Big Data: The New Face of Humanitarian Aid

Data-driven solutions to some of the world's most difficult and dangerous challenges

Robert Alexander Hoekman
510 Global, The Hague, the Netherlands

Key messages

- Big data is data at a scale beyond what most traditional processing applications can handle.
- In humanitarian organizations, big data can empower the responses to disastrous events with greater precision and speed.
- 510 Global is the data initiative of the Netherlands Red Cross (NLRC), which uses data analysis across a wide range of humanitarian activities.
- Analysis by 510 Global of an actual typhoon – Typhoon Haima in the Philippines – indicates that it is possible to use machine learning to construct reliable damage predictors.
- 510 Global operates with several institutions and corporations and is developing an arsenal of tools to to assist with today's and tomorrow's humanitarian needs.

Turn on CNN or stream a Ted Talk, and chances are it won't take long to hear the phrase "big data" mentioned in all sorts of contexts. It has become impossible to avoid the term when looking at the horizon of technological development and discussing our sociotechnical future. Its presence is also felt in humanitarian organizations, where big data can empower the responses to disastrous events with greater precision and speed. All this buzz is for good reason, for data gives us the opportunity to more fundamentally understand our surroundings. For those of us who observe the world from a data-driven perspective, it quickly becomes apparent that there are inherent patterns to most things in life. One must simply set the right boundaries and conditions to see them. Chaos is indescribable until observed – and like an atom, once observed, its properties are defined and therefore predictable. From "spooky" quantum action to hurricanes and flood dynamics, data is crucial to understanding and predicting the behavior of our external environment, and it is this greater understanding that 510 Global considers its ace in the hole.

> "Data is crucial to understanding and predicting the behavior of our external environment"

510 Global is the data initiative of the Netherlands Red Cross (NLRC). In humanitarian work, informed decision-making affects the health and safety of thousands of people, and it was in response to this need that 510 Global was formed. Made up of data scientists, humanitarian experts, information managers and researchers, 510 Global team their skills across a swathe of humanitarian activities.

Big data is simply data, but at a scale beyond what most traditional processing applications can handle, creating challenges for capturing, storing and analyzing it. It can be collated from a variety of sources, including openly available data, crowdsourced data, and private databases. Its most advantageous property is its sheer scale, allowing opportunities for advanced analytics and predictive models. These analytic toolsets themselves are also sometimes referred to as big data today.[1]

It is in the decision-making challenges that humanitarian activities face that big data can have significant positive impact. As seen recently with events in the United States with Hurricane Harvey, first responders needed to know the distribution of damage and risks across a vast area. Which communities are hardest hit and most isolated from help, which communities need emer-

gency resources such as food and water, which communities are at most risk, who do they help first and where are they needed most: these aren't questions with simple answers. A community with comparatively low levels of flooding may see high casualty rates due to other factors.[2] For decision-makers to be empowered, we therefore need to know more, and to know it earlier. This is where big data comes in, as seen in the following case of Typhoon Haima in the Philippines.

"Missing Maps is an open, collaborative project in which you can help to map areas where humanitarian organizations are trying to meet the needs of vulnerable people. The team at 510 coordinates mapathons in The Netherlands, in which volunteers come to map for good causes! We use a tasking manager to select and prepare the areas to be mapped, giving an introduction, helping people with mapping, organizing the validation of preliminary mapping results, and supporting people who live in the mapped area to improve and sustain their local maps."

Koos Krijnders, Missing Maps

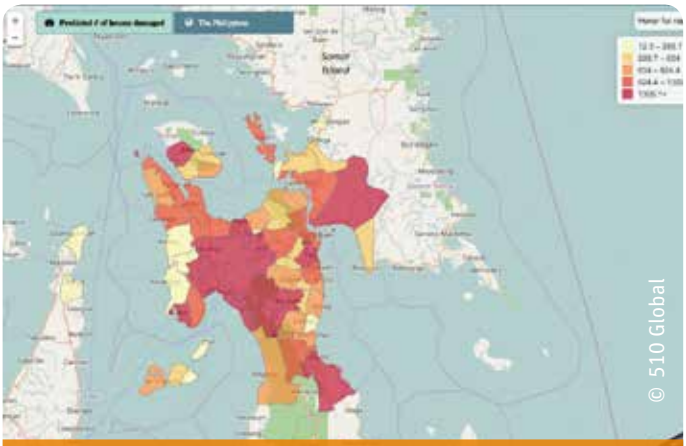
Fine-tuning big data for success: A case study in the Philippines

When the worst happens and a natural disaster strikes, it is crucial that local institutions and organizations have information on the resulting damage. To this end, any information given needs to be as accurate as possible while also being timely and directly appropriate for use. This is a clear example of the utility advantages of big data, as it can increase both the accuracy and speed of delivery of this information. Some of the work 510 Global does in this area can be seen in the creation of a Priority Index Model (PIM) which was tested in the case of Typhoon Haima, in the Philippines. Here, the main challenge to rescue and response operations was found to be the scarcity of personnel and resources.

Collecting data with Missing Maps. First responders often lack the necessary information to make effective decisions about resource allocation and location.

"One of the challenges with disaster response is scarcity of resources: not each affected family can be helped. Therefore it is essential to identify priority areas, by assessing damage and finding vulnerable people that are affected the most. Currently, damage assessments and identification of the most vulnerable is a time-consuming process, which can take weeks to complete, due to logistics, safety constraints, or workload."

Maarten Van der Veen, 510 INITIATOR



Visual Priority Index Model: Mapped damage to houses

To meet this challenge, 510 developed a methodology by which we can identify areas at greatest risk and therefore most urgently in need of humanitarian response. This methodology is reliant on largely open secondary data (received through collaborations with UCL, NASA and others) from affected areas as well as *in situ* data (rainfall, wind speed). Data from previous events is also considered. The aim is to integrate machine learning, which can be used to generate fast and accurate damage predictions for various events and countries. It does this by automatically applying what has been learned in past events to new incoming data on a current disaster. Through testing the effectiveness of the "machine", we can fine-tune its prediction capacity in order for it to get smarter. In the case of Typhoon Haima, data included population, poverty, house wall and roof type, and geographical features, as well as direct impact data from the event (rainfall, typhoon path, wind speed.) To assist in generating a valid index for damage, data from four previous typhoons was used.

The completed priority index model (PIM) was successfully released only 24 hours after the typhoon first hit the coastline. Four days later, the first official (non-PIM) damage counts were released and shared. Further analysis led us to conclude that it is indeed possible to use machine learning to construct reliable damage predictors. There is still work to be done – for example, by addressing the coping capacity of communities accurately, and by obtaining more accurate data on building damage. These elements impact the total accuracy of the model, and fine-tuning and further research are underway to address these. The process still needs to be scripted end-to-end. Eventually our goal is to scale up this methodology and have it available for PIM-generation in numerous countries.

Operating securely with big data and compliance with GDPR

New developments such as the PIM used for Typhoon Haima require the use of large amounts of data. Though currently personal data is not used, there will likely be some cases where data that we use will contain sensitive or personal information. We therefore have to address the challenges of using data in our work and avoid the potential dangers inherent in the new information age. The wide-ranging adoption of data-driven projects means new organizational frameworks and structures to ensure data privacy and security. With open data, remote-sensed satellite data, assessment data, and privately sourced data, the vast scope of information creates its own management and regulatory challenges. First and foremost, we must ensure that our foundational principles are upheld through the ethical and safe treatment of the information we gather ourselves and through external parties. In addition, new regulations and initiatives such as the EU General Data Protection Regulation (GDPR) are coming up on the horizon, and these bear an increasing relevance to humanitarian work as we come to depend increasingly on our predictive capacities to push the barriers of success. To this end, we need to future-proof our project structures in order to be prepared and in a position to handle the data we need without placing individuals or communities at risk.

Other projects with the 510 team

The 510 team is composed of volunteers, students and staff in a startup environment with the space and support to develop and test innovative technologies and ideas. Once ideas have been developed and tested successfully, they can be brought back into the Red Cross operations in the form of products or services. 510 operates with several institutions and corporations, ranging from Missing Maps to NASA Jet Propulsion Laboratory (NASA JPL), collaborating on, as well as supporting, research development and implementation. We are creating and developing

Radar satellite data: Using radar, we can penetrate dense cloud coverage and obtain an unobstructed view of the topography of a flooded area. This is crucial to understanding and predicting future floods.

an arsenal of tools to assist with today's and tomorrow's humanitarian needs. Among our ongoing projects, we are creating predictive damage modeling of floods and earthquakes, which should be able to identify at-risk areas and provide in indication of potential damage and areas of safety. We work to this end with NASA JPL, which provides radar satellite data capable of breaching cloud cover and gathering crucial intelligence on the topography of the flooded area. This project also involves several Dutch universities, from which our students are drawn.

Another large coverage project we are developing is the creation of a Community Risk Assessment (CRA) tool. This collates open data on a selected group of indicators based on vulnerability and coping capacity alongside geo-climatological data to prioritize areas in need for first responders should a disaster strike. With support of the Dutch Coalition for Humanitarian Innovation (DCHI), 510 is exploring in partnership with DSM how nutrition data can be integrated into this tool. We use the INFORM risk framework as a well-established index, and so have a standardized means of measuring risk and a consistent understanding of which communities need to be targeted for relief or for ongoing humanitarian programs.

> "We have an opportunity as a data-literate society to provide increasingly responsive and 'smart' relief and aid"

Migrating projects and institutions into a data-driven world

We are beginning to understand why data-driven solutions are not just where we are headed, but where we need to go. We have an opportunity as a data-literate society to provide increasingly responsive and "smart" relief and aid. The new direction of big data usage, however, does pose certain key questions which affect those indirectly involved in relief work. The implications of accurate and predictive analysis on drought, famine, and agricultural capacity are extensive and should interest anyone from policy-makers to the food security sector. This goes hand in hand with new risks in data storage, processing, and collection, as data responsibility becomes a constant presence in our lives. These are topics that will play a vital role in the success of migrating our projects and institutions fully into a data-driven world, and we must be ready for them.

Correspondence: ***The Team at 510,***
510 Global, Leeghwaterplein 27, 2521 CV Den Haag, the Netherlands ***Email:*** *support@510.global*

References

01. Boyd D, Crawford K. Six Provocations for Big Data. SSRN Electronic Journal, 2011. doi:10.2139/ssrn.1926431.
02. US Department of Commerce, National Oceanic and Atmospheric Administration. NOS Hurricane Harvey Response. NOAA's National Ocean Service. Version current August 31, 2017. Internet: https://oceanservice.noaa.gov/aa-updates/nos-harvey-response-083117.html.