

## Creating humanitarian big data units

Global Pulse has done a fine job of making humanitarian big data visible both within and outside the UN. But it's a big job, and they won't be able to do it on their own. So. What, IMHO, would another humanitarian big data team need to be and do? What's the landscape they're moving into?

### Why should we care about humanitarian big data?

First, there's a growing body of evidence that data science can change the way that international organisations work, the speed that they can respond to issues and degree of insight that they can bring to bear on them.

And NGOs are changing. NGOs have to change. We are no longer organizations working in isolation in places that the world only sees through press releases. The Internet has changed that. We're now in a connected world, where I work daily with people in Ghana, Columbia, England and Kazakhstan. Where a citizen in Liberia can apply community and data techniques from around the world, to improve the environment and infrastructure in their own cities and country.

We have to work with people who used to be outsiders: the people who used to receive aid (but are now working with NGOs to make their communities more resilient to crisis), and we have to work with data that used to be outside too: the tweets, blogposts, websites, news articles and privately-held data like mobile money logs and phone top-up rates that can help us to understand what is happening, when, where and to whom.

UN Global Pulse was formed to work out how to do that. Specifically, it was set up to help developing-world governments use new data sources to provide earlier warnings of growing development crises. And when we say earlier, we mean that in 2008 the world had a problem. Three crises (food, fuel and finance) happened at once and interacted with each other. And the first indicator that the G20 had was food riots. The G20 went to the UN looking for up-to-date information on who needed help, where and how. And the UN's monitoring data was roughly 2 years out of date.

### What have we done so far?

So what are the NGOs and IAs doing so far? The UN has started down the route to fix this with a bunch of data programs including Global Pulse and FEWSnet. Oxfam connected up to hackathons last month; the Red Cross has been there for a while. The World Economic Forum has open data people, as does the World Bank. And other groups as different as the Fed and IARPA are investigating risk reduction (which is the real bottom line here) through big data techniques.

## What should we be doing?

But what do the NGOs need to do as a group? What will it take to make big data, social data, private data, open data and data-driven communities useful to risk-manage for crises?

### 1. First, ask the right questions.

When you design technology, the first question should be “what is the problem we’re trying to solve here?” Understand and ask the questions that NGOs do and could ask, and how new data could help with them. There is data exhaust, the data that people leave behind as they go about their lives: focus on the weak signals that occur in it as crises develop. Reach out to people across NGOs to work out what those questions could be.

### 2. Find data sources.

We cannot use new data if we don’t have new data.

Data Philanthropy was an idea from GFDI to create partnerships between NGOs, private data owners like the GSMA mobile phone authority and other data-owning organisations like the World Economic Forum. Data Commons was a similar idea to make data (or the results of searches on data – we want to map trends, not individuals) available via trusted third parties like the UN. It’s gone a long way politically but still has a lot of work to be done on access agreements, privacy frameworks and data licensing.

Keep encouraging the crismapping and open data communities to improve the person-generated data available to crisis responders, to improve the access of people in cities and countries to data about their local infrastructure and services, and to voice their everyday concerns to decision makers (e.g. via Open311). Encourage the open data and hacker movements to continue creating user-input datasets like Pachube, Buzz and CKAN. All this is useful if you want to understand what is going wrong.

### 3. Find partners who understand data.

Link NGOs to private organisations, universities and communities who both collect and process new types of data. Five of these recently demonstrated Global Pulse led projects to the General Assembly:

- Jana’s mobile phone coverage allowed us to send a global survey to their population of 2.1 billion users in over 70 countries. There are issues with moving from household surveys that need to be discussed, but it allowed us to collect a statistically significant sample of wellbeing and opinion faster and more often than current NGO systems (an authoritative survey I read recently had 3500 data points

from a 5,000,000 person population. Statistical significance: discuss).

- Pricestats used data from markets across Latin America to track the price of bread daily rather than monthly. Not so exciting in 'normal' mode or in countries where prices are regularly tracked. Incredibly useful during recovery or for places where there is no other price data gathered.
- The Complex Systems Institute from Paris tracked topics emerging in food security related news since 2004. This showed topic shifts from humanitarian issues to food price volatility (with children's vulnerability always being somewhere in the news). More of a strategic/ opinion indicator, but potentially incredibly useful when applied to social media.
- SAS found new indicators related to unemployment from mood changes in online conversations – several of which spiked months before and after the unemployment rate (in Ireland and the USA) changed. This gave new indicators of both upcoming events and country-specific coping strategies.
- Crimson Hexagon looked at the correlation between Indonesian tweets about food and real food-related events. The correlations exist, and mirrored official food inflation statistics. Again, useful if gathered data isn't there.

And reach out to the communities that are forming around the world to process generated data, from the volunteer data scientists at Data Without Borders to the interns at Code for America and the GIS analysis experts connected to the Crisismappers Network.

#### **4. Collect new data techniques and teach NGOs about them.**

There is a whole science emerging around the vast ocean of data that we now find ourselves swimming in. It has many names, Big Data and Data Science being just two of them, but it's basically statistical analysis of unstructured data from new sources including the Internet, where that data is often very large. Learn about them, play with them (yes, play!), and teach people in NGOs about how to use them. The list of things you probably need to know include data harvesting, data cleaning (80% of the work), text analysis, learning algorithms, network analysis, Bayesian statistics, argumentation and visualization.

And build a managed toolkit of open-source tools that NGOs and analysts in developing country can use. For free. With support. Which doesn't mean "don't use proprietary tools" – these have a major part to play too. It just means that we should make sure that everyone can help protect people, whatever the funds they have available are.

#### **5. Design and build the technologies that are missing.**

Like Hunchworks. Hunchworks is a social network-based hypothesis management system that is designed to connect together experts who each have part of the evidence needed to spot a developing crisis, but don't individually have enough to raise it publically. It's a safe space to share related evidence, and give access to the data and tools (including intelligent agents automatically searching data for related evidence) needed to collect more. It's still in alpha, but it could potentially help break one of the largest problems in development analysis: namely, the silos that form between people working on the same issues and the people that need to see their results.

## **6. Localize.**

Build labs in developing countries. Build analysis capacity amongst communities in developing countries. People respond differently to economic stress, and environments, data sources and language needs are different in different countries. The labs are there to localize tools, techniques and analysis, and to act as hubs, collectors and sharing environments for the types of minds needed to make this work a reality. No one NGO can afford to do this in all countries, so connections between differently-labelled labs will become vital to sharing best practice around the world.

## **7. Publicise and listen.**

Be there at meetups and technology sessions, at hackathons and in Internet groups, listening and learning to do things better. And never ever forget that this isn't just an exercise. It's about working better, not building cool toys – if the answer to a problem is simple and low-tech, then swallow your pride and do it - if the answer is to share effort with others to get this thing worker faster to protect people around the world, then do that too. We do not have the luxury of excessive time or meeting-fuelled inaction before the next big crisis strikes.