Chapter 5.2
Crowdsourcing to gather data

Kerri Wazny

Learning objectives

To understand the fundamentals of crowdsourcing and its relevance to health emergency and disaster risk management (Health EDRM), including :

- What crowdsourcing is.
- How crowdsourcing differs from related terms.
- Strengths and limitations of crowdsourcing.
- Things to consider when designing a study that would use crowdsourcing to gather data.

What is crowdsourcing?

- Crowdsourcing is a method to harness the knowledge, creativity or sheer manpower of a large number of people at the same time.
- It has increased greatly in the past decade.
- Many think the method is underutilized and underexploited.
- Technological advances have facilitated its impact and feasibility, including great accessibility of mobile phones and advances in machine learning and artificial intelligence (AI) have facilitated ways of processing large amounts of data obtained through crowdsourcing.

Strengths and limitations of crowdsourcing

Strengths

- Provides answers to questions that might not be answered in other ways.
- Lower operational and data collection costs.
- Greater sample size.
- Enables researchers to receive data in real time.

Limitations

- Crowd is often self-selected and might not be generalizable, especially if there are few contributors.
- Security and protection issues need to be considered.
- False information might be submitted.

Models of crowdsourcing

Crowd processing: use of large number of people to process information independently, which become partially aggregated for quality assurance.

- Examples: ReCAPTCHA, GalaxyZoo, BioGames.

Crowd rating: use of large number of people to vote or provide their opinion.

- Examples: TripAdvisor, Hollywood Stock Exchange.

Crowd solving: use of large number of people to solve a problem, where the best submission is the 'winner'.

- Examples: FoldIt, Crowdmed, Innocentive.

Crowd creation: use of large number of people to co-create.

- Example: Threadless.

Methods of crowdsourcing

- Needs to involve a clear call for submissions or tasks, which might be:
  - Voluntary or remunerated.
  - Conducted using technology to enable low-cost and speedier data transmission.
- Crowd can be formed of laypersons or experts.
- Responses may be aggregated or compared against each other in competition form.

Crowdsourcing terms

Term	Definition
Participatory epidemiology	Using participatory methods in epidemiology, which could range from designing the study to participatory methods in data collection (the latter would likely be in line with crowdsourcing) *(31)*.
Wisdom of the crowd (that is, collective intelligence)	A phrase coined by Surowiecki *(32)*, describing a form of crowdsourcing that relies on having an intelligent crowd and follows four 'rules' to ensure crowd intelligence: diversity, aggregation, decentralization and independence. Not all crowdsourcing requires a wise crowd, but all 'wisdom of the crowd' activities are crowdsourcing.
Citizen science	Non-professionals conducting science-related activities *(33)*. While crowdsourcing refers to *how* the activity is conducted, citizen science refers to *who* is doing it, and *what* they are doing. Often, crowdsourcing and citizen science are performed in tandem.
Health 2.0	The use of Web 2.0 technologies to actively participate in one's health *(33)*. These could facilitate crowdsourcing (for example, through using wearable sensors to transmit data en masse), but may also be used individually for personal tracking.
Open-sourcing or peer production	Open sourcing is the development of data or materials that will become freely available, where there is often no clear 'call' to work. In crowdsourcing, an organization would initiate the work *(15)*.
Outsourcing	Crowdsourcing can be defined as a niche form of outsourcing *(2)*. However, unlike outsourcing more generally, there is no contract for crowdsourced work *(9)*.

Example of crowdsourcing used in health research and emergency situations: BioGames

BioGames

- Uses power of large crowds and gamification to analyze malaria smears.
- Accessible via Android device or computer.
- Players 'kill' malaria parasites on blood smears using a syringe.
- Collect healthy cells.
- Gamers have been able to reach 99% accuracy.
- Educational version also created.
- In the future, gamers or machine-learning algorithms might pre-screen positive or negative marked cells, with unclear samples sent to experts for diagnosis.

Examples of crowdsourcing used in health research and emergency situations: OpenZika, Frontline SMS and Ushahidi

OpenZika

- Called for people to volunteer their spare computing power.
- Ran 92,000 simulations of potential drug candidates for Zika.
- All data are open access.

Frontline SMS and **Ushahidi**

- Open-source participatory epidemiology programmes.
- **Frontline SMS:** enables users to request needs via SMS.
- **Ushahidi:** response to the aftermath of the Haiti earthquake in 2010.

Case study: *Use of Ushahidi in Haiti*

January 2010 – 7.0 magnitude earthquake struck Haiti

- Mass destruction in populous areas.
- Ushahidi deployed within four days.
- Provided vital info to responders.
- Opened an SMS service for Haitians to text their needs to a free SMS number, including for food, aid and medicine.
- Location visualized geographically using cell phone tower triangulation, Google Earth and Google Street Maps.
- Reports were triaged, and volunteers were able to text back.
- Over 25,000 text messages received, 3600 of which were actioned.

Ushahidi

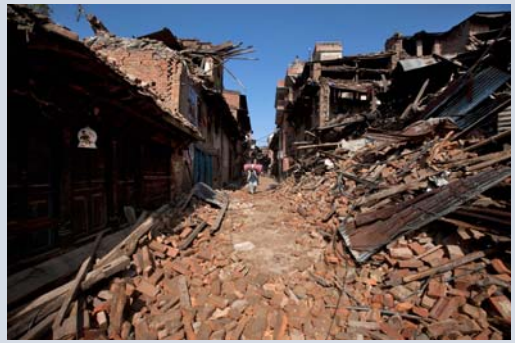
Examples of crowdsourcing used for mapping

In disaster relief settings, the most common use of crowdsourcing is for **mapping**. For example:

- Ushahidi, Frontline SMS, Missing Maps and Humanitarian Open Street Maps.
  - Create maps for disaster preparedness
  - Are able to work with crowdsourcing maps to enhance mapping capabilities to coordinate a response.
- This can be particularly helpful in countries prone to disasters where there is a a lack of accurate maps containing basic geographic information.

Case study: *Open Cities for disaster risk management in Nepal*

- Nepal is particularly exposed to disasters caused by natural hazards but most houses in the capital city, Kathmandu, do not meet minimum requirements for earthquake safety.
- Local stakeholders began using Open Street Map in 2012 to collect exposure data and map schools and health facilities and 2256 schools and 350 health facilities were mapped.
- In April and May 2015, two high magnitude earthquakes hit Nepal.
- Existing information was crucial in informing humanitarian responders and supporting recovery efforts.



Example of crowdsourcing used in participatory epidemiology

MoBuzz

- Participatory epidemiology application to combat dengue in Sri Lanka.
- Multi-component crowdsourcing application.
- Uses predictive technology and machine learning algorithms to detect:
  - Weather
  - Vector
  - Human data
- Produces hotspot maps for public and health officials.
- Civilians engaged to report breeding sites, symptoms and bites.
- Information is communicated widely to the public and health officials.

MOBUZZ
MOBILE AD BUZZAR

Example of crowdsourcing used in geographical sciences

Sapelli

- Used citizen science and crowdsourcing to map poaching in sub-Saharan Africa.
- Icon interfaces on a smartphone app.
- Suitable for people with low literacy.
- CyberTracker.
- Could be tailored to report a variety of relevant health outcomes:
  - Disease monitoring
  - Water and sanitation hygiene risk factors
  - Violence

What to consider when designing a study using crowdsourcing: crowd composition and knowledge

- Consider what type of crowd is needed, including whether tasks may require special knowledge or might rely on information from lay persons.
- Health-related crowdsourcing exercises requiring special knowledge include Innocentive and Crowdmed, which tackle complex pharmaceutical problems and the winner is rewarded financially.
- Examples in which lay persons can be very accurate include BioGames.
- The more diverse the crowd, the higher the probability of obtaining a 'smart' crowd.

What to consider when designing a study using crowdsourcing: hosting platform

- Several platforms exist to reach lay persons, including Amazon Mechanical Turk and Crowdflower.
- Consider whether people affected by the emergency will have sufficient access on their devices.
- Applications to consider include:
  - SMS
  - Specialist data collection tools for first responders (e.g., Open Data Kit)
  - OpenStreetMaps for remote helpers
- Consider whether data generated is comparable with current data management and storage systems.

What to consider when designing a study using crowdsourcing: crowd accessibility

- Crowd may be located in hard-to-reach area, which might be challenging for advertising the call.
- May be barriers to entry, such as cultural sensitivities.
- Specialist communities with needed knowledge might be difficult to reach, for example diaspora communities who are able to read messages from the affected population.

What to consider when designing a study using crowdsourcing: Renumeration

- Crowdsourcing primarily uses volunteered information.
- However, use of platforms such as Amazon Mechanical Turk to process tasks may require renumeration to the crowd.
- If crowd members are to be paid, it is very important to consider a pilot study for budgeting the costs of the submissions.

What to consider when designing a study using crowdsourcing: desired output

- The study or programme that will use crowdsourcing needs a clear question or purpose.
- This requires careful consideration of the type of task and the best way to combine submissions.
- May be ethical issues relating to sensitivity of data collected.
- Care is needed in how such data are collected, processed, stored and analyzed.

What to consider when designing a study using crowdsourcing: advertising the call

- Essential to ensure that the right crowd is reached.
- Call might be issued through:
  - Mass or social media campaign
  - Word of mouth
  - Targeted enrolment
- Important factors include:
  - Literacy
  - Local customs and culture
  - Reach of different media modalities

Study design, analysis and quality assurance

Study design and analysis

- Balance is needed between precision, speed and cost.
- The crowdsourcing needs to be appropriate to answer the research question.

Quality Assurance

- Important to obtain multiple measurements of the same thing, which can be triangulated to verify one another.
- Surveys might need to include questions designed specifically to identify 'malicious participants'.

Key messages

- Although crowdsourcing is still a nascent field, it has huge potential for Health EDRM.
- Crowdsourcing can be a low-cost, rapid alternative to traditional data collection methods.
- Crowdsourcing can be used to solve a variety of problems.
- Several open-source applications exist which can be used for crowdsourcing studies.

Further readings (1)

CrowdsourceEM Toolkit. www.crowdsourceem.org
Online crowdsourcing toolkit for emergency managers to use during disasters and emergencies.

Haklay M, et al. Identifying success factors in crowdsourced geographic information use in government. 2018.
World Bank review of volunteered geographic information (VGI) projects.

Heinzelman J, Waters C. Crowdsourcing crisis information in disaster-affected Haiti. Washington, DC: US Institute of Peace. 2010.
Discussion of crowdsourced information-sharing used in Haiti after the 2010 earthquake.

Mavandadi S, et al. Distributed medical image analysis and diagnosis through crowd-sourced games: a malaria case study. PloS ONE. 2012: 7(5): e37245
Discussion of the use of games for crowdsourced medical image analysis.

Further readings (2)

Wazny K. Applications of crowdsourcing in health: an overview. Journal of Global Health. 2018: 8(1): 010502.
Review of studies of the use of crowdsourcing in health care.
Ushahidi. 10 years of innovation. 10 years of impact. 2018. www.ushahidi.com/blog/2018/10/31/10-years-of-innovation-10-years-of-globalimpact
The creators of Ushahidi describe its history and impact.
Zook M, et al. Volunteered geographic information and crowdsourcing disaster relief: a case study of the Haitian earthquake. World Medical and Health Policy. 2010; 2: 7-33.
Description of the use of crowdsourced volunteered geographic information (VGI) tools in Haiti after the 2010 earthquake.

References

This chapter: Wazny K. Chapter 5.2: Crowdsourcing to gather data.
Models of crowdsourcing: Journal of Global Health. 2017: 7(2): 020602.
Crowd creation – Crowdsourcing information systems: Paper at the Australasian Conference on Information Systems 2011.
Crowdsourcing terms – Participatory epidemiology: Emerging themes in Epidemiology. 2017: 14(1): 2.
Crowdsourcing terms – The wisdom of crowds: New York: Random House. 2004.
Crowdsourced health research studies: Journal of Medical Internet Research. 2012: 14(2): e46.
Case study - BioGames: PLOS ONE. 2012: 7(5): e37245; Science Translational Medicine. 2014: 6(233): 233ed9; paper at the SPIE BiOS 2015.
OpenZika Project: www. worldcommunitygrid.org/research/zika/news.do
Ushahidi: World Medical and Health Policy. 2010: 2: 7-33.
Case study – Nepal: Remote Sensing. 2018: 10(1239): 1-7.
MoBuzz: documents.worldbank.org/ curated/en/387491563523294272/Identifying-Success-Factors-inCrowdsourced-Geographic-Information-Use-in-Government.
Sapelli: Trends in Parasitology. 2018: 34(9): 727-9; paper at the Proceedings of 17th European Conference on Computer Supported Cooperative Work. 2019.

Contact information

Health EDRM Research Network Secretariat
WHO Centre for Health Development (WHO Kobe Centre)
Email: wkc_tprn@who.int