

New development data

I've spent almost two years now thinking about and doing [humanitarian crisismapping](#) – about the sources, analysis and communication of information available to organisations like the [UN](#) before, during and after a natural disaster (fire, earthquake, tsunami, floods, snows etc), but that's too diverse and uncertain for a small number of people in these organisations to distill into usable, searchable knowledge in the timeframes involved in any disaster (seconds, minutes, hours, days before it's all over and all responders can do is help people to recover from its effects).

There have been some amazing things done in those years. The [UN cluster system](#), although imperfect and not mandatory to join, does have a structure that field organisations can join to improve their combined effort in an area (nutrition, health, water/sanitation, emergency shelter, camp coordination, protection, early recovery, logistics, telecomms). Teams like UN OCHA, the standby task force, crisiscommons, sahana, humanitarian open street map, humanity road and mapaction are building maps and situation reports from data across the internet and reported from disaster zones, have helped open up new data sources ranging from near-real-time satellite images to gazetteers and citizen journalists, and built new tools, processes and techniques to do this.

But quietly, behind the big stories like Haiti and Libya, there's also been a lot of work on how to map and use new sources of development data. The same sources, tools and techniques can also be used to map [water availability](#), [health outbreaks](#), [famines](#) and [migrations](#).

But there's more. The crisis mappers have been using people (lots of people, all over the world) to hunt through open data streams like twitter, facebook, website updates, news streams and satellite images, and creating tools to help with searching and tagging information in these (this is exactly the type of search through data that happens as part of [big data](#) analysis). But also out there are slower forms of datastream on the internet, like website contents, updates to blogs and open data from governments, companies and institutions. And data with a lower psychological distance but more uncertainty as to intent, like [trends](#) in the [searches](#) made on google. And if you're part of an organization, you also have access to your [institutional data](#) (and incidentally should be asking yourself what the barriers are to making this data, a cleaned version of it, or results from analyzing it, [openly available](#)), and to data from other institutions and companies. You might also have access to sensor outputs and people's knowledge and opinions.

And more. A lot of the big data systems out there are designed to infer people's behaviours and status from the text (twitter, facebook etc) and data (phone positions etc) they output. Which is great, but a bit one-way and unfocussed. Which is why crowdsourcing systems like ushahidi and open311 become important. And just one of the reasons why the hunchworks system is being built.