived trustworthiness in wikipedia. In *Proceedings of the 2008 ACM conference on Computer supported cooperative work* (CSCW '08). ACM, New York, NY, USA, 477–480.

Kogan, M., Anderson, J., Palen, L., Anderson, K. M., & Soden, R. (2016) Finding the Way to OSM mapping practices: Bounding large crisis datasets for qualitative investigation. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (CHI '16). ACM, New York, NY, USA, 2783–2795.

Korzybski, A. (1958). *Science and Sanity: An introduction to non-aristotelian systems and general semantics*. Englewood, New Jersey: Institute of General Semantics.

Laniado, D., & Tasso, R. (2011) Co-authorship 2.0: Patterns of collaboration in wikipedia. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia* (HT '11). ACM, New York, NY, USA, 201–210.

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: A methodology for information quality assessment. *Information & Management*, *40*(2), 133–146. doi:10.1016/S0378-7206(02)00043-5

Ludwig, T., Kotthaus, C., Reuter, C., Van Dongen, S., & Pipek, V. (2017). Situated crowdsourcing during disasters: Managing the tasks of spontaneous volunteers through public displays. *International Journal on Human-Computer Studies (IJHCS)*, *102*, no. C, 103–121. doi: 10.1016/j.ijhcs.2016.09.008

Ludwig, T., Reuter, C., & Pipek, V. (2015). Social haystack: Dynamic quality assessment of citizen-generated content during emergencies. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *21*(4), Article 17.

Luyt, B., & Tan, D. (2010). Improving wikipedia's credibility: References and citations in a sample of history articles. *Journal of the Association for Information Science and Technology*, *61*, 715–722.

Mashhadi, A., Quattrone, G., & Capra, L. (2013) Putting ubiquitous crowd-sourcing into context. In *Proceedings of the 2013 conference on Computer supported cooperative work* (CSCW '13). ACM, New York, NY, USA, 611–622.

McConchie, A. (2013) Introducing "map gardening". Retrieved September 15, 2016 from http://mappingmashups.net/2013/05/25/introducing-map-gardening/

Mooney, P., & Corcoran, P. (2012) How social is openstreetmap? In *Proceedings of the AGILE 2012 Conference on Geographic Information Science*, 282–287.

Neis, P., & Zipf, A. (2012). Analyzing the contributor activity of a volunteered geographic information project — the case of openstreetmap. *ISPRS International Journal of Geo-Information*, *1*, 146–165. doi:10.3390/ijgi1020146

Palen, L., Soden, R., Anderson, T. J., & Barrenechea, M. (2015) Success & scale in a data-producing organization: The socio-technical evolution of openstreetmap in response to humanitarian events. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (CHI '15). ACM, New York, NY, USA, 4113–4122.

Poiani, T. H., Rocha, R. S., Degrossi, L. C., & Albuquerque, J. P. (2016) Potential of collaborative mapping for disaster relief: A case study of openstreetmap in the Nepal earthquake 2015. In *2016 49th Hawaii International Conference on System Sciences (HICSS)* (pp. 188–197). IEEE.

Reuter, C., Kaufhold, M.-A., & Ludwig, T. (2017). End-user development and social big data – towards tailorable situation assessment with social media. *In* F. Paternò & V. Wulf (Eds.), *New Perspectives in End-User Development* (pp. 307–332). New York, NY: Springer.

Royal, C., & Kapila, D. (2009). What's on wikipedia, and what's not? assessing completeness of information. *Social Science Computer Review*, *27*(1), 138–148. doi:10.1177/0894439308321890

Samoilenko, A., & Yasseri, T. (2014). The distorted mirror of wikipedia: A quantitative analysis of wikipedia coverage of academics. *EPJ Data Science*, *3*(1). doi:10.1140/epjds20

Sehra, S., Singh, J., & Rai, H. (2017). Assessing openstreetmap data using intrinsic quality indicators: An extension to the QGIS processing toolbox. *Future Internet*, *9*(2), 15. doi:10.3390/fi9020015

Soden, R., & Palen, L. (2014) From crowdsourced mapping to community mapping: The post-earthquake work of openstreetmap Haiti. In Proceedings of the 11th International Conference on the Design of Cooperative Systems (COOP '14), 311–326.

Soden, R., & Palen, L. (2016, May). Infrastructure in the wild: What mapping in post-earthquake Nepal reveals about infrastructural emergence. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (pp. 2796–2807). ACM.

Stvilia, B., Gasser, L., Twidale, M. B., & Smith, L. C. (2007). A framework for information quality assessment. *Journal of the Association for Information Science and Technology*, *58*, 1720–1733. doi:10.1002/asi.20652

Stvilia, B., Twidale, M. B., Smith, L. C., & Gasser, L. (2008). Information quality work organization in wikipedia. *Journal of the Association for Information Science and Technology*, *59*, 983–1001. doi:10.1002/asi.20813

Towne, W. B., Kittur, A., Kinnaird, P., & Herbsleb, J. (2013) Your process is showing: Controversy management and perceived quality in wikipedia. In *Proceedings of the 2013 conference on Computer supported cooperative work* (CSCW '13). ACM, New York, NY, USA, 1059–1068.

Vandecasteele, A., & Devillers, R. (2015). Improving volunteered geographic information quality using a tag recommender system: The case of openstreetmap. *In* J. J. Arsanjani, A. Zipf, P. Mooney, & M. Helbich (eds.), *OpenStreetMap in GIScience*. Switzerland: Springer.

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communicable ACM*, *39*(11) (November 1996), 86–95. doi:10.1145/240455.240479

Warncke-Wang, M., Ranjan, V., Terveen, L., & Hecht, B. (2015) Misalignment between supply and demand of quality content in peer production communities. In *Proceedings of ICWSM 2015: Ninth International AAAI Conference on Web and Social Media*.

Yaari, E., Baruchson-Arbib, S., & Bar-Ilan, J. (2011). Information quality assessment of community-generated content – A user study of wikipedia. *Journal of Information Science*, *37*(5), 487–498. doi:10.1177/0165551511416065

Zielstra, D., & Zipf, A. (2010) A comparative study of proprietary geodata and volunteered geographic information for Germany. In *AGILE International Conference on Geographic Information Science*. Guimarães, Portugal.

Zook, M., Graham, M., Shelton, T., & Gorman, S. (2010). Volunteered geographic information and crowdsourcing disaster relief: A case study of the Haitian earthquake. *World Medical & Health Policy*, *2* (2), 7–33. doi:10.2202/1948-4682.1069

About the Authors

Jennings Anderson is a PhD Student in Computer Science at CU Boulder. His work explores the production of Volunteered Geographic Information, especially in times of disaster. He uses open source technologies to build interactive, web-based visualization and analysis systems from massive geospatial datasets.

Robert Soden is a PhD Candidate in Computer Science at CU Boulder. His research examines the various ways in which scientific and engineering understandings of climate change and disaster shape and constrain societal responses to these challenges. Prior to starting his PhD, Robert was a consultant to the World Bank.

Brian Keegan is an assistant professor in the Department of Information Science. He uses computational methods to analyze and theorize about how large-scale social systems respond under stress.

Leysia Palen is Professor of Computer Science, and Professor and Founding Chair of the Department of Information Science at the University of Colorado Boulder. She was awarded the 2015 ACM Computer Human Interaction (CHI) Social Impact Award for contributions the area of crisis informatics.

Ken Anderson is a Professor of Computer Science and the Associate Dean for Education in the College of Engineering and Applied Science at the University of Colorado Boulder. He received his PhD in Computer Science in 1997 at UC Irvine. His research focuses on the design of data-intensive systems.