Chapter 5
Healthsites.io: The Global Healthsites Mapping Project

**René Saameli, Dikolela Kalubi, Mark Herringer, Tim Sutton
and Eric de Roodenbeke**

5.1 Introduction

The world is facing an increasing number of complex natural or man-made humanitarian crises. In order to respond to these growing challenges, humanitarian actors are deploying more and more innovative technologies and approaches to support relief aid (Haselkorn and Walton 2009). Geo-information are an example where humanitarian practices have dramatically evolved in the recent years with the emergence of the phenomena called by Goodchild (2007) Volunteered Geographic Information (VGI), or more generally by Burns (2014) "digital humanitarianism". Indeed, the democratization of access to Global Positioning System (GPS), satellite imagery and web mapping platforms, such as OpenStreetMap (OSM), and more recently mobile phone data collection tools have enabled large numbers of remote and on the ground individuals to produce geographical information to support humanitarian action. This technological evolution has offered greater opportunities of use for humanitarian actors. It has enabled the collection and sharing of large amount of data in short period of time at a fraction of the costs of the traditional data collection and

R. Saameli (✉) · D. Kalubi
International Committee of the Red-Cross, Lausanne, Switzerland
e-mail: rsaameli@icrc.org

D. Kalubi
e-mail: dikolela.kalubi@epfl.ch; dkalubi@icrc.org

M. Herringer · T. Sutton
Healthsites.io, London, UK
e-mail: mark@healthsites.io

T. Sutton
e-mail: tim@kartoza.com

E. de Roodenbeke
International Hospital Federation, Bernex, Switzerland
e-mail: ederoodenbeke@ihf-fih.org

map-making methods (Haworth and Bruce 2015). However, the full scope of opportunities offered by VGI is still underused by traditional humanitarian actors. In order to understand this situation, Richards and Veenendaal (2014) have analyzed comprehensively the gap between the United Nations World Food Programme Crisis Mapping Operations and Crowdsourcing Technology. They conclude that crowdsourcing captured a large amount of data, but not sufficiently the required ones for the agency operations and with not the needed quality. Haworth and Bruce (2015) also highlight this need of enhancing data quality assurance to enhance the relevance of VGI for disaster management. Data quality in VGI relies heavily on Linus's law, which implies the more observers, the more likely an error will be identified. As shown by Haklay et al. (2010) in their analysis of Open Street Map data quality, this law seems to work well for spatial accuracy. However, Mooney and Corcoran (2012) have also discovered serious quality issues with tags or annotate objects in OSM. Such attribute data is usually much more needed than accurate geographical coordinates to support humanitarian management.

Geo-information on health facilities in disaster areas are a good example of the challenges of the use of VGI for humanitarian action. The most comprehensive healthsite geodatabase based on VGI is probably OSM, but information on services offered are still largely incomplete and questionable in terms of reliability. Other health geodatabases with comprehensive set of helpful attributes for health workers exist, but these databases are not easily shared outside of the health organizations which have gathered them, or are only regional in their coverage. OSM and these restricted datasets complement each other in terms of geographical coverage and in terms of the information they contain, however they are almost never readily available in a consolidated, freely and accessible way. Data exchange between VGI communities and health organizations is usually unidirectional and punctual. Traditional humanitarian agencies tend to task digital communities only for specific tasks lasting a rather short period of time (Burns 2014).

In order to address this issue, the Global Healthsites Mapping Project has been launched in 2015 to create an online interactive map, Healthsites.io, of every health facility in the world and make the details of each location and services easily accessible. A team of freelance developers, researchers, the International Committee of the Red-Cross (ICRC), and the International Hospital Federation (IHF) have joined their competences and networks in order to provide a single point of reference for healthcare workers, aid agencies, contingency planners, government agencies, and citizens who need access to a highly curated global dataset of healthcare facilities. In order to meet this aim, the project team has to address three major challenges:

1. Integrate multiple unstructured datasets in one unique database
2. Enhancing the reliability of data
3. Foster sharing and updating of information

In Sect. 5.2, this paper will present the approaches that are currently developed to address these three challenges. In Sect. 5.3, the paper will analyze the potential impacts, risks, and the perspectives of this project.

5.2 Healthsite.io Approach

5.2.1 Datasets Integration

Due to its open data license, its large number of entities, its worldwide coverage, and its large community, OSM was chosen as the main dataset of the project. In addition, several databases from trusted partners were chosen. All this data gathered represented more than 150,000 health localities worldwide. In order to integrate all these datasets with different structures and values, a data model, based on the Entity–Attribute–Value model, and inspired by the OSM data model was chosen. The OSM data model enables users to store information about anything. A FullTextSearch index has been implemented to enable searching for textual data. However, this flexibility of storing information represented a risk potentially hampering the goal of having a curated database with easily accessible information. In order to address this issue, 15 core attributes (Table 5.1) that are relevant for both the public and health professionals were defined by the International Hospital Federation. For most of these core attributes, defined values were set to enable comparisons. Indeed, an essential question such as "What type of health facilities is it" can vary greatly according to the national or organizational classification system.

Each time a record is created in Healthsites.io, a globally unique identifier is created and assigned to that record. This is specific to the Healthsites.io database and is used to provide a canonical point of reference for that record.

The upstream id is used to create a back reference to the original source data (e.g., from a national healthsites dataset).

5.2.2 Validation Process

In order to enhance the reliability of health facilities data, several validation processes have been planned to be added to the project. There is an automatic verification ranging from a simple: "email address should look like an email address" to more complex which even rely on external services like: "check if the Locality address is similar to results returned by external geocoding services" or "check if the telephone number is correct by manually calling the number and verifying".

There will also be a validation process based on user reputation (Fig. 5.1) to assess the reliability of data through a Locality Validity Index (LVI). This reputation-based process consists of four steps. The first step takes place during data integration. Depending on source of the data, the LVI obtains a score ranging from 0 to 10; 0 being for data added by a new user, 10 for data from a community trusted user, and 5 if it is a batch of data coming from a Ministry of Health or healthcare organizations. The reliability of the user is based on the monitoring of his activity on the platform and the crosscheck of his activity by other users. Second, once the data is displayed on the map of Healthsite.io, any user can complete missing attributes data and modify

Table 5.1 HealthSite.io database core attributes

Healthsite attribute name	Value
uuid	Universally unique identifier
upstream_id	Unique identifier placed on the source location
date_modified	Date time stamp of data modification
name	Name of the facility local name and in English
geom	Lat/long
source	Name and or website of the organization supplying the data
physical_address	Address
contact_number	Number including international dialing code (+250 252 588888)
nature_of_facility	Clinic without beds, clinic with beds, first referral hospital, second referral hospital or General hospital, tertiary level including University hospital
scope-of-service	All type of services, specialized care, general acute care, rehabilitation care, old age/hospice care
operation	24/24 & 7/7; open only during business hours; other (specify)
inpatient-service	Number of full-time beds, number of part-time beds
ancillary-services	Operating theater, laboratory, imaging equipment, intensive care unit, emergency department
activities	Medicine and medical specialties, surgery and surgical specialties, maternal and women health, pediatric care, mental health, geriatric care, social care
staff	X full-time equivalent doctors and Y full-time equivalent nurse, level of competency of the health workers
ownership	Public, private not for profit, private commercial
raw_data_archive_URL	A link to the raw data file (.csv, etc.)

or validate existing attributes data through a tweet channel. Third, the more users confirm the validity of information, the higher the LVI becomes. Within this step, if an authoritative user such as a staff of health organization, verifies information, the LVI becomes even higher. Finally, if time goes without anyone validating the record, the LVI progressively decreases.

5.2.3 Updating

Over the last years, World Health Organization (WHO) have developed and supported several projects, systems and guidelines for national and regional authorities to map their health facilities. However, in many countries where humanitarian workers intervene, this data once collected is often not updated or remains not easily accessible. In
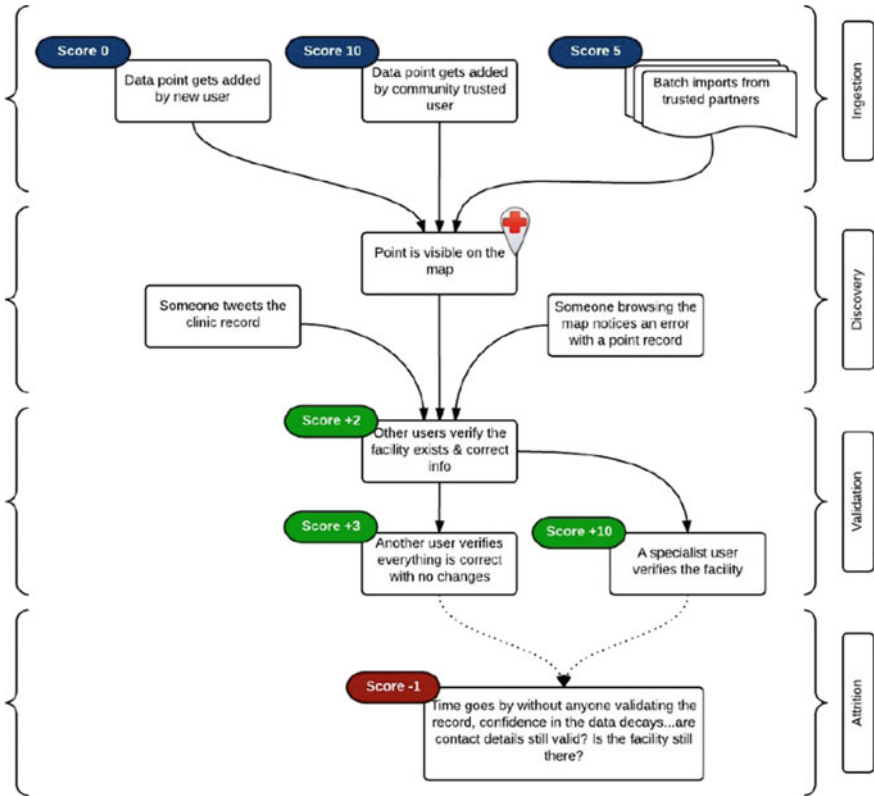
Fig. 5.1 Reputation based validation process

the absence of research on the cause this situation, one can only assume that authorities at regional and national level lack of resources to carry out the monitoring of healthcare services. In order to overcome this issue, the Healthsite.io project has been designed to have a bottom-up approach of the monitoring and data collection. Indeed, it is OSM community who provides the localization of health facilities and health workers in the field who provide related attribute information, such as the type of services available. These two sources of data are consolidated in Healthsite.io and data quality verified through the LVI. Once data quality is good enough, data are then sent back to OSM in order to be shared widely through the community of OSM users.

5.2.4   Opportunities, Risks, and Perspectives

This project aims to democratize access to health geodata. By enhancing the quality and accessibility of geodata, this project offers several opportunities to achieve social, health and humanitarian impacts in areas where data is scarce. A better knowledge of locations of health facilities and associated services can help to map and identify population underserved by healthcare services (Munoz and Källestål 2012) (Blanford et al 2012). Such information can contribute to develop and advocate better evidence-based health policy. It can also help healthcare organizations to better plan their activities, vaccination campaigns based on the correlation between vaccination rate and distance to primary healthcare center (Al Taiar et al. 2010). For relief workers, knowing where the health facilities and associated health service are located are crucial information to know to prepare contingency plans, but also to respond promptly and meaningfully during the emergency.

The project also bears some limitations and risks. Several studies (Haworth and Bruce 2015) have showed that questions about liability of VGI are often not clearly answered "Who is responsible if harm results from reliance on volunteered information: the initial contributor, the host or the organization responsible for the website or product relying on VGI". The question of moral and legal reliability is highly relevant in the case of public exposure and dissemination of potentially sensitive data such health facilities and workers location. In order to address this risk, a governance committee composed of ICRC, IHF and developer team has been set to monitor data quality and the risk of exposure of data in sensitive contexts.

The capacity of the project to meet its goal to build global, complete and high-quality database of world health facilities depends on the contributions of VGI communities and health experts. In comparison with other similar VGI initiative, this project has the advantage of having the endorsement of major actors such as Humanitarian OSM team, the ICRC, and IHF. This support enables a strong promotion among digital humanitarians and health experts.

Finally, the model of Healthsite.io of taking OSM data to enhance its attribute by health experts before exporting back completed and validated data to OSM is a truly innovative approach. It fosters bidirectional exchange between health experts and VGI communities. In addition, it also a new approach to quality issue of attribute data in OSM. If this model proves to be successful, it could be replicated in other essential domain, such water services, where accessibility to detailed and reliable data is much needed for humanitarian operations. This approach would be a new way to maintain updated comprehensive thematic database without having to spend resources usually required for such tasks

Research on how to measure quality of VGI, specifically attribute data, are recent and few. Healthsite.io offers an interesting concrete case to measure the reliability of data jointly validated by VGI community and authoritative experts. Such research could help to identify opportunities to enhance VGI validation process. In Healthsite.io, this process is mainly based on user reputation where also only few research on VGI applications have been done.

Furthermore, research on motivation of VGI and health experts contributing to such project could help to develop strategies to ensure long-term updating of the database.

References

Al-Taiar, A., Clark, A., Longenecker, J., & Whitty, C. (2010). Physical accessibility and utilization of health services in Yemen. *International Journal of Health Geographics, 9*, 38.

Blanford, J., Kumar, S., Luo, W., & MacEachren, M. (2012). It's a long, long walk: Accessibility to hospitals, maternity and integrated health centers in Niger. *International Journal of Health Geographics, 11*, 24.

Burns, R. (2014). Rethinking big data in digital humanitarianism: Practices, epistemologies, and social relations. *GeoJournal, 80*(4), 477–490.

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

Haklay, M., Basiouka, S., Antoniou, V., et al. (2010). How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. *Cartographic Journal, 47*(4), 315–322.

Haselkorn, M., & Walton, R. (2009). The role of information and communication in the context of humanitarian service. *IEEE Transactions on Professional Communication, 52*(4), 325–328.

Haworth, B., & Bruce, E. (2015). A review of volunteered geographic information for disaster management. *Geography Compass, 9,* 237–250.

Mooney, P., & Corcoran, P. (2012). The annotation process in OpenStreetMap. *Transactions in GIS, 16*(4), 561–579.

Muenoz, U., & Källestål, C. (2012). Geographical accessibility and spatial coverage modeling of the primary health care network in the Western Province of Rwanda. *International Journal of Health Geographics, 11*, 40.

Richards, S., & Veenendaal, B. (2014). *Understanding the gap between the United Nations World Food Programme crisis mapping operations and crowdsourcing technology*. Conference Article. Curtin University. http://espace.library.curtin.edu.au:80/R?func=dbin-jump-full&local_base=gen01-era02&object_id=235814.