An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap

A. Yair Grinberger1 | Moritz Schott2 | Martin Raifer3 | Alexander Zipf2,3

1Department of Geography, The Hebrew University of Jerusalem, Jerusalem, Israel
2GIScience Research Group, Institute of Geography, Heidelberg University, Heidelberg, Germany
3Heidelberg Institute for Geoinformation Technology, Heidelberg, Germany

Correspondence
A. Yair Grinberger, Department of Geography, The Hebrew University of Jerusalem, Mt. Scopus, Jerusalem 919051, Israel.
Email: yair.grinberger@mail.huji.ac.il

Funding information
Alexander von Humboldt-Stiftung, Klaus Tschira Stiftung

Abstract
Organized mapping activities within OpenStreetMap frequently lead to the production of massive amounts of data over a short period. In this article we utilize a novel procedure to identify such large-scale data production events in the history of OpenStreetMap and analyze their patterns. We find that events account for a significant share of OpenStreetMap data and that organizational practices have shifted over time towards local knowledge-based events and well-organized data imports. However, regions in the “Global South” remain dependent on remote mapping events, pointing to uneven geographies of representation. We also find that events are frequently followed by periods of increased activity, with the exact nature of effects depending on contextual elements such as previous events. These findings portray organized activities as a significant and unique component which requires consideration when using OpenStreetMap data and analyzing their quality.

1 | INTRODUCTION

Technological developments such as the Web 2.0, sensing technologies, and geographic information systems (GIS) greatly increase the accessibility of geo-information today, thus diminishing the divide between producers/experts and users. Accordingly, more populations engage today with digital mapping activities, leading to...
the formation of bottom-up crowdsourced digital mapping initiatives (Goodchild, 2007). These initiatives provide readily accessible and open data that are increasingly used in multiple fields of practice and research, including humanitarian aid, ecological monitoring, and navigation (Arsanjani, Mooney, Zipf, & Schauss, 2015). In his seminal contribution, Goodchild (2007) named these initiatives volunteered geographic information (VGI) and described them as leading to a "growing reversal of the traditional top-down approach to the creation and dissemination of geo-information" (p. 218). This process, according to Goodchild, relies on the fact that "[a]nyone with an internet connection can select an area on the Earth's surface and provide it with a description" (p. 212).

The quotes above emphasize the role of the individual in producing VGI. This is in line with one of the main justifications for VGI projects—the ability to achieve "collective intelligence", a crowd that is smarter than its smartest member, through the independence of contributors (Spielman, 2014). Yet, the socially constructed nature of VGI implies an interdependence between the data product, the processes that led to its creation, and the context in which these take place (Elwood, 2008; Sieber & Haklay, 2015). In their systems perspective approach for the evaluation of VGI projects, Fast and Rinner (2014) identify the project (i.e., the key conceptual and methodological decisions made by the initiators) as one of the contextual levels of organization shaping the data, alongside with participants and infrastructure. They also acknowledge an increasing institutional involvement in VGI, especially by governmental agencies (supported by the findings detailed in Haklay, Antoniou, Basiouka, Soden, & Mooney, 2014). This exposes organizations as an additional level of organization located between "project" and "participants".

Given that organized activities undermine the principle of independence required for achieving collective intelligence, or at least transform its meaning, studying VGI projects such as OpenStreetMap (OSM) requires also understanding the role and impacts of these activities. The literature offers various explorations of organized activities in OSM (see Section 2) yet, to the best of our knowledge, it still lacks a systematic analysis of the patterns and implications of this level of organization. In fact, two recent analyses—one of the VGI literature (Yan et al., 2020) and one of research related to the OSM project (Sehra, Singh, & Rai, 2017)—do not mention this issue in their lists of trends and themes. This article aims to address this knowledge gap regarding the role of organized activities in OSM by exploring a specific yet informative case: large-scale data production events, that is, events which produce large volumes of data for a specific region over a short period of time. Such events hold the potential to significantly affect the data and, due to the interconnectedness of production procedures and data, also contribution patterns. As such, studying large-scale events holds the potential to contribute knowledge on the roles of organized activity in OSM. In this article we identify such events via a global analysis of the entire history of OSM contributions. Studying the characteristics and spatial and temporal patterns of events, along with the dynamics following them, we provide an extensive account of the role of organized activities in OSM and their impacts.

We continue as follows. Section 2 discusses the OSM project and the current knowledge regarding the organization of mapping activities within it. Section 3 then presents an event identification procedure used to identify a global set of events taking place throughout the history of OSM. Section 4 uses the results of this procedure to analyze the nature of the events, defining a typology of events and exploring their spatial and temporal patterns and effects on contribution activities. Section 5 then turns to discuss the practical and theoretical implications of our results for OSM. Section 6 concludes by offering directions for future research.

2 ORGANIZED ACTIVITIES WITHIN THE OSM PROJECT

This article studies large-scale data production events, as a form of organized activity, within OSM, a prominent VGI project and community dedicated to creating an editable and open map of the surface of the Earth (Mooney & Minghini, 2017). In OSM, spatial entities are represented as nodes (single points, e.g., a tree), ways (collections of nodes forming a line or a polygon, e.g., a road, a building), or relations (collections of nodes, ways, and
other relations, e.g., a bus route consisting of a way representing the route and a collection of bus stop nodes). Additionally, it is possible to add tags to entities, containing semantic information such as land use or name using “key–value” pairs (e.g., landuse = residential). Anyone owning an OSM account can create a new entity, delete an existing one, or edit entities’ geometries and tags (some less frequent operations such as posting notes on the map are possible as well).

The OpenStreetMap foundation maintains the project and while at times intervening in the data through the acts of Working Groups (cf. Bittner, 2017), it does not control the data (OpenStreetMap, 2020e). Guidelines for contributors are available in the project’s Wiki pages (https://wiki.openstreetmap.org) which are also open for editing, thus making it a negotiated product which offers a relatively loose structure for data and mapping activities (Ballatore & Mooney, 2015; Mocnik, Zipf, & Raifer, 2017). Hence, OSM is not characterized by a strong structural mechanism, keeping the project open for different kinds of interactions, some of them of an organizational nature, thus making OSM a “community of communities” (Anderson, Sarkar, & Palen, 2019, p. 2).

As noted above, governmental agencies have become increasingly active within OSM (Haklay et al., 2014). However, governmental agencies are far from the only type of organization involved in OSM today. They are joined by local communities (Hristova, Quattrone, Mashhadi, & Capra, 2013; Mooney, Minghini, & Stanley-Jones, 2015; Perkins & Dodge, 2008), humanitarian organizations (Dittus, Quattrone, & Capra, 2017; Eckle & de Albuquerque, 2015) and even business corporates such as Facebook and Apple (Anderson et al., 2019). In these cases, data are produced in an organized fashion, thus contradicting the original bottom-up conceptualization of OSM as a VGI project.

Institutional engagements with OSM are frequently embodied in (usually) time-constrained organized contribution activities which aim to enrich the database for a specific region. Such events take multiple forms: mapping parties, where mappers meet together to do some on-the-ground mapping (OpenStreetMap, 2020k); mapathons, where mappers meet (mostly physically, but sometimes virtually) to jointly map from afar a certain area (“armchair mapping”; OpenStreetMap, 2020j); and organizations can coordinate global remote mapping events, as usually done by the Humanitarian OpenStreetMap Team (HOT) when responding to disaster and hazard situations (OpenStreetMap, 2020g). These last two remote mapping events rely on utilizing auxiliary geographic information such as satellite images to map from afar. Another type of organized contribution common within OSM is a bulk import of an external data set, usually from an authoritative source (Zielstra, Hochmair, & Neis, 2013). While an import can be carried out by a single user, this event still represents a type of organized contribution, with the organization taking place outside of OSM (the import itself could also entail some coordination efforts, cf. Grinberger, 2018; OpenStreetMap, 2020i).

These three primary types of events—local, remote, and data import (Coetzee, Minghini, Solís, Rautenbach, & Green, 2018)—are grouped together here under the term “large-scale data production events.” Several studies have attempted to study the effects of such events. Imports, for example, were found to introduce quality issues that affect the feasibility of future imports (Zielstra et al., 2013) and to have a limited effect on the behavior of established mappers or on retaining new users (Juhász & Hochmair, 2018). Studies of humanitarian remote mapping efforts show that coordination practices and event frequency are important in engaging newcomers during events and retaining them (Dittus, Quattrone, & Capra, 2016a, 2016b). Yet, apart from high-profile cases, these events generally rely on a relatively stable community of experienced mappers (Dittus et al., 2017). Finally, local mapping parties, despite early evidence (Perkins & Dodge, 2008), are generally found to produce a meaningful amount of data while cultivating a lasting behavioral effect only for light and medium mappers (Hristova et al., 2013; Mooney et al., 2015). Similarly, remote mapping events were found to impact mostly the activity of newcomers (Schott, 2019).

Despite these efforts, knowledge about the nature and effects of large-scale events in OSM remains limited. This is evident in the ongoing (and sometimes heated) discussion within the OSM community and project regarding data imports (cf. https://lists.openstreetmap.org/pipermail/imports/2013-January/001715.html; Juhász & Hochmair, 2018; OpenStreetMap, 2020i; Zielstra et al., 2013). Below we present a global analysis of such events throughout the history of OSM, facilitating the production of new knowledge on their characteristics, patterns, and impacts on contribution patterns.
3 | METHODOLOGY AND DATA

3.1 | Event identification procedure

Identifying a large-scale event is not a straightforward task, given the relative nature of the term “large”—for example, 1,000 contributions made within a span of one month will not form a unique event in a region where the average monthly number of contributions is 800, but will make a significant effect in a region where the monthly average is 100. Hence, in this article we utilize a relative definition of large-scale events which considers the trajectory of data production activities within a given region. Such an approach requires defining a naïve model describing the dynamics of data production in the absence of an event. Gröching, Brunauer, and Rehrl (2014) provide such a model based on empirical patterns identified in OSM data. Their model uses measures of the growth of the database (difference in the number of features between consecutive time periods) and of progress (relative growth) to identify four stages: no data; start (small number of active contributors); growth (low, medium, and high, differentiated by growth values); and a state of saturation. This translates into an S-shaped cumulative development trajectory with a relatively moderate rate of change when no event is introduced (Figure 1). We define large-scale events as sudden increases in the size of the database that greatly exceed growth patterns predicted by the S-shaped curve (Figure 1).

Accordingly, we base our event identification procedure on fitting a logistic curve:

$$\tilde{y}_{r,t} = a + \frac{b}{1 + e^{-c(t-d)}}$$

(1)

to empirical data describing the size of the database over time, and finding significant and positive errors ($e_{r,t} = y_{r,t} - \tilde{y}_{r,t}$). Here $\tilde{y}_{r,t}$ is the predicted cumulative value for region $r$ at time $t$ (measured as the time since the first contribution to the region), and $a$, $b$, $c$, and $d$ are parameters.

The work by Gröching et al. (2014) relies on counting the number of entities in a region. Mapping activities, however, also include other types of contribution actions such as deleting objects and modifying their geometries or tags (Mooney & Minghini, 2017). While it is probable that most events will include the addition of new

**FIGURE 1** A synthetic example of the effects of an event on cumulative values and errors
entities to the database, there is no reason to assume that creating data will be the primary focus of all events. Furthermore, the level of effort varies between contribution types: deleting an object may be done with one simple operation, but creating or editing an entity requires multiple operations such as locating nodes, tagging entities, or modifying individual nodes. To account for this, our procedure counts the number of contribution operations per spatial unit and time period instead of the number of entities or contributions. Deletions are counted as one operation, creations as the number of nodes added plus the number of tags added, and editing operations as the number of new tags and nodes plus the number of those deleted.

Events lead to significant estimation errors. These errors also affect estimation errors for consecutive periods until predicted values "catch up" with observations (Figure 1). To control for such effects, we use time-lagged errors, \( e_{rt}^* = e_{rt} - e_{rt-1} \), which eliminate the temporal trend within errors (Figure 1). We compute standard scores for lagged errors and identify events as time periods in which the error is significantly large and positive at 99% confidence.

### 3.2 Data extraction and preprocessing

The procedure detailed above first requires delineating regions and producing time-series data describing the cumulative number of contribution operations. As theoretically the spatial extent of an event can range from subnational to international, national boundaries do not present a useful geographic system for this analysis. Instead, a data-driven approach based on a quad-tree-like process was used to produce a spatial system. First, the globe was divided into four equal regions. Using the OSM History Database tool (OSHDB; Raifer et al., 2019), which allows querying and aggregating OSM history data in a flexible manner, the number of entities in each quadrant (according to the state of the database in April 2019) was counted. An entity in this case consisted of node and way elements, as defined in the OSM conceptual data model (OpenStreetMap, 2020c; relations were excluded from the analysis). The procedure then recursively divided each cell into equally sized sub-quadrants until none of them included more than 50,000 entities. The resulting division thus contains cells of varying sizes and number of entities, allowing for some cells containing a relatively low number of entities (in cases where the division of one cell produced some sub-quadrants with at least 50,000 entities and others with less). To reduce the probability of incorrect identification of events, cells with less than 20,000 entities were excluded from the analysis, resulting in 10,136 cells. Figure 5 presents the outcome of this process.

Using the OSHDB tool again, we extracted all contributions within each cell along with their time-stamp and computed the number of operations included in them. To create a less noisy and smoother curve, the estimation procedure utilized a temporal resolution of 1 month. While this may have led to a state in which some non-event operations were attributed to an event, it reduced the sensitivity of the procedure to the day-to-day or week-to-week fluctuations that are part of OSM data. The cumulative number of operations was calculated per month and cell, producing the basic information for the curve fitting procedure. The OSHDB query also produced additional information per spatiotemporal unit, including the number of contributions by type (i.e., creation, deletion, tag editing, and geometry editing) and the maximal share of all operations in a given month carried out by one user only. The fact that until the introduction of version 0.5 of the OSM API in October 2007 object changes were not recorded through versioning (OpenStreetMap, 2020b) introduces uncertainties into the data. Hence, only data from November 2007 and onwards were used in the fitting procedure. The fitting itself was carried out using a Levenberg–Marquardt nonlinear least-squares procedure and implemented in R using the nlsLM function from the minpack.lm package (Elzhov, Mullen, Spiess, & Bolker, 2016). Python packages such as pandas, numpy, and scipy were used for subsequent analysis.
4 | RESULTS

4.1 | Event identification results

To assess the results of the curve fitting procedure, we computed normalized root-mean-square error (NRMSE) values (computed as a percentage of the median observed value) for each cell in our spatial system. This analysis shows that the procedure generally produced reasonable results as NRMSE values were relatively low and followed a long-tail distribution (Figure 2) with a median value of 12.45% and interquartile range of 26.15%. In some extreme cases, almost all contribution operations were made during a single month, leading to an almost perfect fitting of the curve. In such cases estimation errors were minimal, hence leading to the identification of small errors as events instead of the true larger contributions. To control for this, we identified false positives through a Box–Cox transformation to the number of operations in events and removing events in which the number of operations was more than two standard deviations below the mean value (i.e., events attributing for less than 15,714 operations). To identify false negatives, the same procedure was performed for the set of observations not identified as events, adding observations in which the number of operations was more than two standard deviations above the mean to the set of events (i.e., months attributing for more than 47,267 operations).

The procedure identified 55,374 events. The spatiotemporal resolution of the analysis led to events spanning more than one cell and/or month being counted more than once, meaning this number is inflated relative to the true number of events. The median number of events per cell was 5.00, the maximum number of events was 46, and 47.89% of the cells recorded three to five different events (Figure 3). When translated to the percentage of months identified as events for each cell, the median value was 3.65% and the maximum value was 33.58%. As might be expected, cells in which more events were identified show a higher total number of operations (Figure 3a). However, there is no clear relation between the frequency of events in a cell and their volumes, expressed in the average number of operations per event (Figure 3b): a higher frequency is associated with smaller events only for cells in which fewer than six events were identified; beyond that the average size of events increases again before stabilizing, suggesting that a high frequency of events in a cell does not make each individual event less meaningful in terms of the volume of contributions.

FIGURE 2  Distribution of NRMSE values
Characterizing events

As discussed in Section 2, there are different types of large-scale data production events (e.g., data imports, local mapping parties, and remote mapping efforts). We used a $k$-means clustering procedure to better differentiate between these types of events, thus producing a typology of events. The procedure used the following variables, extracted as part of the OSHDB query (see Section 3.2):

- **Creations ratio**, the share of creations (contributions adding new entities to the database) out of all contributions during the event;
- **Deletions ratio**, the share of deletions (contributions deleting existing entities);
- **Tag changes ratio**, the share of tag changes (contributions adding a key-value pair to an existing entity or changing an existing tag);
- **Geometry changes ratio**, the share of geometry edits (contributions changing the geometry of an existing entity);
- **Maximal share**, the maximal share of operations made by one user out of all contribution operations (i.e., we computed the number of operations made by each of the users active during the event and divided the maximum of these by the total number of operations).

The first four variables characterize contributions during the event, representing the extent to which they were focused on adding data to the database or on altering existing entries (through deletions, tag changes, or geometry edits). The fifth variable represents the extent to which the event was the result of a centralized effort carried out by one user (or a few users), as in a data import, or a more distributed effort where multiple users were making meaningful contributions, in terms of the volume of operations. We used a combination of clustering metrics and visual analysis to determine the value of the $k$ (number of clusters) parameter. First, we computed solutions for values of $k$ ranging between 2 and 12. Second, we computed three clustering metrics for each solution: the Davies–Bouldin score (DBS), the Calinski Harabasz score (CHS) and the mean silhouette coefficient (MSC). All three agreed that four clusters presented the best solution (DBS = 0.90; CHS = 40,420.04; MSC = 0.41). Yet, when we surveyed the characteristics of clusters in this solution, we discovered that the intra-cluster variance within one of the clusters was large, hence limiting the interpretability of the results. We therefore decided to choose a
value greater than 4 for $k$. The three measures did not agree on which solution was the second best (for DBS it was $k = 8$ with a score of 0.92; for CHS it was $k = 6$ with a score of 39,200.08; and for MSC it was $k = 7$ with a mean coefficient of 0.40). We chose to set $k = 6$ after comparing boxplot figures for all three second-best solutions, as it presented the most coherent set of clusters.

Table 1 presents the characteristics of each cluster using two additional variables: the mean number of active users during the event (total users) and the mean time-step value, the time-step representing number of months since the first contribution recorded in the region. Of the six clusters, four represented highly centralized contributions (clusters 1–4) characteristic of a data import, showing average maximal share values above 84%. The other variables help differentiate between these imports. Cluster 2 is almost entirely focused on creating new entities and is the second earliest type of event on average, pointing to an intervention within a relatively immature database, hence termed “early import.” Clusters 3 and 4 seem to operate within a more mature data environment (average time-step values of 93.18 and 98.29) and hence shift their focus from creation to editing and deletions. Cluster 4 is especially focused on updating geometries and hence is named here “geometry update,” while the other is termed “late import.” The last import cluster (cluster 1) is unique as, beyond being the earliest type of event on average, it focuses almost entirely on tagging activities. As the results in the next sections show, this “tag import” cluster signifies a unique phenomenon within the history of OSM.

The other, more distributed, two clusters (5 and 6) represent community mapping events. The greatest differences between the two exist in the number of active users and editing and creation contributions: cluster 6 involves more users and is characterized by a focus on creating new entities rather than editing existing ones. This pattern fits with the profile of remote mapping events which mobilize larger communities that contribute based on little local knowledge. Accordingly, we identify cluster 6 with remote mapping events and cluster 5 with local knowledge events, that is, events in which the mappers show more knowledge about the area, evident in higher rates of tag and geometry changes. This cluster may include mapping parties along with mapathons and other remote mapping events.

4.3 | Event patterns

Based on this characterization of events, we now turn to examine the role of events in the entire OSM database, their patterns, and their relations with spatial contexts and the state of the database. Table 2 explores the weights of events in the context of the entire history of OSM mapping activities, starting from November 2007. The first insight that emerges from this table is that events are far from a negligible phenomenon within the OSM project: a significant share across all measures is attributed to events. While these numbers are somewhat inflated due to the temporal resolution of the analysis, they still point to the significant role of events in shaping the OSM database, with about 60% of creation contributions being attributed to events and more than 50% of all operations. The earliest types of imports (tag imports, early imports) present the most significant weights, especially in the context of creations and tagging of entities. Yet these results are certainly biased by temporal trends: events introduced later to the project necessarily make a smaller impact due to the increase in the size of the database. Given this, the finding that remote events are responsible for more than 12% of all creations despite becoming a common practice 2.5 years after early imports (on average) suggests that over the more recent time periods this type of event is becoming more significant, as do the other types of later events (late imports, geometry imports, local knowledge events).

Investigating the temporal trends of event weights (Figure 4) confirms this hypothesis. Excluding earlier noisy months, the share of contribution operations attributed to events out of all operations peaks during May 2010. From then on it starts declining until reaching a low of close to 30% in early 2013. The subsequent increase happens despite a decrease in the relative weight of early imports and is driven by an increase in the number of contributions attributed to remote and local knowledge events. This may be at least partially driven by the increasing
| Cluster | n  | Creations ratio | Deletions ratio | Tag changes ratio | Geometry changes ratio | Max. ratio | Users | Time-step | Name                  |
|---------|----|----------------|----------------|-------------------|------------------------|------------|------|-----------|-----------------------|
| 1       | 4,549 | 4.07% | 2.08% | 90.57% | 3.27% | 90.44% | 9.34 | 51.58 | Tag import                |
|         |      | (7.83%) | (6.59%) | (13.86%) | (6.32%) | (14.33%) | (8.9) | (37.14) |                        |
| 2       | 15,974 | 89.94% | 2.66% | 2.38% | 5.03% | 93.67% | 10.24 | 73.84 | Early import              |
|         |      | (8.33%) | (4.9%) | (3.24%) | (5.28%) | (8.59%) | (8.09) | (39.13) |                        |
| 3       | 12,865 | 56.99% | 9.35% | 12.49% | 21.17% | 84.96% | 17.01 | 93.18 | Late import               |
|         |      | (10.69%) | (13.5%) | (10.55%) | (11.1%) | (11.53%) | (10.72) | (33.07) |                        |
| 4       | 6,565  | 21.78% | 8.96% | 18.29% | 50.97% | 86.27% | 17.37 | 98.29 | Geometry import          |
|         |      | (12.81%) | (14.85%) | (13.63%) | (17.33%) | (11.6%) | (11.85) | (33.12) |                        |
| 5       | 7,957  | 36.63% | 8.16% | 18.18% | 37.03% | 43.02% | 31.89 | 101.19 | Local knowledge event    |
|         |      | (14.22%) | (14.03%) | (13.34%) | (15.96%) | (14.31%) | (30.32) | (32.06) |                        |
| 6       | 7,464  | 80.31% | 3.65% | 4.39% | 11.65% | 35.16% | 55.35 | 103.15 | Remote event              |
|         |      | (11.95%) | (5.16%) | (5.17%) | (8.27%) | (17.5%) | (83.93) | (26.81) |                        |
involvement of business corporations within OSM (Anderson et al., 2019). A shift of practices thus emerges which implies that events remain a significant mode of contribution in OSM today.

It is plausible to expect that this shift in practices would follow a geographical pattern: all event types, from remote events, usually associated with humanitarian efforts, through local knowledge events, and down to bulk imports, are inherently tied to specific regions. To explore this, Figure 5 presents the weight of each event type per cell (percentage of all operations in the cell). Several related patterns are indeed evident in this figure. First, the weights of imports are higher in developed regions and certain sections of South America: even early imports, despite being the most evenly distributed type of event, have more weight in developed countries such as the USA, France, Japan, and New Zealand; late imports and especially geometry imports have little weight, if any, in the sub-Saharan region and in far-east Asia. Data in these latter regions are mainly attributed to remote events and, to a lesser extent, to local knowledge events. Therefore, it seems that less developed regions do not experience this shift in practices the same way regions in the Global North do.

The patterns in Figure 5 can also be used to assess the validity of the event identification results, as they point to known events in OSM history. For example, the impacts of tag import events are especially clustered in the
USA. These events mostly take place between September 2009 and January 2010 (Figure 4), thus fitting with the 2009 automated fixup (OpenStreetMap, 2020m) which aimed to remove four superfluous tags defined for each of the ~170 million nodes created during the 2007 import of the 2005 Topologically InteGrated Encoding and Referencing system database (TIGER) produced by the US Census Bureau. Tag imports are therefore a unique phenomenon in OSM history and not a generic type of event. Beyond tag imports, early imports also produced much of the data in the USA due to the 2007 TIGER import (OpenStreetMap, 2020l; Zielstra et al., 2013). The procedure also successfully identified other imports of national or local data sets, such as imports in the Netherlands and France (OpenStreetMap, 2020h). Also, the procedure manages to identify the contributions of humanitarian remote mapping efforts, such as HOT’s mapping efforts after the 2015 Nepal earthquake (OpenStreetMap, 2020a).

4.4 Effects of events

An event is a singular occurrence that greatly affects the data. Due to the inseparability of data and production processes, it is probable that events will also affect contribution behaviors in the months following them. One way to explore this point could be to compute a measure of change, such as the percentage change in the value of a certain measure (e.g., number of operations) in the period after the event in relation to a similar period before it, that is:

FIGURE 5 Events’ weights (percentage of all operations) per cell, by event type: (a) remote events; (b) local knowledge events; (c) early imports; (d) late import; (e) geometry imports; and (f) tag imports. The two higherscale maps for each event type present regions discussed in the text as regions where weights are especially high or low. Cells where the weight is zero are not shown. A dynamic version of this map is available at https://giscience.github.io/OSM_Events
\[ c_{i,t,k} = 100 \left( \frac{k_{Tj,i} - k_{Tj,i}}{\bar{k}_{T,i}} - 1 \right) \]  

where \( c_{i,t,k} \) is the percentage change in the value of measure \( k \) (e.g., change in the number of operations or the number of active users) for an event taking place at time \( t \) in cell \( i \); \( \bar{k}_{Tj,i} \) is the median value of \( k \) in cell \( i \) for a certain time period \( T \), where \( -T \) indicates time before the event (e.g., if \( T = 6 \), \( -T \) refers to \([t - 6, t - 1]\)) and \( T \) time after the event (e.g., \( T = 6 \) refers to \([t + 1, t + 6]\)).

One issue with this measure is that its values may be excessively large for early time-steps: as the volume of activity during these time-steps is low, a small increase in absolute numbers may translate to a large percentage increase. To control for this, we compare events' \( c_{i,t,k} \) values with values for corresponding months in regions in which no event was registered. We call this measure the relative median change \( r_{c_{i,t,k}} \) and its computation includes:

- identifying “control” observations (months in which no event was registered) and computing \( c \) values for them \( (c_{i,t,n}) \);
- computing median \( c_{i,t,n} \) values per time-step for control observations (i.e., \( \bar{c}_{t,n} \)); and
- computing the relative median change as the difference between each event observation \( c_{i,t,k} \) and the \( \bar{c}_{t,n} \) value for the respective time-step:

\[ r_{c_{i,t,k}} = c_{i,t,k} - \bar{c}_{t,n} \]  

To avoid biasing the results, only observations for which no event was registered \( T \) months before and after the observation were included in the analysis. Also, events of the same type taking place on consecutive months were considered as one event. In such cases, \( -T \) ended one month before the first month in which an event was identified and \( T \) started one month after the last month in which an event was identified. Figure 6 uses boxplots to represent the distribution of \( r_{c_{i,t,k}} \) values for number of operations, number of contributions by type (creations, deletions, tag changes, geometry changes), and number of active users for \( T = 6 \) months (black boxes) and \( T = 12 \) months (grey boxes), by event type.

One prominent result is that for almost all combinations of measures and event types, the median value is positive, meaning that in more than 50% of cases the volume of activity increased after the event. Additionally, in most cases the range of positive values is greater than the range of negative values, providing further support to the impression that events of all types frequently act to encourage activity and have no intrinsic negative effect. Over most measures, the widest ranges and hence the highest values for \( T = 6 \) are registered for remote events, early imports, and tag imports. Notice the distinction here between early types of imports (early, tag) and later ones (late, geometry). This suggests that not only the database has matured but also the community, developing more sound import practices that require less correction afterwards (see OpenStreetMap, 2020i). In this sense, imports not following good practices can be seen as catalysts for community engagement (in addition to the natural increase in activity over time attributed to the growth of global and local communities). The increase in activity following remote events can be explained by humanitarian mapping projects tending to be mission-centric campaigns with no clear deadline (Dittus et al., 2017) and that organizations such as HOT practice validation procedures. The 7.3% increase in the median number of active users for \( T = 12 \) in relation to \( T = 6 \) for this type of event further supports this. Beyond this, differences between the \( T = 6 \) and \( T = 12 \) results are visible mostly for tag imports. As this is the earliest type of event (see Table 1), this result may still be related to the volume of activity prior to event. However, given the relative nature of the measure, it is still probable that this import did serve as an incentive for users to correct the data.
Another way in which an event can affect data production dynamics is by affecting the probability of the emergence of future events (as discussed in the context of imports by Zielstra et al., 2013): if an event is deemed successful, the community may be interested in replicating it; if unsuccessful, the community may strive to change practices and organize an event of a different type, countering the effects of the previous event, or decide to avoid organizing any type of event.

To explore this possible effect, the first type of event in each cell was isolated. Then the frequency with which initial event types were followed by events of all types (considering all subsequent events) was computed (Table 3). The results do show a certain form of path dependency: apart from tag imports, all event types were followed by events of the same type in more than 40% of cases. Additionally, imports are more frequently followed by imports of all types; for example, early imports are frequently followed by early imports, tag imports, and late imports and least frequently by local knowledge and remote events. Hence it is possible to differentiate between “import” regions and “community” regions.

**FIGURE 6** Distribution of relative median change values by event type for \( T = 6 \) months (black boxes) and \( T = 12 \) months (gray boxes). Numbers indicate the median value.
Interestingly, remote events are almost as frequently followed by local knowledge events as they are by additional remote events. It is probable that in many cases these subsequent events are part of validation efforts made by organizations such as HOT, as suggested above. The data support this: in 47.59% of the 1,866 cases in which a remote event was the first event registered in a cell, the next event was a remote or a local knowledge event, with a median time difference of 6 months. Thus, it seems that remote events show only limited success in encouraging organized activity beyond the scope of a specific, time-dependent, context. Finally, it is interesting to note that local knowledge events frequently follow the later types of imports (late, geometry) and are also frequently followed by these types of events. This suggests a common causal origin: both organizing the event or identifying available data and importing it while following good practices require a motivated local community. These results, along with the findings from Figure 6, suggest that the effects of events and the processes leading to their creation emerge from intricate relations between local context, the global community, and previous contribution patterns.

| First event type          | # events | Tag import (%) | Early import (%) | Late import (%) | Geometry import (%) | Local knowledge event (%) | Remote event (%) | No event (%) |
|---------------------------|----------|----------------|------------------|----------------|--------------------|------------------------|-----------------|-------------|
| Tag import                | 257      | 31.13          | 60.70            | 40.86          | 48.64              | 24.12                  | 17.51           | 3.11        |
| Early import              | 4,667    | 31.86          | 60.98            | 45.96          | 31.54              | 22.86                  | 27.11           | 6.00        |
| Late import               | 1,899    | 10.80          | 52.34            | 66.61          | 36.44              | 41.02                  | 31.49           | 1.90        |
| Geometry import           | 600      | 11.00          | 37.83            | 66.50          | 49.17              | 48.00                  | 24.00           | 2.17        |
| Local knowledge event     | 839      | 11.80          | 30.63            | 66.15          | 43.38              | 57.33                  | 28.25           | 2.26        |
| Remote event              | 1,866    | 6.38           | 29.69            | 24.65          | 18.01              | 40.73                  | 43.25           | 15.70       |

Note: The three highest values in each row are in bold.

**TABLE 3** The frequency with which each event type, when being the first one recorded for a cell, was followed by each of the types of events

5 | DISCUSSION

In this article we have set out to assess the role of organized activity in OSM by identifying large-scale data production events and characterizing them in terms of their spatial and temporal patterns, contextual relations, and effects. Our findings show that organized activity is indeed a major component of OSM as a VGI project, with large-scale events being a frequent phenomenon encompassing much, and possibly the majority (considering the share of operations and creations attributed to events; Table 2), of data production. As such, the results here can offer some empirical insights relevant to the discussions within OSM mentioned in Section 2. First, it seems that in terms of activity following an event, this phenomenon does not have an intrinsic negative effect. Not only was it rare for activity volumes to decrease after such events, the most contested type of events (early imports) have led to more community engagement by increasing the need for controlling for adverse data effects.

This is most evident perhaps in the case of the 2007 import of TIGER data in the USA, which is notoriously known for its adverse effects on data quality. On the one hand, this import may have fueled the suspicion towards bulk imports. Yet, phrasing in the OSM Wiki entry on TIGER (e.g., “Enough editing has occurred since the original upload”; OpenStreetMap, 2020), the numerous discussions on the American OSM community mailing list (https://lists.openstreetmap.org/pipermail/talk-us), and the still ongoing TIGER fixup efforts suggest that these issues have
served as a call for action for the American OSM community, perhaps even becoming part of its identity. In at least one case this import directly motivated a second organized effort: the 2009 automated TIGER fixup, identified in our results as a cluster of tag imports. Another case where such an early import was followed by a tag import deleting and changing redundant data was registered for the city of Tel Aviv-Jaffa, Israel. In this case, a December 2012 import of address data (labeled here as an early import) was immediately followed by an automated correction in January 2013 (a tag import) in an effort coordinated by the local community (Grinberger, 2018).

This last case relates to another finding: the frequent co-occurrence of late imports, geometry imports, and update events. This finding suggests that imports may be an indication for a functioning and vibrant local community (as in the case of Tel Aviv-Jaffa) which can sustain any possible adverse effects. Nevertheless, local knowledge events tend to present a limited effect on non-organized activity or on the number of active users. This is not to say that field mapping events or local mapathons hold no value. First, it is possible that these are required to keep the community engaged and hence their impact may lie in retaining mappers and avoiding a decrease in activity. Second, these events may work in other ways not captured by our data, such as increasing the cohesiveness of the local community and improving coordination and communication. These possible impacts can be further studied through analyses of changes to collaboration networks within the local community (Mooney & Corcoran, 2014; Truong, de Runz, & Touya, 2019), of the effects of dedicated interventions (Juhász & Hochmair, 2018), or of changes to retention rates and the “life cycles” of contributions (Bégin, Devillers, & Roche, 2017, 2018).

The results provide some indications that remote mapping events encourage further activity, but the findings are somewhat contradictory and need to be interpreted in relation to the contexts producing the events themselves (i.e., organized validation efforts after humanitarian mappings events). For example, in the 12 months following the response to the 2010 Haiti earthquake (OpenStreetMap, 2020f), the median number of monthly operations across the 20 cells covering Haiti was 99.50 (range 11.5–4,061) and the median number of active users per month was 2.5 (range 1–7.5). Yet, this event was a major milestone in the institution of HOT (OpenStreetMap, 2020g), which later led the 2016–2017 mapping efforts following Hurricane Matthew, which included Haiti as well.

The findings here support the critique that production processes cannot be separated from the data in VGI (Sieber & Haklay, 2015), pointing to complex relationships between communities, spaces, previous data contributions, and sociocultural contexts. The maturity of the data encourages a shift in organizational practices as the weight of initial imports decreases while new modes of organization, such as community-based events and imports that enhance data rather than just producing them, are emerging. Accordingly, it is likely that in the future community-based events will become an even more important component of data production procedures. This shift towards community events will not necessarily work to support the main objective of OSM (see https://www.openstreetmap.org/about): representing local knowledge. Remote events, which are more effective than local knowledge events in terms of scale and number of involved mappers (see Tables 1 and 2), rely on external information in the shape of satellite images, and thus integrate local knowledge only to a limited extent. The geographical focus of remote events on less developed spaces in sub-Saharan Africa and south-east Asia points to a possible adverse effect on local representation, despite the general enrichment of representation through such events. From this it follows that not all spaces enjoy the shift in organizational practices alike and that organized activity, despite being more community-driven, cannot solve the digital divide typical of VGI (Sui, Goodchild, & Elwood, 2013) and may even enhance it.

This points to an important theoretical gap that requires attention. Despite contribution behaviors receiving attention as a topic related to understanding the data and their quality (Sehra et al., 2017; Yan et al., 2020), studies of this topic tend to focus on individual mappers, their behaviors, and the interactions between them. The findings here suggest that this may not always be the best resolution for studying OSM data. In their discussion of the relevance of actor-network theory for studying Web 2.0 cartographies, Bittner, Glasze, and Turk (2013, p. 940) mention “asymmetries between actors because bigger actors (e.g., international organizations or companies) succeed in stabilizing many transactions in black boxes, which allow them to appear as a huge and single homogenous actor.” The concept of “black boxes” (actors jointly operating so smoothly as to form a seemingly unified entity) did
not receive enough attention in the study of OSM data quality. An event can be read as an attempt to homogenize contributions; this is clear for imports that originate from a specific organizational environment, but true also for local and remote events that follow common goals, guidelines, and practices. As such, events act to introduce specific dimensions of performance into the project. Studying the behaviors of individuals without considering this may lead to “not seeing the wood for the trees”; the total effect of an assemblage may hide behind the individual actions of its members. Hence, quality measures that treat mapper groups as assemblages acting alongside individual mappers can expose important information regarding asymmetries of knowledge and representation. For example, conceptualizing the 2007 TIGER import as one assemblage leading to the creation of the fixup assemblage (by individual actions and automated fixes) can expose whether these black boxes steered attention away from other mapping issues, thus affecting the quality of the data. Analysis of co-contribution networks identifying dense clusters of cooperation (see Ma, Sandberg, & Jiang, 2015), in which tags or entity types are conceptualized as nodes, holds the potential for identifying such assemblages, alongside spatiotemporal clustering of contribution behaviors. Analyses of this kind should also consider additional contextual elements, such as the aggregated expertise of users (Muttaqien, Ostermann, & Lemmens, 2018) or activities beyond mapping, such as involvement in the development of the project and the discourse around it (Plennert, 2018) or the status of the OSM Wiki and folksonomy (Ballatore & Mooney, 2015; Mocnik et al., 2017).

Such approaches are especially required in the context of the corporate editors phenomenon: individuals hired by corporates to map according to the corporates’ needs. Anderson et al. (2019) list 10 corporates that contribute through automatic identification procedures (e.g., Facebook’s artificial intelligence-assisted road tracing projects; OpenStreetMap, 2020d), in response to users’ feedback, or via remote/on-the-ground mapping. Some of these corporations (e.g., Mapbox, Apple, Grab, Facebook) contribute also to areas which did not experience the shift towards community events, thus opening up the integration of more local knowledge into OSM. Yet, this knowledge is collected and incorporated only if it fits with the corporate’s needs and practices. Hence, they bring into the project their own perspective, at times through massive contributions. Quality measures relating to these groups of mappers as one entity can help generate knowledge regarding the effects of their actions on the data and on contribution patterns, a knowledge which is still lacking given the recency of this phenomenon.

6 | CONCLUSIONS

This article explores the role and effects of large-scale data production events in the history of the OSM project, events in which many data operations are carried out over a short period of time. As these events require some form of organization, we use this to consider the significance of organizational context in OSM. Our results show that organizational contexts play a significant role in the development of the data in OSM, yet that their specific effects change over space and time: as the project and the data mature, partially through events and the responses to them, practices shift towards community-based events; yet this does not automatically imply an increase in the extent of representation of local knowledge, as evident in event dynamics in regions outside of the “Global North.”

The insights emerging from these findings point to theoretical issues that require consideration. First, organizational activities being not just a byproduct of OSM but a crucial and constitutive element of it, as the findings above show, calls for integrating organizational contexts in a formal way into analyses of OSM. This research makes only a first step in this direction. Formal measures based on conceptualizing organized activity as a unique entity and utilizing techniques such as network-based analyses of mapping interactions (Moony & Corcoran, 2013, 2014; Stein, Kremer, & Schlieder, 2015), inquiries into the identity of mappers active in different spaces (Quinn, 2017), and studies of more specific effects on the nature of representation still need to be developed.

Second, the results expose complex interactions between practices, data, context, physical spaces, and digital environments during data production. These dynamics picture OSM as an ontogenetic digital space (Kitchin & Dodge, 2011), a space that is always in a state of becoming through the practices of participating mappers. New
contributions, whether creations, edits, or deletions, continuously alter these spaces while relying on their existing spatial forms and imperfectly citing previous contribution practices. Such spaces extend themselves through interactions beyond any specific scale and are constantly engaged with other spaces, such as physical and cultural ones. OSM research has yet to consider the implications of this state for analysis (perhaps with the exception of Schott, 2019). Conceptualizing OSM in this way calls for relational modes of inquiry that focus on the relations that produce space. The network-based studies mentioned above certainly mark a step in the right direction. Yet more work is required in order to integrate dynamics, meaning, and practices into such methods. Future endeavors should explore issues such as how networks and community structures change following major contributions (see Grinberger, 2018), how meaning is being introduced into OSM projects through activities such as tagging and adding Wiki entries (Ballatore & Mooney, 2015; Mooney & Corcoran, 2012), what trajectories mappers follow through OSM, and what practices they employ. While the VGI literature relates to some of these issues, a holistic inquiry into the subject, one that will probably extend beyond quantitative and computational approaches, can serve to further enhance our understanding of OSM and possibly also other crowd-based mapping endeavors.

ACKNOWLEDGEMENTS
The authors would like to thank Mr. Rafael Troilo for his support in developing and running the OSHDB query. We would like to also thank the three anonymous reviewers for their helpful comments on earlier drafts of this manuscript. Yair Grinberger’s work was financially supported by the Alexander von Humboldt Foundation. Martin Raifer’s work was financially supported by the Klaus Tschira Stiftung. Early versions of this manuscript were presented at the Geographical and Cultural Aspects of Geo-Information (GeoCultGIS) 2019 AGILE workshop and the 2019 State of the Map conference.

CONFLICT OF INTEREST
A. Yair Grinberger, Mortiz Schott, and Martin Raifer are members of the OpenStreetMap community, actively engaging in community-related activities. Additionally, Martin Raifer is a member of the OpenStreetMap Foundation.

ORCID
A. Yair Grinberger https://orcid.org/0000-0002-7393-9903
Moritz Schott https://orcid.org/0000-0003-3018-2565
Martin Raifer https://orcid.org/0000-0002-2106-8560
Alexander Zipf https://orcid.org/0000-0003-4916-9838

ENDNOTES
1 While it is possible that such events will emerge through a spontaneous reaction to external events, the volume of contributions suggests some level of organization.
2 Using the OSM History Database tool, there is no way to distinguish a change to a key’s value from a case where a key was deleted and then added again. Thus, we analyzed key–value pairs as one object, meaning that changes in key’s values were counted as two operations—deletion and addition. We believe that due to our focus on large-scale production events this double-counting of tag operations did not significantly affect our results.
3 The code for this procedure and all following procedures is available at https://github.com/GIScience/OSM_Events.
4 Plotting the sum of squared distances (SSD) of samples to their closet cluster center by value of $k$ provided further support to this decision as the “elbow” (the section of the line just before it flattens) appeared after $k = 4$, at $k = 6$ or $k = 7$.

REFERENCES
Anderson, J., Sarkar, D., & Palen, L. (2019). Corporate editors in the evolving landscape of OpenStreetMap. ISPRS International Journal of Geo-Information, 8(5), 232. https://doi.org/10.3390/ijgi8050232
Arsanjani, J. J., Mooney, P., Zipf, A., & Schauss, A. (2015). Quality assessment of the contributed land use information from OpenStreetMap versus authoritative datasets. In J. J. Arsanjani, A. Zipf, P. Mooney, & M. Helbich (Eds.), OpenStreetMap in GIScience (pp. 37–58). Cham, Switzerland: Springer.

Ballatore, A., & Mooney, P. (2015). Conceptualizing the geographic world: The dimensions of negotiation in crowdsourced cartography. International Journal of Geographical Information Science, 29(12), 2310–2327. https://doi.org/10.1080/13658816.2015.1076825

Bégin, D., Devillers, D., & Roche, S. (2017). Contributors’ withdrawal from online collaborative communities: The case of OpenStreetMap. ISPRS International Journal of Geoinformation, 6(11), 340. https://doi.org/10.3390/ijgi6110340

Bégin, D., Devillers, D., & Roche, S. (2018). The life cycle of contributors in collaborative online communities: The case of OpenStreetMap. International Journal of Geographical Information Science, 32(8), 1611–1630. https://doi.org/10.1080/13658816.2018.1458312

Bittner, C. (2017). OpenStreetMap in Israel and Palestine: “Game changer” or reproducer of contested cartographies? Political Geography, 57, 34–48. https://doi.org/10.1016/j.polgeo.2016.11.010

Bittner, C., Glasze, G., & Turk, C. (2013). Tracing contingencies: Analyzing the political assemblages of web 2.0 cartographies. GeoJournal, 78, 935–948. https://doi.org/10.1007/s10708-013-9488-8

Coetzee, S., Minghini, M., Solis, P., Rautenbach, V., Green, C. (2018). Towards understanding the impact of mapathons - Reflecting on younmappers experiences. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII-4/W8, 35–42. http://dx.doi.org/10.5194/isprs-archives-xlII-4-w8-35-2018

Dittus, M., Quattrone, G., & Capra, L. (2016a). Analysing volunteer engagement in humanitarian mapping: Building contributor communities at large scale. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing, San Francisco, CA (pp. 108–118). New York, NY: ACM.

Dittus, M., Quattrone, G., & Capra, L. (2016b). Social contribution settings and newcomer retention in humanitarian crowd mapping. In E. Spiro & Y.-Y. Ahn (Eds.), Social informatics (pp. 179–193). Cham, Switzerland: Springer.

Dittus, M., Quattrone, G., & Capra, L. (2017). Mass participation during emergency response: Event-centric crowdsourcing in humanitarian mapping. In Proceedings of the 20th ACM Conference on Computer-Supported Cooperative Work and Social Computing, Portland, OR (pp. 1290–1303). New York, NY: ACM.

Eckle, M., & de Albuquerque, J. P. (2015). Quality assessment of remote mapping in OpenStreetMap for disaster management perspectives. In Proceedings of the ISCRAM 2015 Conference, Kristiansand, Norway.

Elwood, S. (2008). Volunteered geographic information: Future research trends directions motivated by critical, participatory, and feminist GIS. GeoJournal, 72, 173–183. https://doi.org/10.1007/s10708-008-9186-0

Elzohov, T. V., Mullens, K. M., Spiess, A.-N., & Bolker, B. (2016). Minimap.pack: R interface to the Levenberg-Marquardt nonlinear least-squares. Retrieved from https://CRAN.r18project.org/package=minimap.pack

Fast, V., & Rinner, C. (2014). A systems perspective on volunteered geographic information. ISPRS International Journal of Geo-Information, 3(4), 1278–1292. https://doi.org/10.3390/ijgi3041278

Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. GeoJournal, 69(4), 211–221. https://doi.org/10.1007/s10708-007-9111-y

Grinberger, A. Y. (2018). Identifying the effects of mobility domains on VGI: Towards an analytical approach. In Proceedings of the VGI-E-LIVE Workshop held in conjunction with the 21st AGILE Conference on Geographic Information Science, Lund, Sweden.

Gröchting, S., Brunauer, R., & Rehrl, K. (2014). Digging into the history of VGI data-sets: Results from a worldwide study on OpenStreetMap mapping activity. Journal of Location Based Services, 8(3), 198–210. https://doi.org/10.1080/17487252.2014.978403

Haklay, M., Antoniou, V., Basiouka, S., Soden, R., & Mooney, P. (2014). Crowdsourced geographic information use in government. London, UK: World Bank Global Facility for Disaster Reduction and Recovery.

Hristova, D., Quattrone, G., Mathadi, A., & Capra, L. (2013). The life of the party: Impact of social mapping in OpenStreetMap. In Proceedings of the AAAI International Conference on Weblogs and Social Media, Boston, MA (pp. 234–243). Menlo Park, CA: AAAI.

Juhász, L., & Hochmair, H. H. (2018). OSM data import as an outreach tool to trigger community growth? A case study in Miami. ISPRS International Journal of Geo-Information, 7(3), 113. https://doi.org/10.3390/ijgi7030113

Kitchin, R., & Dodge, M. (2011). The transduction of space. In R. Kitchin & M. Dodge (Eds.), Code/spaces: Software and everyday life (pp. 65–80). Cambridge, MA: MIT Press.

Ma, D., Sandberg, M., & Jiang, B. (2015). Characterizing the heterogeneity of the OpenStreetMap data and community. ISPRS International Journal of Geo-Information, 4(2), 535–550. https://doi.org/10.3390/ijgi4020535

Mocnik, F.-B., Zipf, A., & Raifer, M. (2017). The OpenStreetMap folksonomy and its evolution. Geo-spatial Information Science, 20(3), 199–209. https://doi.org/10.1080/10095020.2017.1368193

Mooney, P., & Corcoran, P. (2012). The annotation process in OpenStreetMap. Transactions in GIS, 16(4), 561–579. https://doi.org/10.1111/j.1467-9671.2012.01306.x
Mooney, P., & Corcoran, P. (2013). Understanding the role of communities in volunteered geographic information projects. In J. M. Krisp (Ed.), Progress in location-based services (pp. 357–371). Berlin, Germany: Springer.

Mooney, P., & Corcoran, P. (2014). Analysis of interaction and co-editing patterns amongst OpenStreetMap contributors. Transactions in GIS, 18(5), 633–659. https://doi.org/10.1111/tgis.12051

Mooney, P., & Minghini, M. (2017). A review of OpenStreetMap data. In G. Foody, L. See, S. Fritz, P. Mooney, A.-M. Olteanu-Raimond, C. C. Fonte, & V. Antoniou (Eds.), Mapping and the citizen sensor (pp. 37–59). London, UK: Ubiquity Press.

Mooney, P., Minghini, M., & Stanley-Jones, F. (2015). Observations on an OpenStreetMap mapping party organised as a social event during an open source GIS conference. International Journal of Spatial Data Infrastructures Research, 10, 138–150. https://doi.org/10.2902/1725-0463.2015.10.art7.

Muttaqien, B. I., Ostermann, F. O., & Lemmens, R. L. G. (2018). Modeling aggregated expertise of user contributions to assess the credibility of OpenStreetMap features. Transactions in GIS, 22(3), 823–841. https://doi.org/10.1111/tgis.12454

OpenStreetMap. (2020a). 2015 Nepal earthquake. Retrieved from https://wiki.openstreetmap.org/wiki/2015_Nepal_earthquake

OpenStreetMap. (2020b). API changes between v0.4 and v0.5. Retrieved from https://wiki.openstreetmap.org/wiki/API_changes_between_v0.4_and_v0.5

OpenStreetMap. (2020c). Elements. Retrieved from https://wiki.openstreetmap.org/wiki/Elements

OpenStreetMap. (2020d). Facebook AI-assisted road tracing. Retrieved from https://wiki.openstreetmap.org/wiki/Facebook_AI-Assisted_Road_Tracing

OpenStreetMap. (2020e). FAQ. Retrieved from https://wiki.osmfoundation.org/wiki/FAQ

OpenStreetMap. (2020f). Haiti OpenStreetMap response. Retrieved from https://blog.openstreetmap.org/2010/01/14/haiti-openstreetmap-response/

OpenStreetMap. (2020g). Humanitarian OSM team. Retrieved from https://wiki.openstreetmap.org/wiki/Humanitarian_OSM_Team

OpenStreetMap. (2020h). Import/catalogue. Retrieved from https://wiki.openstreetmap.org/wiki/Import/Catalogue

OpenStreetMap. (2020i). Import/guidelines. Retrieved from https://wiki.openstreetmap.org/wiki/Import/Guidelines

OpenStreetMap. (2020j). Mapathon. Retrieved from https://wiki.openstreetmap.org/wiki/Mapathon

OpenStreetMap. (2020k). Mapping parties. Retrieved from https://wiki.openstreetmap.org/wiki/Mapping_parties

OpenStreetMap. (2020l). TIGER. Retrieved from https://wiki.openstreetmap.org/wiki/TIGER

OpenStreetMap. (2020m). TIGER fixup/node tags. Retrieved from https://wiki.openstreetmap.org/wiki/TIGER_fixup/node_tags

Perkins, C., & Dodge, M. (2008). The potential of user-generated cartography: A case study of the OpenStreetMap project and Mapchester mapping party. North West Geography, 8(1), 19–32.

Plennert, M. (2018). The social construction of technological stasis: The stagnating data structure in OpenStreetMap. Big Data & Society, 5(2), 1–18. https://doi.org/10.1177/2053951718790591

Quinn, S. (2017). Using small cities to understand the crowd behind OpenStreetMap. GeoJournal, 82(3), 455–473. https://doi.org/10.1007/s10708-015-9695-6

Raifer, M., Troilo, R., Kowatsch, F., Auer, M., Loos, L., Marx, S., ... Zipf, A. (2019). OSHDB: A framework for spatio-temporal analysis of OpenStreetMap history data. Open Geospatial Data, Software & Standards, 4(3), 1–12. https://doi.org/10.1186/s40965-019-0061-3

Schott, M. (2019). Under the spell of community happenings: Analysing the effects of mapping events on OpenStreetMap contributors (Unpublished MA thesis). Heidelberg, Germany: Heidelberg University.

Sehra, S., Singh, J., & Rai, H. (2017). Using latent semantic analysis to identify research trends in OpenStreetMap. ISPRS International Journal of Geo-Information, 6(7), 195. https://doi.org/10.3390/ijgi6070195

Sieber, R., & Haklay, M. (2015). The epistemology(s) of volunteered geographic information: A critique. Geo: Geography & Environment, 2(2), 122–136. https://doi.org/10.1002/ge2.10

Spielman, S. E. (2014). Spatial collective intelligence: Credibility, accuracy, and volunteered geographic information. Cartography & Geographic Information Science, 41(2), 115–124. https://doi.org/10.1080/15230406.2013.874200

Stein, K., Kremer, D., & Schlieder, C. (2015). Spatial collaboration networks of OpenStreetMap. In J. J. Arsanjani, A. Zipf, P. Mooney, & M. Helbich (Eds.), OpenStreetMap in GIScience (pp. 167–186). Cham, Switzerland: Springer.

Sui, D., Goodchild, M. F., & Elwood, S. (2013). Volunteered geographic information, the exaflood, and the growing digital divide. In D. Sui, S. Elwood, & M. F. Goodchild (Eds.), Crowdsourcing geographic knowledge (pp. 1–12). New York, NY: Springer.

Truong, Q. T., de Runz, C., & Touya, G. (2019). Analysis of collaboration networks in OpenStreetMap through weighted social multigraph mining. International Journal of Geographical Information Science, 33(8), 1651–1682. https://doi.org/10.1080/13658816.2018.1556395
Yan, Y., Feng, C.-C., Huang, W., Fan, H., Wang, Y.-C., & Zipf, A. (2020). Volunteered geographic information research in the first decade: A narrative review of selected journal articles in GIScience. *International Journal of Geographical Information Science, 34*(9), 1765–1791. https://doi.org/10.1080/13658816.2020.1730848

Zielstra, D., Hochmair, H. H., & Neis, P. (2013). Assessing the effects of data imports on the completeness of OpenStreetMap: A United States case study. *Transactions in GIS, 17*(3), 315–334. https://doi.org/10.1111/tgis.12037

**How to cite this article:** Grinberger AY, Schott M, Raifer M, Zipf A. An analysis of the spatial and temporal distribution of large-scale data production events in OpenStreetMap. *Transactions in GIS.* 2021;25:622–641. https://doi.org/10.1111/tgis.12746