Unpacking Data Preparedness from a humanitarian decision making perspective: toward an assessment framework at subnational level

Marc van den Homberg
510.global, Netherlands Red Cross
marcjchr@gmail.com

Jannis Visser
510.global, Netherlands Red Cross
jannisvisser@redcross.nl

Maarten van der Veen
510.global, Netherlands Red Cross
mvanderveen@redcross.nl

ABSTRACT

All too often the collection as well as analysis of data for humanitarian response only starts once a disaster hits. This paper proposes a framework to assess Data Preparedness on five dimensions: Data Sets, Data Services and Tooling, Data Governance, Data Literacy, and Networked Organizations for Data. We demonstrate for one dimension, i.e. Data Sets, how it can be quantified. First step is to determine which Data Sets must be collected before a disaster strikes so that as many as possible decision-makers' information needs are covered. Subsequently, a *Data Sets Preparedness Index* can be calculated based on Completeness, Recency and Accuracy & Reliability. We tested the index for Malawi and The Philippines and show how it can be used to direct data collection and determine when data analysis for e.g. predicting severity becomes meaningful. The index can be modified for reporting on global policies such as the Sustainable Development Goals.

Keywords

Data preparedness, humanitarian response, information requirements.

INTRODUCTION

Humanitarian decision-makers are working in stressful, high-pressure conditions where information is often lacking, distorted or uncertain. These conditions are known to introduce or enforce biases (Comes, 2017): cognitive biases resulting from simplifications to deal with complex problems and/or motivational biases resulting from the desire for a specific result. Examples are overestimating the number of affected people to push funding in a certain direction or underreporting for political reasons. The availability of accurate, reliable and timely information can reduce these biases and lead to improved decision making. It is therefore essential to know what the decision-maker's information needs are (Gralla et al, 2015). A case study on the 2014 floods in Bangladesh (van den Homberg et al., 2016) identified seven clusters with in total 71 information needs. This study did not only identify the information needs, but also mapped the data sets that became available from the start of the floods onwards, on the needs. 15 data sets with in the order of 40 to 60 indicators each were identified and they could meet only 27% of the information requirements in time (and 62% if timing constraints were not considered). This data and information gap enhances the biases in humanitarian decision making.

The rapidly changing and new information environment consisting of mobile services, social media, crowdsourcing and collaborative digital spaces offers several opportunities to narrow this gap as much as possible both before and during a disaster. One can more easily engage with communities and collaborate among different stakeholders. Governments can use the new digital technologies to open and share their data. Humanitarian donors push their recipients to do so as well using International Aid Transparency Initiative (IATI). Increasingly paper-based processes get digitized. These trends lead to a rapid increase in availability of data. In addition, the key global agreements that entered into force in 2015, i.e. the Sendai framework for Disaster Risk Reduction (DRR), the Paris Climate Agreement and the Sustainable Development Goals (SDGs), all require data collection on a number of indicators of which several are relevant also for humanitarian contexts.

Data Preparedness will be pivotal for seizing these new opportunities. We define Data Preparedness as all activities, that can be done before a disaster hits, to pre-stage data with sufficiently high data quality (that matches the prospective information needs of responders) and to develop capacities to collect data on affected communities and areas once a disaster hits to ensure a timely, efficient, and effective response. Data Preparedness can be considered an element of information management preparedness. OCHA defined already in 2009 an information management preparedness matrix, that covered general preparedness, capacity, data standards, data sets and tools and needs assessments (OCHA ROAP, 2009). Raymond and Al Achkar (Raymond and Al Achkar, 2017) proposed a data preparedness cycle, departing from the risks that the use of new technologies and data poses. The cycle has therefore a strong emphasis on legal, ethical and regulatory rules and norms. However, there is not yet to the best of our knowledge a Data Preparedness Framework that can be used to assess and quantify the level of data preparedness and that has been tested in practice.

METHODOLOGY

The objective of this paper is to develop such an assessment framework and test it in practice. It should be possible to apply the framework to an individual organizational but also to the humanitarian ecosystem of a country. Our methodology consists of desk research combined with case studies on typhoons in The Philippines and floods in Malawi. The case studies combine remote with in-country activities. The remote activities involved the organization of mapathons as part of Missing Maps and identifying data providers and data sharing platforms. For the Philippines, the in-country activities consisted of 32 interviews as part of an MSc study (van Lint, 2017) and a Surge Information Management Support deployment to Manila shortly after Typhoon Haima at the end of 2016, whereby feedback from humanitarian decision-makers of The Philippines Red Cross and OCHA was obtained. The in-country activities in Malawi consisted of a Data Preparedness Mission in February 2017, whereby a workshop was held within the Malawi Red Cross. Over 20 stakeholders (mainly government and NGOs) were interviewed and one week of mapping with about 15 enumerators was organized.

The research described is part of a larger Data Preparedness project led by The Netherlands Red Cross 510 initiative. 510 aims to drive the smart use of (big) data for faster and more (cost) effective humanitarian aid. The team works with knowledge institutes, private sector and other NGOs to ideate, research, implement and scale up new data-driven humanitarian aid innovations and to increase the capacity in the sector to work with and understand data. In this project, subnational disaster management data on several hazard prone developing countries is gathered and integrated into an easily accessible online dashboard, i.e. the Community Risk Assessment tool. This tool feeds into another tool that predicts severity and priority areas for humanitarian aid.

The outline of the paper is as follows. In the next section, we present the framework components at a general level. Subsequently, we focus on one component, i.e. the Data Sets. We describe how one can determine which data to collect and collate and develop, subsequently, a method to numerically measure progress. This method is applied for The Philippines and Malawi. We discuss these first test results and describe the implications for further use. Future work aims at developing an assessment framework for the other four components.

A FRAMEWORK FOR DATA PREPAREDNESS

Our framework for Data Preparedness consists of five components. For each component, we describe a set of questions that can be answered at an organizational or ecosystem level. We used this set of questions also in an interactive round table session, where groups of people rotate from one table to another, dealing at each table with one of the components.

Data Sets: Which data in relation to disaster management does your organization collect? Do you use a framework with indicators for this and what are your information needs? Which gap is there between your information needs and the data that is available to you? Do you have an overview of those data providers that will be important for you once a disaster strikes? For example, during the Typhoon Haiyan the international community did not have the automatic reflex to request data on cities from mayors.

Data Services and Tooling: Which services does your organization offer based on its data sets? Which tooling (software, hardware, but can also be paper-based) does your organization use to collect, analyze, and share data? Which tooling does your organization use for collaboration with other organizations and/or dissemination (like geospatial sharing platforms and collaborative digital tooling)?

Data Literacy: Do you have training programs for your employees in relation to data? Do you face obstacles in terms of lack of data literacy at several hierarchical levels within your organization? How do you assess the level of data literacy within your organization or possibly also of the partners you work with? Do you have an HR policy that attracts data literate staff?

Data Governance: What is the mandate of your organization in terms of data for disaster management and/or the business rationale? Do you have specific guidelines in place in relation to data collection, analysis and sharing? How do you safeguard privacy and ensure sensitive data is handled responsibly? How are data harms prevented from occurring?

Networked Organizations for Data: With which organizations do you coordinate or collaborate in terms of data? With which organizations do you share data or from which organizations do you get data? Do you have an open data policy and are you actively sharing data online? Have you reached agreements with others for datasets that cannot be shared openly? We note that these latter points overlap with Data Governance.

Table 1 Defining the different components when going from the preparedness to the response phase

Data Preparedness Component	Preparedness	Response
Data Sets	Framework: Risk indices More stable data, mostly secondary data	Framework: Crisis impact and Operational Environment indicators Highly dynamic data, mostly primary data
Data Services and Tooling	Services: early warning, identifying most vulnerable and hazard prone areas Tooling: Geospatial sharing platforms, formal communication channels (email, phone), mobile data collection, cloud based file sharing, dashboard technologies for data analysis/visualization.	Services: Identifying priorities: areas and people most affected. Tooling: similar, but for communication more informal, instant messaging and more use of collaborative digital spaces.
Data Literacy	Formal capacity building trajectories within organizations can be set-up and data champions appointed. Capacity building in assessment methodologies of especially the local actors is essential for obtaining granular data. Data collaboratives organize open data events, data uploading sessions as well as peer-to-peer learning.	High rotation of staff and the emergence of actors new to the humanitarian field. This requires an agile approach to learning and capacity building in different contexts to get people up to speed quickly.
Data Governance	Fundamental issues of data governance can be addressed. Formal, legal and regulatory data sharing frameworks can be developed.	Pragmatic, ad hoc solutions might be used. Risks for data harms might get aggravated.
Networked Organizations for Data	Data Collaborative More formal and long term network. Majority of national and local stakeholders involved.	Coordinated Data Scramble. Largely informal, flexible, short-term. Majority of international humanitarian organizations in case of level 3 disaster

Table 1 shows how the meaning of each of the components evolves when going from the preparedness to the response phase by giving a few examples. While for sudden onset disasters the preparedness phase is truly disparate from the response phase, for protracted and slow onset crises this distinction becomes more arbitrary. The table indicates that for both slow and sudden onset disasters, the data preparedness components are part of a continuum. The highly dynamic, new set of data on the crisis impact and operational environment in the response phase will be most valuable when integrated with pre-disaster more static baseline data. The

Networked Organizations for Data component describes the degree of coordination among multiple organizations. The term “Networked Organization” refers to organizations with a dominant focus on emergent dynamics in collaboration and less reliance on formal hierarchical structures (Treurniet, 2014). Typically, organizations in the response and preparedness phase are networked but do not have a formal structure around collecting, analyzing or sharing data for disaster management. For several of the larger disasters in the recent years, a Coordinated Data Scramble has been setup by the international community (Campbell 2016; Verity 2016) in the response phase. These scrambles rely on *contingent coordination* (Herranz, 2008), where some opportunistic directive influence guides network behavior, but reliance on emergent behavior is still quite high. After the response phase, this group can grow into a Data Collaborative (Verhulst, 2015), a network of organizations that exchange data for disaster management with a network management regime that relies on *active coordination*. Active coordination implies a more deliberate design of the network, including its constituent partners as well as the interaction and incentive mechanisms among the partners (Herranz, 2008). Concretely this means that for example data sharing agreements have been signed or that a facilitator has been appointed. Treurniet and van Buul (Treurniet and van Buul, 2015) described how the level of information sharing depends on the dynamics and type of collaboration within these networked organizations. The information sharing can move from stovepiped data (sharing information via non-aligned reports across actors), shared data (a common frame of reference/uniform format for basic information exchange) to shared situation assessment (full common frame of reference/uniform format for all information exchange). We can use this matrix of level of information sharing versus level of collaboration to assess data preparedness on the Networked Organizations for Data dimension.

The framework is mostly useful for measuring data preparedness at output and -but to a lesser degree- outcome level. The framework cannot be used to determine the impact on sensemaking and decision making processes. Our underlying assumption is that data preparedness leads to data with higher quality in the response phase, which will debias the humanitarian decision making process. For example, the decision on where to place an operational hub for UN clusters in country can become more transparent by having a geographical assessment in place beforehand. In the remainder of this paper we zoom in on the first of the five pillars: Data Sets.

DATA SETS

Determining which data to collect and collate

As explained in the previous paragraph, a crucial part of data preparedness is pre-staging as many data sets as possible and making sure they match the information needs of responders.

Obviously, a first step is to make an inventory of the information needs for different disaster contexts. The Interagency Standing Committee (IASC) Assessment Task Force developed the Multi-Sector Initial Rapid Assessment framework (MIRA) and groups information needs into two pivotal categories, i.e. the Crisis Impact and the Operational Environment, with twelve subthemes (MIRA, 2015). Gralla et al. (Gralla et al., 2015) identified eight clusters of information requirements of field-based decision-makers along the response timeline, when going from the early stage of the response towards recovery. In the Bangladesh case that was introduced earlier (van den Homberg et al., 2016), the two main categories of MIRA were used to group the 71 information needs as identified through semi-structured interviews and focus group discussions. The cluster of Crisis Impact consisted of baseline information about the vulnerability and livelihood of communities in hazard prone areas, damage and needs assessments and information about the disaster situation such as which areas are inundated and how long the flood will last. The Operational Environment comprises coordination and capacity information as well as information on the degree of humanitarian access.

Second step is to acquire the corresponding data sets. Information requirements can be met by combinations of data sets; this is not a completely straightforward relationship. The IASC developed in 2010 *Guidelines Common Operational Datasets (CODs) in Disaster Preparedness and Response* (IASC, 2010). CODs are 'best available' datasets that ensure consistency and simplify the discovery and exchange of key data. The core CODs are Administrative Boundaries, Population Statistics, Humanitarian Profile. Country-specific CODs are a subset of the CODs that are specific to each country’s risk profile. Examples included: datasets related to demographics, geography, infrastructure (structures that could be impacted or used during relief operations such as schools, health facilities and refugee camps) and activities. Whereas these guidelines are very useful to increase the interoperability of data sets and improve data sharing, they do not aim at providing a framework to bring the different data sets together in an insightful and coherent way. A risk framework can do so. Risks materialize into crisis impact and affect the operational environment. As an example, the number of people affected (part of Crisis Impact) has a direct relation with the number of people that are exposed (before the disaster hits). Similarly, the hospitals and doctors available in the disaster area (part of Operational

Environment) can be derived from the hospitals present in the area before in combination with how they were impacted. We have therefore decided to collect and collate data on the indicators of risk before a disaster hits. Given the mandate of humanitarian organizations, our focus hereby is at risk at the community level. As a matter of fact, most of the CODs can also be categorized under one of the risk components.

In recent years, many risk indices have been developed. Most of these indices consist of a weighted or unweighted combination of indicators measuring different components of risk, resilience, or vulnerability. The World Bank report, *Unbreakable, Building the Resilience of the Poor in the Face of Natural Disasters*, lists 13 of them (Hallegatte, 2017), where each index is developed with different or sometimes to some extent overlapping user groups and application domains in mind. For example, the World Risk Index (WRI, 2016), as developed by the Institute for Environment and Human Security, United Nations University, and the University of Bonn, seems to be mostly used for development, Climate Change Adaptation (CCA) and DRR, whereas INFORM, developed by the EU Joint Research Center and UN OCHA (INFORM, 2016), is mostly targeting humanitarian aid. At the community level, NGOs, Red Cross local chapters and other mostly local organizations do regularly Participatory Capacity and Vulnerability Analysis (PVCA) assessments. These give relevant information on some of the risk components, but the analyses are often not digitized and can be both quantitative and qualitative. PVCAs tend to be an instrument to raise awareness and build capacity at community level and are not so much an instrument to get very accurate numbers on risk.

The INFORM, WRI and PVCA report at different administrative levels. Whereas PVCA reports at the community level, the WRI and INFORM report at the national level. INFORM is now working on going towards subnational level under an EU ECHO grant that OCHA and UNDP have received at the end of 2016. INFORM's main components are Vulnerability, Lack of Coping Capacity, and Hazards & Exposure, see Figure 1. These three categories are always the same for several countries. At subnational level, specific indicators per component are determined per country based on a consultative process with key stakeholders and data providers. This allows integration of data at the subnational levels, from regional, provincial, municipal up to neighborhood level. We have therefore decided to create a Community Risk Assessment based off the INFORM framework. This Community Risk Assessment distinguishes itself from INFORM through its reach in terms of administrative levels and through its application in a prioritization model. The reach in terms of administrative levels is limited by the availability of data; Figure 1 shows which data sets are available (yes) or not (no) at administrative level 2. The prioritization model predicts high priority areas based on a combination of (open) secondary risk data on hazard prone areas, weather information on the imminent hazard (such as wind speeds and rain falls) and data from past similar disasters. It uses machine learning and data mining algorithms.

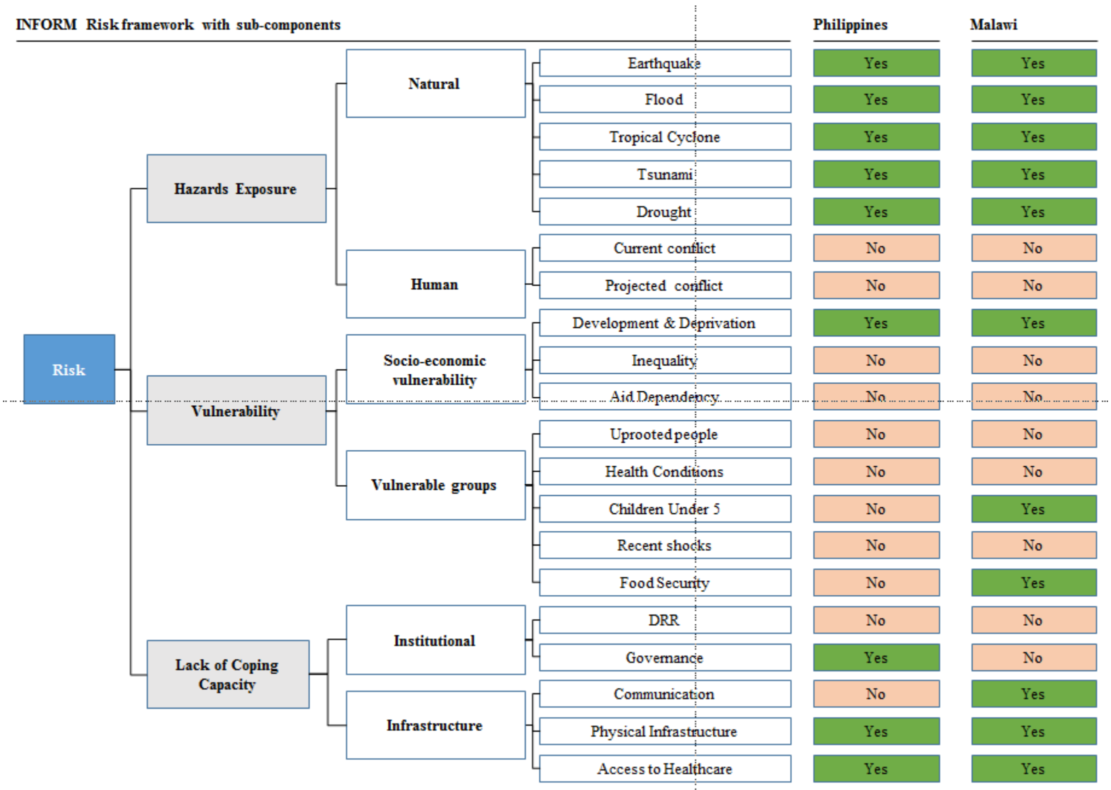

Figure 1 Indicator Framework inspired by INFORM with available data sources for Philippines and Malawi. Yes and No refer to whether the data was available at administrative level 2.

Calculating the Data Sets Preparedness Index

We develop a Data Sets Preparedness Index (DSPI) by selecting specific Data Quality dimensions. Several international organizations such as the UN, IMF and the EU have developed different data quality assessment frameworks. The Post2015 Data Test initiative (Post2015, 2014) drew from these frameworks to develop a Quality Assessment Framework specifically for the monitoring of SDGs. We extracted and transformed some of the indicators in this framework so that we would have a basic set to start with and that would be quantifiable, i.e. *Completeness, Recency* and *Accuracy & Reliability.* The formula for DSPI is as follows and will be explained in more detail below.

$$DSPI = Completeness * Recency * Accuracy \,\&\, Reliability$$

Completeness: The first component of the DSPI to consider is Completeness. If for each subcomponent (see **Erreur ! Source du renvoi introuvable.**igure 1) at least one data source is available then Completeness is defined as 100%. For each subcomponent i that has no corresponding data source in the considered country, the Completeness-score decreases, according to the weight of the subcomponent in the total risk-score. The weight of subcomponents towards a higher-level component is always divided equally. Hazards, Vulnerability and Coping Capacity each constitute one third of overall risk, Natural and Human Hazards each constitute 50% of overall Hazard, etc.

$$Completeness = \sum_i weight_i * I_i$$

where I_i is a function that results to 1 if datasource *i* is available and 0 otherwise.

Naturally, if the considered use case is different, the definition of Completeness will have to change accordingly. When collecting data on SDGs, Completeness will be defined by taking all considered SDGs and their corresponding indicators as a starting point. The principle of measuring completeness between 0% and 100% as is done here, will remain the same.

Granularity level: With the first component of the DSPI defined, it is insightful to take a step back and define the concept of granularity level. Namely, one can look at different levels to Data Preparedness. Some sources may be available on municipal level, while others are available only on provincial level. This means that when we are assessing Community Risk at a provincial level we will have more data sources and thus higher Completeness, then when we are assessing at a municipal level. Depending on the context, both perspectives can be relevant, and thus the Community Risk Assessment toolbox allows for various levels of administrative granularity. Correspondingly, the same goes for the Data Sets Preparedness Index. For each source, the deepest available granularity level is noted down, so that Completeness can be computed as a function of granularity.

$$Completeness(g) = \sum_i weight_i * I(g)_i$$

where *g* is granularity level, and $I(g)_i$ results to 1 if .data source *i* is available on granularity level *g* or deeper.

Recency: After completeness, a second very important criterion of a data source is its Recency. Recency is a combination of when the data set was last updated and how long a data set remains representative of the reality. The more recent the source is, the higher the Recency score should be. How long a data set remains representative can also be termed "retention period". Some data sets stay valuable longer than others. For example, the retention period of a source about the geospatial distribution of earthquake risk is generally deemed longer than the retention period of a research on Good Governance Index of local municipal government, which might very well have changed considerably after a new local government was elected. Scoring the retention period per source is a subjective matter. In our primary calculations, this was done to the best of our abilities. But in future applications of this framework, this can be determined more accurately for example through averaging expert opinions.

In conclusion, the more recent a data set is and the longer its retention period is, the higher the Recency score is. Mathematically, this is defined as follows for data source *i*.

$$Recency_i = Avg\left(\frac{\max(10 - years\ passed_i, 0)}{10}, \frac{\min(retention\ period\ _i, 10)}{10}\right)$$

The first part of this formula resolves to 1 if 'years passed = 0', so if the source is from 2017, and results in 0 if the source is more than 10 years old. The second part resolves to 1 if the retention period is 10 years or higher, while it resolves to 0 if the retention period is 0. The two parts are averaged out to get to the Recency score, which is again a score between 0 and 1 (i.e. between 0% and 100%). The total Recency score is again the weighted average of all individual Recency scores with the same weights as earlier used in the Completeness measure. Note that only Recency of available sources are included in this weighted average, so that incompleteness is not counted twice. It might be necessary at some stage to differentiate the cut-off timescales between indicators.

Accuracy & Reliability: Lastly, data sources can – independently from their Recency – also vary greatly in their Accuracy & Reliability (Post2015, 2014). In terms of source types, a census is generally considered more accurate than a survey for example, simply because it measures the entire population instead of a sample. Besides this, the Accuracy & Reliability can also be assessed by looking at the source and its publishing organization. This is currently a mostly qualitative assessment. In our preliminary calculations, we have done this to the best of our abilities by ranking each source on a scale from 1 to 5. Like the retention period, this part can be further formalized in future research by defining subcriteria, such as used in the Post2015 framework (Post2015, 2014) and by using expert opinions to assess these criteria. The 1-5 scale is subsequently linearly transformed to a 0-1 scale per source, after which the weighted average is taken to obtain the overall Accuracy & Reliability score. Note again that only Accuracy & Reliability of available sources is included in this weighted average, so that incompleteness is not counted twice.

Data Sets Preparedness Index: Finally, the three components of Data Preparedness are combined into the Data Sets Preparedness Index, also a score between 0 and 1. For granularity level g, this yields:

$$DSPI(g) = Completeness(g) * Recency * Accuracy\ \&\ Reliability$$

We use multiplicative aggregation here (versus taking the arithmetic mean for example), because in an ideal situation all three components should be equal to 1. If only 50% of the sources is available, and thus the Completeness score is 0.5, then we want the DSPI also to be 0.5 at most (with perfect Recency and Accuracy & Reliability). If on top of that, the Recency and Accuracy & Reliability are also not perfect, then this should further decrease the DSPI-score.

For interpretation purposes, some feeling for the magnitude of DSPI is insightful. For example, if all three components are average and have a score of 0.5 each, then the DSPI resolves to $(0.5)^3 = 0.125$. While this low score is justified, because there is a lot of progress to be made on three fronts, a score of 0.125 can be interpreted as reasonable.

Applying the Data Sets Preparedness Index

Calculation example: Table 2, 3 and 4 show results for the DSPI calculations for The Philippines and Malawi. For Malawi, 0 represents national level, 1 region level (there are three regions), 2 district level (28 districts) and 3 National Assembly constituency or Traditional Authority (TA) level (usually around 5 or 6 per district). For the Philippines 0 represents national level, 1 regions (17), 2 provinces/cities (81), 3 municipalities (1,489) and 4 barangays (42,029). If we look at the consolidated results in Table 4, we can see that the Completeness decreases when going down in administrative levels from 2 to 3 for Malawi and from 3 to 4 for The Philippines. Although indicators differ per country, the DSPI can be used to compare also across countries how far a country is in having sufficiently granular and qualitative data on their set of indicators. The index could then be used to track progress of countries in making sure they get more accurate, detailed, timely data on communities and in this way, we could link it more easily to indicators used in key global agreements such as Sendai and the SDGs or other models like INFORM.

Application: The above calculations lead to an immediate conclusion for the Community Risk Assessment toolbox. Namely, for the Philippines it is currently meaningful to develop the risk framework at administrative level 3 (municipal), but not yet at administrative level 4 (barangay) where the DSPI drops sharply. For similar reasons, in Malawi, the risk framework is for now developed only at administrative level 2 (the district level), but not yet at administrative level 3 (TA level), see Figure 2. In general, we have from these initial calculations derived an initial DSPI threshold level of 0.10. While rolling out the Community Risk Assessment toolbox to new countries, after the initial data scramble a DSPI is calculated. If this DSPI is lower than 0.10, the data preparedness is currently considered too low to meaningfully include the country.

Even though a higher DSPI is preferable to a lower one, the framework and calculation are still very insightful in case of the latter. First, the calculation is digitally stored and as new sources become available they can quickly be added, so that the progress can be monitored, and the "gap" between the current situation and the minimum threshold can be measured. Second, the bottom-up calculation of the framework also quickly gives insight into where the gaps are (see Table 3, which can even be expanded to one lower level of subcomponents as seen in Figure 1). All scores can easily be subdivided into the subcomponents of the risk framework, thereby giving focus to initiatives to bridge the gap.

Shortcomings: Some parts of this calculation, especially in the Recency and the Accuracy & Reliability part are not perfect yet and should be further developed. As such, they should in part be seen as an exemplary calculation to make the somewhat abstract concept of a DSPI more concrete, and as a starting point of further discussion.

Table 2 Calculation of the Data Sets Preparedness Index for administrative level 3 in The Philippines

Component	Completeness	Recency	Accuracy &Reliability	DSPI
Hazards & Exposure	0.50	0.70	0.60	0.21
Lack of Coping Capacity	0.58	0.39	0.63	0.14
Vulnerability	0.19	0.58	0.70	0.08
Risk	***0.42***	***0.54***	***0.63***	***0.14***

Table 3 Specification of DSPI per subcomponent for administrative level 2 in The Philippines

Component	Completeness	Recency	Accuracy &Reliability	DSPI
Hazards & Exposure	0.50	0.70	0.60	0.21
Natural	1.00	0.70	0.60	0.42
Human	-	-	-	-
Lack of Coping Capacity	0.58	0.39	0.63	0.14
Infrastructure	0.67	0.53	0.50	0.18
Institutional	0.50	0.20	0.80	0.08
Vulnerability	0.25	0.58	0.70	0.10
Socio-economic vulnerability	0.50	0.58	0.70	0.20
Vulnerable Groups	-	-	-	-
Risk	***0.44***	***0.54***	***0.63***	***0.15***

Table 4 Overview of the Data Sets Preparedness Index results for The Philippines and Malawi

	Philippines			**Malawi**	
Administrative level	**2**	**3**	**4**	**2**	**3**
Completeness	0.44	0.42	0.08	0.46	0.19
Recency	0.54	0.54	0.52	0.55	0.67
Accuracy &Reliability	0.63	0.63	0.63	0.61	0.60
DSPI (product)	***0.15***	***0.14***	***0.02***	***0.15***	***0.08***

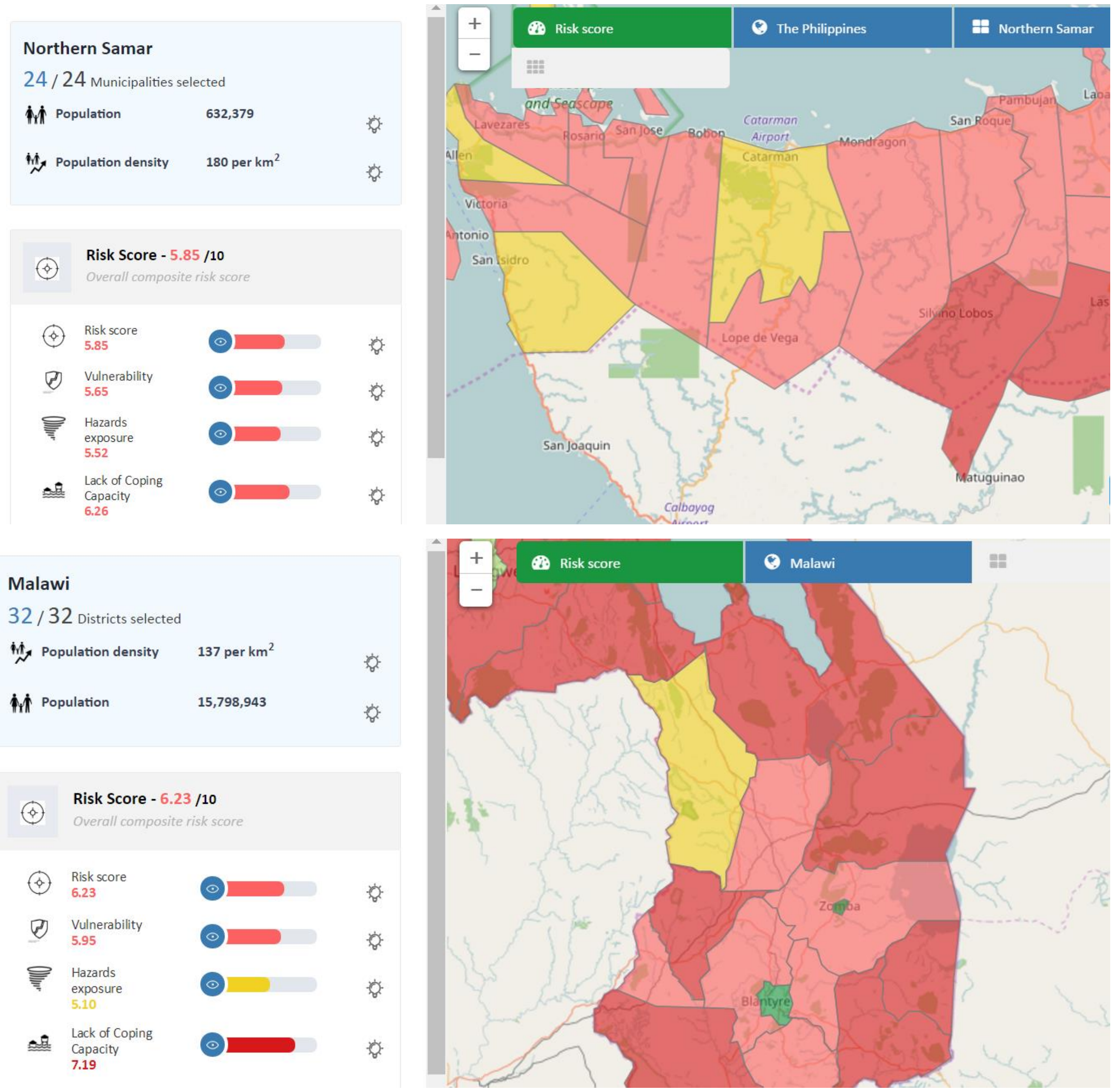

Figure 2 Community Risk Assessment dashboard for The Philippines (top) and Malawi (bottom), showing that the lower level of Data Preparedness in Malawi only allows a visualization up to the district level, whereas in The Philippines it is possible up to the municipality level.

DISCUSSIONS AND CONCLUSIONS

Data Preparedness Framework

This paper proposed a framework to assess Data Preparedness on five dimensions: Data Sets, Data Services and Tooling, Data Governance, Data Literacy, and Networked Organizations for Data. Whereas this framework is tailored to humanitarian aid, the description of the five components across the spectrum from preparedness to response also allows application in the development context. We have used this framework for the first time during a Data Preparedness Mission, where a team goes into a country, works with all the local stakeholders intensively and plugs all collected data sets into a dashboard. Possibly, data preparedness could become part of more generic preparedness missions, such as the ones that the United Nations Disaster Assessment and Coordination (UNDAC) teams execute (UNDAC, 2013; UNDAC 2017). Quantifying the different data preparedness components makes it easier to characterize the current state, to define the desired state and to focus data preparedness activities. We showed for one dimension, i.e. Data Sets, how it can be quantified.

Data Sets

First, we explained how one can use a risk indicator framework to determine which data to collect and collate. We developed a dashboard to bring all the data together in an accessible and insightful way. Red Cross National Societies can play a pivotal role around the communication and engagement with communities in relation to the dashboard. However, Red Cross National Societies are not considered apt to be the host of the data and to fulfill an administrator type of role for the such a data platform. This should be the local/national government. Subsequently, and this is the main contribution of this paper, we designed a prototype DSPI with as objectives to give direction to data collection, to track progress, to create more efficiency in the data collection and sharing and to define when it becomes possible to use the Severity and Priority tooling. The DSPI can still be extended to include other data quality dimensions. For example, Data Accessibility could be added as a factor to the equation. We saw in Malawi that especially at the lower administrative levels data could be available but not at all easily accessible. The data was not yet uploaded to geospatial sharing platforms, but kept on individual computers. Reasons for this could vary from low priority, lack of time and infrastructure (poor internet for example) to do so. The Post2015 Data Set initiative (Post2015, 2014) mentions several other subcomponents for Accessibility and clarity.

Limitations and future applications

Our research has until now solicited only limited user feedback on the DSPI. We will extend the number of case studies and application areas. 510 has advanced discussions going on with INFORM to see how the DSPI in combination with the Community Risk Assessment and Prioritization toolbox can be linked to the roll-out of the INFORM Sub-National Models. It will also be important to see how the DSPI can be linked to INFORM's reliability score. This score is a measure of reliability for each country. It is represented on a scale from 0-10 and includes missing data, out of date data, and conflict status. Countries with lower Reliability Index scores have risk scores that are based on more reliable data (INFORM, 2016).

DSPI gives direction to data collection but does not solve the issue of a lack of data at grassroots level. Alkire and Samman (Alkire and Samman, 2014) have developed ten technical criteria to evaluate different data collection methods at the household level specifically for monitoring on the SDGs. The DSPI data quality components are part of these ten criteria and can be used to assess different data collection methods. For example, social media data might score very high on Recency but the Completeness might be low if only a small percentage of the people in the area of interest use social media. We have also started mapping data providers on the indicators, since this will give an intuitive and visual representation of the data ecosystem. Such an institutional map can uncover if there are too many data providers for a certain indicator or rather too little and can subsequently be used in a data collaborative to align data collection among different organizations.

In conclusion, we developed a DSPI that can be used to monitor progress of countries on making sure they get more accurate, detailed, timely data on communities paving the way for more transparent humanitarian decision making and more inclusive reporting on global development agreements.

ACKNOWLEDGMENTS

We would like to thank Andrew Thow from OCHA and Luca Vernaccini from the EU Joint Research Center for providing us their valuable insights on the subnational INFORM index. We are grateful as well for the humanitarian decision-makers in the Philippines and Malawi who took the time for a semi-structured interview with us and more specifically the Philippines and Malawi Red Cross for hosting and supporting our mission. The Prevent a Disaster crowdfunding campaign of the Prinses Margriet Fund supported this work financially.

REFERENCES

Alkire, S. and Samman, E. (2014) 'Mobilising the Household Data Required to Progress toward the SDGs'. OPHI Working Paper 72, Oxford University.

Campbell, Helen, The Coordinated Data Scramble for Humanitarian Emergency Response, July 2016, https://www.youtube.com/watch?v=HmsF743VRX0, Accessed 22 Mar. 2017

Comes, Tina. "Cognitive biases in humanitarian sensemaking and decision-making lessons from field research." *Cognitive Methods in Situation Awareness and Decision Support (CogSIMA), 2016 IEEE International Multi-Disciplinary Conference on*. IEEE, 2016.

Gralla, Erica, Jarrod Goentzel, and B. V. de Walle. "Understanding the information needs of field-based decision-makers in humanitarian response to sudden onset disasters." *Proceedings of the 12th International Conference on Information Systems for Crisis Response and Management (ISCRAM).* 2015.

Hallegatte, Stephane, et al. *Unbreakable: Building the Resilience of the Poor in the Face of Natural Disasters*. World Bank Publications, 2016.

Herranz, Joaquin. "The multisectoral trilemma of network management." *Journal of Public Administration Research and Theory* 18.1 (2008): 1-31. History of the coordinated data scramble (DH Network, June 2016), https://app.box.com/s/k9kk1oidqngxv301ibna3p2cm97qff68

IASC Guidelines Common Operational Datasets (CODs) in Disaster Preparedness and Response, November 2010, https://www.humanitarianresponse.info/en/applications/tools/category/operational-datasets, Accessed 22 Mar. 2017

INFORM index for risk Management (2016). Using INFORM to assess crisis and disaster risk within a country or region, Guidance Note, November 2015, http://www.inform-index.org/Portals/0/InfoRM/2016/INFORM%20Subnational%20Guidance%20Note.pdf, Accessed 22 Mar. 2017

INFORM index for Risk Management (2017), INFORM results 2017, http://www.inform-index.org/Portals/0/InfoRM/INFORM%20Global%20Results%20Report%202017%20FINAL%20WEB.pdf, Accessed 22 Mar. 2017.

Multi-sector initial rapid assessment Guidance, July 2015, https://www.humanitarianresponse.info/en/programme-cycle/space/document/multi-sector-initial-rapid-assessment-guidance-revision-july-2015, Accessed 22 Mar. 2017.

OCHA ROAP, 2009, IM preparedness, matrix of indicators. http://www.pacificdisaster.net/pdnadmin/data/original/OCHAROAP_PreparednessIndicatormatrix_V2.pdf, Accessed 22 Mar. 2017

Post-2015 Data Test Unpacking the Data Revolution at the Country Level, Southern Voice, July 2014, http://www.post2015datatest.com/wp-content/uploads/2014/07/Post2015-Data-Test-Methodology-Guide.pdf, Accessed 22 Mar. 2017.

Raymond, N., Al Achkar, Z., Data preparedness: connecting data, decision-making and humanitarian response, Signal Standards and Ethics Series 01, 2017.

Treurniet, W. "Shaping comprehensive emergency response networks." *Network Topology in Command and Control: Organization, Operation, and Evolution: Organization, Operation, and Evolution* 26 (2014).

Treurniet, Willem, and Kim van Buul. "Four archetypal networked organisations." *Palen LA Comes T. Buscher M. Hughes AL, ISCRAM 2015 Conference Proceedings-12th International Conference on Information Systems for Crisis Response and Management, 24-27 May 2015, Kristiansand, Norway*. Information Systems for Crisis Response and Management, ISCRAM, 2015.

UNDAC Disaster Response Preparedness Missions Part I Strategy and Framework, http://www.pacificdisaster.net/pdnadmin/data/original/UNDAC_Mission_PartI_strategy_framwrk.pdf, Accessed 22 Mar. 2017.

UNDAC, United Nations Disaster Assessment and Coordination Field Handbook, 6th edition, 2013.

United Nations University Institute for Environment and Human Security (2016), Overview of updated indicators for the WorldRiskIndex 2016, http://weltrisikobericht.de/wp-content/uploads/2016/09/Indicator-update-for-the-WorldRiskIndex_2016.pdf Accessed 22 Mar. 2017.

Van den Homberg, 2016, “Data for disaster management – mind the gap”, essay in the Responding to Natural Disasters in the MENA Region and Asia: Rising to the Challenge? series of the Middle East Institute, http://www.mei.edu/content/map/data-disaster-management-mind-gap

Van den Homberg, M., Monné, R., Spruit, M. 2016, “Bridging the Information Gap: Mapping Data Sets on Information Needs in the Preparedness and Response Phase”, in Tech4Dev UNESCO online conference proceedings.

van Lint, Sandra "Sense-making of the Netherlands Red Cross Priority Index model – case typhoon Haiyan, Philippines."

Verhulst, Stefaan G., and David Sangokoya. “Data Collaboratives: Exchanging Data to Improve People's Lives.” Medium, GovLab, 22 Apr. 2015, medium.com/@sverhulst/data-collaboratives-exchanging-data-to-improve-people-s-lives-d0fcfc1bdd9a. Accessed 22 Mar. 2017.

Verity, Andrej. “I Scramble. You Scramble. We All Scramble for Data!” Verity•Think, 15 June 2016, blog.veritythink.com/post/145957370879/i-scramble-you-scramble-we-all-scramble-for. Accessed 22 Mar.

2017.