

Workshop summary

Data analysis tools for the response to humanitarian health crises

Introduction

In response to the rapid evolution of disruptive information technologies, and the accelerated frequency of events that threaten human populations (wars, climate change, economic crisis), Crisis Ready, in association with the Universidad del Rosario and the Universidad Nacional de Colombia, organized a workshop to discuss, hand in hand with policymakers, the importance of the availability and timely use of high-quality data in response to humanitarian crises, with an emphasis on health emergencies. The participants of this workshop included multilateral organizations (OIM, PNUD, OPS, Red Cross, iMMAP), national NGOs (Karisma), academic research groups, government agencies (National Institute of Health, National Department of Statistics), and private corporations (Procalculo). The main goal was to discuss the need to build a network composed of multiple actors that facilitates the response to humanitarian crises.

Goals of the workshop:

- Build a network of key actors surrounding humanitarian crisis response.
- Identify the role data analysis can play at the different junctures throughout the development of a crisis.
- Identify the information, analyses, and data access needed that will facilitate resilience in times of humanitarian crises.

Panel 1: Use and nature of data

In order to answer questions about data representativity, resolution, bias, privacy, and possible uses, we first need to address two key questions:

- (1) Where did the data come from?
- (2) How is the data being aggregated?

Where did the data come from?

This question involves the type of data (e.g. satellite imagery, vector mobility data, population densities, etc) and the resolution and granularity levels.

Taking as an example cellphone mobility data, understanding the way data is collected is enough to infer the representativeness of the sample as well as possible biases (Nishant Kishore, Harvard University). Data that is collected through publicity banners in cellphone apps (e.g. Rappi) is by definition a sample of the app's users and it carries with it the biases that this subpopulation

represents (e.g. Who is more likely to be using Rappi?). This characteristic is not unique to cellphone-sourced data. Representativeness can also be inferred for remote sensing data, and it plays an equally important role in later interpretation. In the case of satellite or drone data, representativeness can be understood in terms of resolution (spatial, temporal, spectral, and radiometric) since sampling a given area will inevitably translate to compromises in terms of the object of sampling and its frequency. As an example, satellite data is notorious for having a higher sampling frequency in more developed areas of the world (i.e higher temporal resolution). What's more, countries like Colombia are often covered by clouds, making the sampling rate even less adequate and data for a given area less representative.

Data representativeness derives from biases built into how data is collected and should inform how this data is used.

How is the data being aggregated?

This question, although highly relevant is not always easy to answer. As an example, not understanding the process of aggregation for Call Detailed Record (CDR) data can bring us to misleading conclusions (Taylor Chin, Harvard University). A cellphone uses its closest cellular tower; CDR data consists of the location of the most often used cellular tower, over a given day, by a given cellular device. This means that the fact that one device shows the same geolocation over two consecutive days, does not necessarily mean that such a device stayed put over that time period. Or moreover, that the person associated with that device did. Beyond the logical conclusion that more than one individual can share a cellular device, the way that the geolocation is generated (the closest and *most often used* cellular tower in a day), dilutes the possible movements associated with that given cellphone throughout a given day. If an individual constantly uses their cell phone at home, day after day, the aggregation system will hide any other movement associated with that device. Each aggregation system translates into a particular level of resolution which will then dictate what can be inferred from the data.

Beyond the importance of aggregation systems to inform the nature of conclusions that can be drawn from data, they help address questions surrounding privacy. It is often the case that aggregation gets equated with anonymity, but in reality, aggregation is rarely enough to ensure privacy. For this reason, it might be necessary to resort to methods like differential privacy (DP) (Nishant Kishore, Harvard University) to complement aggregation schemes. However, by aggregating along markers of inequality (e.g. gender) it is possible to preserve a degree of privacy while allowing for the possibility of conducting a more granular analysis that could inform policy with a differentiated approach (e.g. taking into account gender dynamics) (Andrew Schroeder, Crisis Ready).

Panel 2: Migration and displacement

This panel addressed the question of how data can be used to deal with migration and displacement crises, and the type of infrastructures and human talent that are necessary to support this endeavor. We started to talk about the distribution of roles along the spectrum of actors (academia, industry, policy-makers) and the wide range of challenges that exist in connecting them.

Andrew Schroeder (Crisis Ready) presented the work that was done in Ukraine to complement official migration records. Using data from Meta Platform, it was possible to get an idea of the cities where there was a higher flux of refugees to better inform resource allocation. The use of data for planning purposes was also exemplified in the work done by UNHCR in Brazil. Where data from the Fundación C aritas is being used to model, in semi-real time, shelter capacity and availability for Venezuelan migrants and refugees. These models can be enriched by additional data signals that expand the range of conclusions that can be drawn. In the case of Ukrainian refugees, Crisis Ready is complementing population growth data with information regarding social connectivity (inferred from Facebook friendships). This allows the model to predict not only the final destination of refugees but the likelihood that they will remain there.

These are both examples of data flowing from vulnerable populations, to relief and government agencies, but this is not always the case. For instance, the application GIFMM “GIFMM contigo” consolidates the offer of relief services and provides this information to Venezuelan migrants and refugees. It is important to highlight that although privacy is important in any exercise of data-use when there is data captured directly from the devices of vulnerable populations, these considerations are particularly relevant.

The second vein of discussion during this panel was regarding the need for robust infrastructure and human talent to support data-driven decision-making. Diana Gualteros (National Health Institute, INS, Colombia) gave an overview of the system in place for epidemiological surveillance developed by the INS. The system monitors 108 signals of an array of public health events as part of an early alert system. These signals are evaluated in the risk analysis room to decide if it is necessary to activate the immediate response teams. This process relies on appropriately trained individuals, making the training branch of the INS an equally important program. Humberto Mendoza (Secretary of Health, Barranquilla Colombia) reminds us that this also means that it is paramount to guarantee the continuity of personnel across changes in governments.

Panel 3: Data use in crisis response and disease transmission

Traditionally, data flows and protocols are established to articulate industry, academia, and the government. The Risk Analysis Room in Cartagena, Colombia is a good example of this. Here, data

on aeronautical traffic, social media posts, and special communities are monitored daily (Luisana del Carmen Cárcamo Marrugo, DADIS Cartagena), and there are protocols in place to involve non-government specialists when the situation demands it. However, the need for new data protocols may arise with new emergencies. This was the case in Chile, where the need for wider collaboration to deal with COVID-19, became apparent, and in response, the Minister of Science and Technology made COVID-19 data publicly available, enabling several academics and researchers to further advance the response to the pandemic. The research presented by Pamela Martinez (University of Illinois Urbana Champaign) was made possible precisely because of the availability of such. Her research explored the differential factors surrounding socioeconomic status during the COVID-19 pandemic. This type of research, which elucidates the link between social factors and health burdens, plays a crucial role in guiding the government's efforts toward specific and focused actions. As an example, Pamela's team analyzed how socioeconomic factors translate to a correlation between infection rates and mortality in Santiago de Chile (i.e. mobility, test rates, case underreporting). This allows decision-makers to focus their efforts on acquiring better data to analyze the relevant factors involved in the contagion, which in turn allows for better resource allocation.

Nevertheless, in order to ensure successful strategies for emergency contingency and crisis mitigation, it is not enough to involve government entities and academic groups. Knowledge translation plays a crucial role when it comes to explaining the flaws associated with data-driven policy-making. It is precisely this problem that brings Felipe González y Andrea Parra (Universidad Nacional, Universidad del Rosario) to redefine information technologies. Usually, information technologies are regarded only as monitoring or surveillance systems. An example of such a system is the OCHA monitoring system, which is able to collect, process, and visualize data (Silvia Echeverri, OCHA). Although this tool is useful for both the general public and decision-makers, sometimes its information is not easily translated into specific policies. Andrea Parra and Felipe González propose the use of information technology as a design engine for public policy. This implies the construction of portable, scalable, and flexible systems that facilitate collaboration between the public sector and academia. This collaboration enables the co-creation of relevant metrics and methodologies that enable decision-makers to act based on data and state-of-the-art analytics. IMMAP realizes an example of such collaborations between actors. Jeffrey Roberto Villaveces (IMMAP) presented a case study where satellite imagery is used to detect new settlements. Even though this project follows traditional paradigms (imagery intake, model training, verification, on-field validation, and publication) one of its branches focus on community mapping over 3D maps. This collaboration between community and academia facilitates knowledge transfer and assimilation.

Panel 4: Privacy and data providers

This panel section discussed the tension between the risks around data privacy on the one hand and the “intent to do good” (Carolina Botero, Fundación Karisma). Evidence was shared of moments when, even though data use was well intended, loose privacy protocols lead to privacy abuse and tension between different stakeholders. As a concrete example, the mayor’s office in Bogota, with the support of several academic groups, used cellular data to train super-spreading models. These models were intended to mitigate contagion without having to turn to strict quarantine measures during the COVID-19 epidemic. In other cities around the country, however, similar data was used by local authorities to identify individuals and pay home visits, which is considered a serious violation of privacy rights. “Normalizing the idea that all data should be available always for any use” is one of the dangerous narratives that the organization Karisma warns about. This is a narrative that can take hold during emergencies, given that as Edna Margarita Valle (Vital Statistics Coordinator, DANE, Colombia) well put it, in moments of crisis, governments are under tremendous pressure to protect and serve the community with little to no available information. It is here where one should be reminded that there is a difference between data used by the government and data used by public entities or academic organizations. One of the key roles played by academia during a crisis is to act as an ambassador for privacy, transforming sensitive data into actionable and anonymized information that governments can use for policymaking. To accomplish this goal, Carlos Ahumada (Meta Platform) brings forward a design paradigm called “privacy by design”. Data aggregation is rarely enough to guarantee anonymity; therefore, it is necessary to include tools like differential privacy or federated learning. The former refers to a technique where data is translated into distribution, which in turn is mixed with noise and later sampled to achieve anonymity without compromising analytic potential. The latter refers to a machine learning technique where several machines or nodes are used, forcing the algorithm to be distributed and only aware of local data, instead of the entire database. Tools like these are important to close the translation gap between data and public policy, but they are only the first step. Zulma Cucunubá (Universidad Javeriana, Imperial College) describes this process as an iterative one, where the first objective is to identify the problem so that models and metrics can be codeveloped. This leads to the next step which is the use of anonymized data for model validation and result dissemination. This communication should already be translated into public policy language and should include the concrete implications of such policies. Finally, a period dedicated to personnel training and knowledge transfer is established. With this, the models stay in the hands of the public entity that will be using them for decision-making and thus, concluding the cycle of knowledge translation.

Design thinking workshop

Given that one of the objectives of the workshop was to understand the network of actors in Colombia who are involved in a crisis situation, a mapping exercise using design thinking methodologies was designed to crowd-source these networks from the participants. The workshop was split into 4 activities:

- Actor-network map: participants were asked to create a list of key actors during an emergency. By using the idea of Jungian archetypes as a way to guide thinking, participants were asked to identify actor roles and interactions between them.
- Process mapping: participants were asked to build a timeline of a “typical” response to a humanitarian crisis.
- Data dashboard design: participants were asked to engage in a mapping exercise where specific information needs were matched with specific types of data visualization.
- Data journey: based on the products of the three previous exercises, participants were asked to consolidate the actors, processes, and information needs in the form of a *data journey*. This had the goal of identifying different paradigms of response to crises and the role that data analytics play in each one.

Based on the workshop the following flow describing the development of a crisis was built as a synthesis exercise.

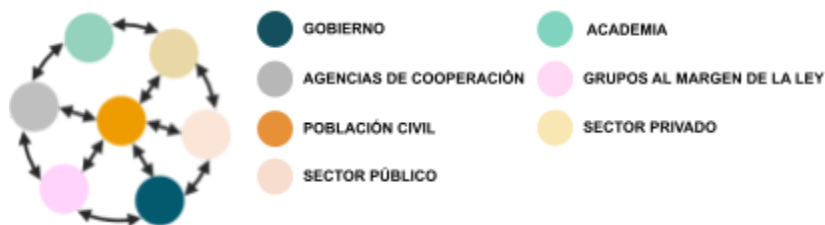
1. Previous context and vulnerability analysis

Before a crisis occurs, there already is a setting where there a series of systematic relationships among actors exist. This refers to the underlying dynamics between actors, a product of the different systems in place. When the time to deploy emergency response comes, these dynamics can influence, overtly or not, actor interactions. As an example, exclusion dynamics towards marginalized populations can translate into exclusion dynamics when deploying relief. The responsibility of the actors involved during this step is to study the existing context and power dynamics to identify vulnerabilities and create comprehensive response protocols. This stage of the crisis will determine what data is readily available, and what protocols and data-sharing agreements are in place to be used in a timely fashion.



2. Triggering event

The triggering event is the event that unleashes a series of calamities. At this stage of the crisis, all actors are interacting with each other. Given the chaos present at the beginning of a crisis, this initial interaction is mainly guided by the protocols established in the previous stage and whatever other systematic factors affect actor coordination. This highlights the importance of the previous stage, where the conscious development of comprehensive and inclusive protocols will allow for a first-response scenario that is smooth and equitable. Academia is usually the only actor that does not directly interact with the general population.



3. Primary response

Primary response is deployed and a period of crisis mitigation starts. The need for situation-specific response starts to play an important role. The availability of real-time data to guide relief efforts (e.g. mobility, vulnerability, and connectivity data...) play a key role in helping to optimize resource allocation. For this to be effective it is essential that during stage (1) the necessary infrastructure for the collection, processing, and analysis of data is in place. Privacy protection protocols developed in stage (1) will also be important, to avoid creating dangerous precedents regarding abuses of privacy during the urgency of primary response.



4. Design and strategy

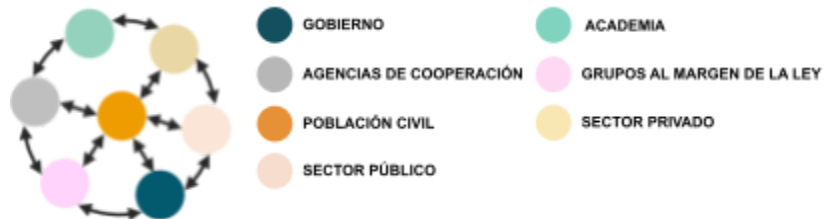
Actors responsible for the strategic design of crisis response collaborate to identify a sustainable mid-term and long-term strategy that represents a comprehensive solution to the crisis. Here there is an opportunity for knowledge co-creation between academia, government, cooperation agencies, and the private sector. The practice of knowledge translation is fundamental as it will determine to what extent state-of-the-art scientific methodologies will be integrated into the process of policymaking. In terms of data use,

academia should act as an ambassador of privacy and the ethical use of data. Participants of the workshop identified the need and difficulty to convince the private sector to share data in the first place. This makes coming up with incentives to encourage data sharing a key necessity.



5. Comprehensive response

The solution designed previously is deployed. A comprehensive solution to a crisis requires that all actors are involved and that there is a dynamic collaboration among them. It is fundamental that the practice of knowledge translation that started in the previous stage continues. The type of data that will be relevant in this stage of the crisis is likely to be different from the data that was identified in stage (2). Here, forecast methodologies and models, as well as data with longer time windows help give a more complete picture of the situation.



6. Evaluation and redesign

Once the crisis is under control, the actors in charge conduct an impact assessment and evaluate the response to the crisis. Lessons are collected and analyzed to update contingency plans and protocols. This is a good moment to re-evaluate existing data-sharing agreements and come up with new ones. During the workshop, it was suggested that the academia and the government should collaborate on creating joint methodologies centered around collaboration weak spots that became evident during the crisis. Participants also emphasized the harm in having actors go back to an insular M.O. once the crisis is over since this perpetuates poor intersectoral collaboration.



Conclusion and next steps

The symposium was the first step in generating a network of actors and identifying the processes around data use for humanitarian crisis response. Participants identified friction points between government goals and privacy needs, and conflicts between timelines of academic research and crisis response needs. But opportunities for creating a robust ecosystem for crisis response were also identified. There is ample room for data analytics to guide crisis response in Colombia, but this will not translate into anything without clear collaboration and protocols, and data-sharing agreements. The need for incentives for the private sector to encourage data sharing remains a key obstacle in making the data necessary for timely analysis available to researchers and governments.

On the other hand, the importance of human talent accompanying information technologies became evident through the workshop. These professionals are tasked with the challenge of translating knowledge between researchers and policymakers. This highlights the need of assigning resources and personnel to intersectoral articulation efforts. This personnel should be trained in public policy as well as data science to close the gap that currently exists.

Participants also highlighted the importance of holding these kinds of network-strengthening events with certain periodicity to solidify ties and establish a smooth dialogue between stakeholders to improve resilience in humanitarian crisis response.

Instituciones participantes

- Agencia Analítica de Datos-AGATA-
- Agencia de la ONU para los Refugiados (ACNUR)
- Crisis Ready
- Cruz Roja Colombiana
- DataLama
- Departamento administrativo distrital de salud (DADIS) Cartagena Secretaría de Salud del Distrito de Bogotá
- Fundación Karisma
- Grupo Interagencial de Flujos Migratorios Mixtos (GIFMM)

- Harvard T.H.Chan School of Public Health, Harvard University
- iMMAP
- Instituto Nacional de Salud (INS)
- Ministerio de Salud y Protección Social de Colombia
- Oficina de las Naciones Unidas para la Reducción del Riesgo de Desastres (UNDRR)
- Oficina de Naciones Unidas para la Coordinación de Asuntos Humanitarios (OCHA)
- Organización Internacional para las Migraciones (OIM)
- Organización Panamericana de Salud (OPS)
- Procalculo
- Profamilia
- Secretaría de Salud de Guainía
- Secretaría de Salud Distrital de Barranquilla
- United Nations High Commissioner for Refugees (UNHCR)
- Universidad de Los Andes
- Universidad El Rosario
- Universidad Javeriana
- Universidad Nacional de Colombia

Event transmission can be found here:

1. <https://www.youtube.com/watch?v=qvJm3ggPT4g&feature=youtu.be>
2. https://www.youtube.com/watch?v=iPM_70oKacU