How much data we have in DAP?

When we started DAP almost a year ago we created the Common Humanitarian Dataset. The CHD, as we call it, intends to create a collection of indicators that are both sustainably collected by authoritative institutions and that are already used by humanitarian actors in their analysis. For example, the CHD has nutrition indicators from FAO, poverty indicators from the World Bank, and basic health indicators from the World Health Organization.

In other words, DAP uses an *indicator based approach* to populate its repository. The idea is that the CHD will be community-driven throughout time, having a community of humanitarian analysis questioning old indicators, adding new ones and so forth. It provides a common-ground for conversation, but still leaving the system open to be populated by any other data that is considered relevant.

The full list of indicators currently available in the CHD can be found here: <http://goo.gl/Id4jiQ>

*(Please note that the list is a refined draft – there is a lot of work to be done there).*

Based on the CHD we partnered with ScraperWiki, a company from the UK specialized in scraping data from the web. Out friends   
**Ian Hopkinson** and **Dave Mckee**, both fantastic data scientists at ScraperWiki, build scrapers to collect data from a number of indicators from the CHD. In short, the scrapers are scripts that go to a certain website and “extract” the data out of those websites. They, then, put the data into a single database using unique identifiers – in our case countries’ ISO3c codes.

Thanks to the work our friends from ScraperWiki have done we now have collected a good amount of authoritative data commonly used to understand countries that have a humanitarian crisis. Here is on Ian’s words:

In total we’ve collected over 300,000 data points for 240 different regions (mostly countries – plus dependent territories such as Martinique, Aruba, Hong Kong, French Guiana and so forth), covering a period of 60 years with 233 different indicators.

From https://blog.scraperwiki.com/2013/10/us-and-the-un/

The graphic bellow shows how much data we have in our system today:

In the charts bellow, you can also see how much data we have per region:

So, what can we do with authoritative data?

The great thing about *most* of the sources we have collected is that they are, for the most part, comparable across time and among countries. (I’ll let Javier Teran -- our statistician -- explain in a later post why some of the sources are not comparable.) For example, you could analyze indicators very commonly used in the context of humanitarian response across a large period of time such as Crude Mortality Rate:

Above you can see that in the 24 countries that OCHA has field operations the CMR has been steadily decreasing over the last 50 years. The Central African Republic and the Democratic Republic of the Congo are notable examples with spikes in XX and XX.

Also, we could combine at two indicators and visualize how, for example, poverty is correlated with migration waves (left) or access to clean water (right):