Abstract

ReliefWeb is one of the largest repositories of "information products" in the humanitarian community and one of the most visited websites of the UN system. By sampling and analyzing ReliefWeb's trove of around 550,000 humanitarian "Reports, analysis documents, infographics, and other content", we identified a set of data-based indicators that are the most referenced by the users of the service: the number of people killed in disasters, the number of those affected by varied health threats, humanitarian financing figures, and the number of refugees, IDPs and asylum seekers. Considering ReliefWeb's size, mandate, and legacy we believe that those indicators can be representative of what data is most commonly used in the humanitarian community.

Based on those findings we are introducing and opening for debate the concept of the Common Humanitarian Dataset (CHD). The CHD envisions to create a single, fundamental dataset of humanitarian-relevant data, that is, data that could be used to understand humanitarian crises in the different stages of their existence. With this paper we are introducing the first public version of the CHD.

Methodology

The first public version of the CHD uses a number of authoritative sources as proxies to its data for a number of reasons.

We collected systematic samples of ReliefWeb's six most recent years of reports (from 2008 through 2013).  The data was collected using ReliefWeb's newly-developed API in combination with an open-source R package created for the purposes of the research. All the source-code and data will be made publicly available on GitHub.

That exercise allowed us to identify two data-sources as able to supply around a third of the need for data in the humanitarian community: UNHCR's "people of concern" database alongside OFDA/CRED's International Disaster Database. Other proxy sources which could cover significant demand for data are WHO, OCHA's Financial Tracking Service, UNICEF, the World Bank Group, and UNDP.