The reciprocity of data integration in disaster risk analysis

David Paulus  
PhD Researcher  
Policy Analysis Section  
TU Delft  
d.paulus@tudelft.nl

Kenny Meesters  
PhD Researcher  
Policy Analysis Section  
TU Delft  
K.J.M.G.Meesters@tudelft.nl

Gerdien de Vries  
Assistant Professor  
Organisation & Governance Section  
TU Delft  
G.deVries-2@tudelft.nl

Bartel Van de Walle  
Professor  
Head of Multi-Actor Systems Department  
TU Delft  
B.A.vandeWalle@tudelft.nl

ABSTRACT

Humanitarian organizations are increasingly challenged by the amount of data available to drive their decisions. Useful data can come from many sources, exists in different formats, and merging it into a basis for analysis and planning often exceeds organizations’ capacities and resources. At the same time, affected communities’ participation in decision making processes is often hindered by a lack of information and data literacy capacities within the communities. We describe a participatory disaster risk analysis project in the central Philippines where the community and a humanitarian NGO worked towards a joint understanding of disaster risks and coping capacities through data integration and IT-supported analysis. We present findings from workshops, focus group discussions and semi-structured interviews, showing the reciprocal effects of the collaborative work. While the community valued the systematically gathered and structured evidence that supported their own risk perceptions and advocacy efforts, the humanitarian NGO revisited established work practices for data collection for analysis and planning.

Keywords  
Reciprocity, Resilience, Disaster risk analysis, Community engagement, Organizational effectiveness, Data integration

INTRODUCTION

To prepare for the events of natural hazards, stakeholders of climate change adaptation (CCA) and disaster risk reduction (DRR) programs try to understand capacities and vulnerabilities of potentially affected communities (Mercer et al., 2010). This requires gathering and understanding data from various sources (Kawasaki et al., 2017). Useful data can come from many sources, exists in different formats, and merging it into a basis for analysis and planning often exceeds organizations’ capacities and resources. Household surveys are frequently conducted by DRR program stakeholders and are a well-established tool to capture field realities. They provide information on demographics, education levels, income types and salary levels (Morin et al., 2016). And they can capture people’s perceptions of risks and capacities in their neighborhoods (Sullivan-Wiley & Short Gianotti, 2017).

Official data from government agencies, e.g. censuses, are increasingly published openly in light of open government initiatives (Janssen et al., 2012). Recently, aid organizations increasingly publish and share data on their activities (IATI, 2017; UNOCHA, 2018), which in turn can be useful information for other organizations (Crowley et al., 2011). Examples are data created through the International Aid Transparency Initiative (IATI)\(^1\) and the Humanitarian Data Exchange (HDX)\(^2\). In combination with organizations’ own data creation efforts, this leads to an increasingly difficult task of data integration and sense making.

\(^1\) https://iatistandard.org  
\(^2\) https://data.humdata.org
Problem statement

While these initiatives are deployed to increase the data availability, analysis and use in decision making processes, these initiatives are often centered around humanitarian organizations, their objectives and requirements. Notwithstanding the objective of humanitarian organization to alleviate human suffering and strengthening communities by increasing their resilience to disruptive events (Weiss, 1999), often the main focus of data collection and processing is to ensure effective humanitarian operations, informed decision making or provide accountability on an organizational level (Comfort et al. 2004; Gralla et al. 2013). Subsequently, with the rise of technologies and data in the humanitarian field, various platforms, tools and systems have been provided to humanitarian organizations to support their needs assessments, operations and impact evaluations (Crowley et al. 2011).

Consultations of affected communities are the fundament of most CCA and DRR programs. Data are gathered from assessments and then analyzed and interpreted by organizations to plan their program implementation. This often moves communities into passive roles as data providers who are excluded from analysis, interpretation and decision making processes (Streefkerk et al., 2014). This stands in contrast with arguments from the social resilience literature (Comes et al., 2017) that emphasizes the importance of building resilience by increasing community capacities, resources and ownership (Maskrey, 1989).

With the increased importance of data and IT-tools in humanitarian operations, such capacities become critical elements for the empowerment of communities to self-organize preparations and responses to disasters (Baharmand et al., 2016; Kapucu, 2008; Kendra et al., 2007). However, communities - in contrast to humanitarian organizations - often have less access to specific technologies and resources to implement and use supportive systems. Their capabilities to employ data-driven assessments remain limited (Piccolo et al., 2018; Streefkerk et al., 2014).

Research objective

We examine how, through the use of modern technologies, this discrepancy can be resolved. Specifically, through (1) building an understanding of information needs and offers of communities and organizations, (2) streamlining data gathering and analysis, and (3) a redistribution of processes around data that builds on and strengthens local capacities.

In the remainder of this paper we first discuss the current state of research around data-driven humanitarian operations which leads to the uncovering of the above briefly outlined research gap: how can communities and humanitarian organizations jointly build an understanding of risks, vulnerabilities and capacities to support community-driven resilience and mitigation efforts? We then describe our research approach, using a prototypical process implemented in a CCA and DRR program in the Philippines, of how such joint efforts could be shaped in practice. This is followed by the description of results, both solely technical as well as reciprocal, meaning the mutual influences between system, community and organization. We conclude the paper with a discussion of the results and prospects for future research.

PREVIOUS RESEARCH

In previous research Van Den Homberg et al. (2014), Gralla et al. (2015) and Comes et al. (2017) investigated information needs of humanitarian organizations and affected communities. We argue that information needs identified by them, also apply to CCA and DRR program stakeholders in our case study. In the field of humanitarian logistics Link et al. (2015) stressed the importance of data integration during disaster preparedness activities and structured humanitarian logistics information into three categories which determine when certain information is needed: during the rapid response, ongoing response or preparedness phase. Horita et al. (2014) developed the humanitarian logistics infrastructure and resource model to integrate volunteered geographic information into humanitarian logistics processes. The usefulness of the integration of data initiatives like IATI and HDX for humanitarian organizations and affected communities was suggested by Paulus et al. (2018). They outlined an information system design approach to address questions that arise during different phases of humanitarian activity. Muhren et al. (2010) and Van de Walle et al. (2016) stressed the role of information on situational awareness, sense making and decision making and formulated fundamental design principles for data systems that support individuals and groups in these activities. Information systems, according to DeLone and McLean (2003) are comprised of a multitude of factors, and not only describe technical artifacts. They also include the human and organizational factors, that should not only be considered as part of the design and performance of
an information system but are in fact an integral part that determine the effectiveness of such a system to achieve organizational goals. Formalized community engagement principles like the empirically-based EnRiCH community resilience framework for high-risk populations have emphasized empowerment and collaboration as crucial drivers of adaptive capacity (O’Sullivan et al., 2014). While the importance of community engagement in disaster preparedness efforts is acknowledged and success factors have been defined (Mays et al. 2014), surprisingly few studies analyzed the reciprocal effects of disaster information systems and community engagement efforts (Ahmed et al. 2012).

From these previous findings we draw that the reciprocal effects of data-driven and IT-supported methods for community-centered disaster risk reduction projects have not been investigated substantially. In this paper we will therefore assess these effects by employing the development, testing and evaluation of a software prototype for data integration and analysis into a DRR program.

SELECTED CASE

The research was conducted within a CCA and DRR program in Jagobiao, an urban poor community of approximately 14,000 people near Cebu City, Central Philippines. The implementing organization had established close ties to the Jagobiao community, local government authorities, local faith-based groups and the local disaster risk management office over the past years. The community, administratively divided into several districts, held frequent self-organized gatherings during which risks and capacities were collected and potential mitigation and supportive measures were discussed. One outcome of these gatherings were hand-drawn maps per district and per hazard with color-coded households according to the perceived risk level for each household. The maps further contained community capacities, for example evacuation centers, hospitals and water wells. Another community-driven activity was the conducting of household surveys of the community’s youth group. The surveys captured per household demographics and characteristics, for example building material, proximity to shoreline and household income. These community-driven activities to capture local knowledge lead to a volume of data that the program stakeholders were unable to process, analyze and interpret effectively.

RESEARCH APPROACH

As per the case description, data integration and analysis became the major concern of the program stakeholders and hindered the development of a common understanding of the community’s risks and capacities. The process towards a more streamlined approach to data integration and analysis through the incremental development of a supportive software prototype is described in this section. We start with describing the process through the lens of Action Design Research (ADR) by Sein et al. (2011). Table 1 summarizes the process.

<table>
<thead>
<tr>
<th>Stages and Principles</th>
<th>Artifact</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stage 1: Problem Formulation</strong></td>
<td></td>
</tr>
<tr>
<td>Principle 1: Practice-Inspired Research</td>
<td>Research project was initiated to identify concrete information challenges and possible information system solutions in the selected DRR project. <strong>Recognition:</strong> Existing work practices, used tools and available data did not satisfy organizational objectives to understand risks, vulnerabilities and capacities of the local community.</td>
</tr>
<tr>
<td><strong>Stage 2: Building, Intervention, Evaluation</strong></td>
<td></td>
</tr>
<tr>
<td>Principle 3: Reciprocal Shaping</td>
<td>Indicator definitions and survey data was either lacking or not systematically created. Continuously during the research project, the stakeholders collaborated to design processes and requirements to address both issues. <strong>Alpha Version:</strong> The prototype should streamline the data integration and analysis process.</td>
</tr>
<tr>
<td>Principle 4: Mutually Influential Roles</td>
<td>The group of stakeholders consisted of researchers, NGO staff, local government agency officials as well as representatives of the local community and a local faith-based organization. Thus, scientific, practical and policy perspectives were taken into account. <strong>Beta Version:</strong> The analysis offered by the prototype should be based on systematically created indicators and data.</td>
</tr>
<tr>
<td>Principle 5: Authentic and Concurrent Evaluation</td>
<td>First evaluations were conducted by the researchers and NGO staff and then with the other stakeholders. Also further international offices of the NGO tested the prototype during development. <strong>Emerging Version and Realization:</strong> Additional requirements for the</td>
</tr>
<tr>
<td><strong>Stage 3: Reflection and Learning</strong></td>
<td></td>
</tr>
<tr>
<td>Principle 6:</td>
<td>The NGO recognized potentials for improving work practices regarding systematic data collection and analysis. The local</td>
</tr>
</tbody>
</table>
Guided Emergence: The government agency reflected on use cases for a broader rollout of the system within its mandate area. The community valued the novel approach as an additional source of evidence for their own advocacy campaigns. System design requirements for the different stakeholders emerged.

Stage 4: Formalization of Learning
Principle 7: Generalized Outcomes
Concrete requirements for a first prototypical system version and for a more advanced later version were captured. Regarding the system as an integral future tool for analysis and planning of DRR programs.

Requirement analysis
Specific requirements for the prototype were continuously generated and adapted from discussions, workshops, interviews and observations with the DRR program stakeholders during field visits to the case area and remotely. General requirements were drawn from information system design literature and previous studies on information system development for DRR and crisis response (Turoff et al., 2004). According to the technology acceptance model, perceived usefulness and perceived ease-of-use are the major influential factors that determine if new information systems will be used (Davis, 1989; Davis et al., 1989). In the case described here, involved stakeholders IT literacy and experience varied strongly: from advanced experience within the government agency to mid-skilled humanitarian NGO staff to local community people with rather low IT literacy. The development process of the prototype needed to take this capacity diversity into account to facilitate adoption (Maiers et al., 2005). It became evident that a successful approach would need to consider processes and tools already implemented and used by the program stakeholders.

Development and Feedback Process
The prototype was developed with three core components: 1) a dashboard acting as the tool for analysis, 2) the data processing layer and 3) an API layer to retrieve external data and allow integration with other applications. A feedback process was set up that included stakeholders of the CCA and DRR program. Table 2 lists the participants and their involvement in the feedback process. Feedback data collection was conducted through semi-structured interviews, focus group discussions, workshops, field observations as well as e-mail and Skype conversations.

As a foundation for the development of the prototype, a dashboard previously developed by UNOCHA was used. It was open to adaptation, fulfilled sufficient aspects for dashboard design (Janes et al., 2013) and already provided some of the main requirements needed for the prototype. This included the interrelation of charts, a map and processing functions for household survey data on demographics, capacities and vulnerabilities of communities. During the parallel development and evaluation process we added the selection of hazards, the simulation of hazard intensities, substantially more charts on household capacities as well as dynamic map features and more interrelations between the charts and the map.

Within the data processing layer, the prototype does not store any data. Rather, the user needs to be in the possession of the data at all time. This eliminates the possibilities of malicious system infiltration and leakages of sensitive data. JavaScript code is sent by the server to the user and executed on the user’s machine. Thus, when using the dashboard, both data and application are running on the user side. The decision on this form of data handling was made due to concerns raised by the community. Some members regarded the data being used by the prototype as too sensitive and personal and expected this additional security feature to prevent malicious actors from accessing it unauthorized. The decision is further supported by recent arguments within the debate on data management in humanitarian operations. Especially in political sensitive contexts, household and demographic data can be highly valuable for the different conflicting parties (Greenwood et al. 2017).

Three datasets are needed to run the prototype:
- Survey data (csv-file): Can be created in Excel or from a KoboToolbox\(^3\) export, or any other survey platform. The internal data structure and terminology of the prototype needs to be followed.
- 3W data (csv-file): Can be created in Excel. It comprises local knowledge from DRR program stakeholders on relevant organizations, groups and networks. To complement the 3W data, information from IATI and HDX are fetched.
- Geographic features (kml-file): Can be created in Google Earth. Includes local knowledge from DRR

\(^3\) https://www.kobotoolbox.org/
program stakeholders on important geographic features in the case area, e.g. evacuation centers, wells and water flows.

The choices for the data creation tools, Microsoft Excel, KoboToolbox and Google Earth, were made during stakeholder discussions and successive testing of the prototype during development. The three tools were well-known by most of the stakeholders. No additional training was needed.

Table 2. Methods of feedback collection and relations to stakeholder affiliations.

<table>
<thead>
<tr>
<th>Group or organization</th>
<th>Interviewee affiliation (n)</th>
<th>Feedback collected via</th>
</tr>
</thead>
<tbody>
<tr>
<td>International humanitarian NGO</td>
<td>Head of mission (1)</td>
<td>Interviews, focus group discussions, e-mails, Skype</td>
</tr>
<tr>
<td></td>
<td>Data analyst (1)</td>
<td></td>
</tr>
<tr>
<td>Local disaster risk management office</td>
<td>Data analyst (1)</td>
<td>Interviews, focus group discussions</td>
</tr>
<tr>
<td>Local faith-based organization</td>
<td>Data analyst (2)</td>
<td>Interviews, focus group discussions</td>
</tr>
<tr>
<td>Local community</td>
<td>Priest (1)</td>
<td>Focus group discussions</td>
</tr>
<tr>
<td></td>
<td>District representatives (5)</td>
<td></td>
</tr>
</tbody>
</table>

RESULTS

We present the results in the following first from a purely technical perspective, followed by the observed reciprocal results of the effects between system prototype development, community and organization.

Technical results

The prototype’s visual user interface is a single-page dashboard made up of a map, charts, tabularized information and buttons as control elements. The three main components are depicted in figures 1-3. Figure 1 shows how the prototype visualizes community demographics. The bars and slices in the charts are clickable to allow the selection of a certain kind of data only. For example, clicking on a single district (i.e. a slice) in the Sitio pie chart, manipulates the gender pie chart in a way that it only shows the gender proportion within the selected district. Selecting multiple districts leads to the gender chart showing the combined gender proportions within all selected districts. All other charts are affected at the same time: the age chart shows the age structure and the affiliation chart shows the job title distribution within the selected districts. This works in all ways, e.g. starting from selecting affiliations or age groups instead of starting from districts. Data categories can be combined by selecting various slices from various charts. Users thereby can answer questions like: how many females above 60 years of age live in Sitio Santa Cruz?

Figure 2 shows the mapping of households at risk of being affected by different hazards within the community. Implemented were the main hazards identified by the stakeholders: flood, fire, typhoon and drought. The colored buttons below the map allow for simulating the intensity of those hazards, which in turn has an effect on the number of households potentially affected and highlighted on the map. Via the map legend, features like evacuation centers, wells and water flows can be added to the map. Clicking on a household dot shows its key information including age structure and income level. Making selections via the demographics charts, for example selecting only those households with small children, renders the map new, only showing households that match the selection criteria.

Figure 3 shows two example charts that visualize capacities and vulnerabilities of households in the community. These had been identified during the workshops and discussions with the stakeholders. Again, these charts are intertwined with the demographics charts and the map. Allowing users to answer questions like: what districts have the most small children and elderly people living in light material housing in low elevation areas?
Reciprocal results

Within the NGO, the work on the systematic creation of indicators and the collection of data lead to a reconsideration of established work practices. Figure 4 depicts how the organization tried to assess risk levels prior to the work on the prototype. The amount of data collected and the complexity of relations between capacities, risks and vulnerabilities blurred the organization’s view on what the key issues were in the community and what caused them. As shown in figure 4, the previous mapping approach included a number of data layers but their integration was lacking. It didn’t allow answering questions like: how many elderly people and children under five live in light-material buildings in typhoon prone areas? This was only possible with the new prototype where combining variables for integrated analysis became possible. The community members valued the prototype and the development approach leading to it, as helpful in their objective understanding of the risks they face as well as a validation of their subjective risk perception. They acknowledged the prototype as a supportive vehicle that could facilitate their own community-driven initiatives to raise awareness of their case at the local government level. The local government agency mentioned that a final version of the prototype could be rolled out to the whole province of Cebu, covering an area of approximately 3 million people. The agency stressed the importance of a flexible system open to future adaptations and implemented through technologies already employed by the agency.
DISCUSSION

Throughout the participatory approach towards designing the system prototype, we uncovered critical elements for the continued engagement of the community in the data management processes. Taking these design considerations into account enables communities to not only gather data themselves but also analyze it and support their own, internal, sense making and decision making processes. The inclusion of the community in the design, development and testing of the system does not only improve adoption but also supports the building of key capacities and data literacy skills. The end goal of data preparedness therefore is not merely the creation of data-sets but rather a collaborative learning process in which the community examines their information needs, capacities and tools needed (Norris et al. 2008). Data preparedness activities therefore should not only focus on making communities visible on maps, but build necessary capacities that lead to community-driven, collaborative programs and an active exchange of data and knowledge.

These capacities are critical for community resilience as they enable them to actively participate in the decision-making processes (Nur et al. 2015; Onencan et al. 2018). They can build a stronger evidence-base by continuously generating and sharing information that places these empowered communities firmly in the local, regional, national and even international stakeholder landscape (Kapucu 2008; Mayunga 2007).

All interviewees stressed that it is of major importance to understand and comply with data rights of communities and affected persons. This does not only include data privacy and security. But also, the rights of data subjects to get access to information about them when requested, to have information about them changed when wrong, to be informed about how information about them is stored, processed and used and to what purpose and result.

The joint disaster risk analysis and prototype development have created a paradigm shift in the role of the community in the overall information management process. Where initially the role of the community was primarily the one as data provider, the new design has led to a more collaborative information gathering, analysis and dissemination approach in which local actors play a more empowered role. Figure 5 depicts this shift, with the former data relationships between DRR program stakeholders on the left side and on the right side the new, joint cooperation to build a common understanding of risks and capacities. The dotted lines represent the newly formed communication and collaboration links between the community and the NGO, and all other previously established forms of collaboration and exchange further exist.

CONCLUSION

The most striking observation was that the community and the humanitarian organizations recognized significant reciprocal effects between the development of the prototype and their work practices. They revisited established data collection and analysis practices while seeing the prototype advance, testing it and giving feedback on it. The stakeholders acknowledged that the existing data gathering, and creation processes were not systematic enough to lead to robust risk and capacity indicators and variables.

We used prototypical software development and a diverse feedback collection process to study what information needs communities and humanitarian organizations have and how streamlined data integration and analysis can
help in addressing these needs. We found evidence for the usefulness of data integration for various stakeholders of DRR programs: humanitarian organizations, local communities, government agencies and faith-based organizations.

The increased availability of data to support humanitarian operations, spurred on by the profusion of information technologies that facilitate data collection, analysis and sharing, have led to a data revolution in the humanitarian field. Increasingly, organizations, donors and policy makers place a strong emphasis on the use of data for effective operations, evidence-based decision making and accountability. At the same time the potential of data in the humanitarian field has also led to an increased attention for communities to become prepared and ‘data-ready’. As a result, more and more datasets become available, strengthening the potential of data to ensure effective aid delivery to those in need.

Throughout our participatory approach towards designing a novel disaster risk analysis prototype, we uncovered critical elements for the continued engagement of the community in the data management processes. Taking these design considerations into account enables communities to not only gather data themselves but also analyze it and support their own, internal, sense making and decision making processes. More important than the design of the system itself however, is the inclusion of the community in the design, development and testing of the system as this does not only improve the probability of adoption but also supports the building of key capacities for local disaster risk mitigation, preparedness, response and recovery efforts.

LIMITATIONS AND FUTURE RESEARCH

The work presented in this paper is limited to a single case study, and in our future work we aim to expand our work to other regions and settings to further validate the findings presented in this study. In addition to the further validation of the work, a key aspect is the continued exploration of the participatory design of data management processes and information systems. Specifically, various approaches towards capacity building and training, also raised during interviews and focus group discussions, need to be further investigated. Another interesting further endeavour is to study how decisions from humanitarian organisations are affected by how the evidence that results from the streamlined data integration and analysis approach is presented and communicated. That is, recent empirical research shows that decision making in complex situations can be influenced by how information is being framed (e.g. de Vries, 2017).

ACKNOWLEDGMENTS

This research was funded by HumanityX, an initiative of the Leiden University’s Centre for Innovation.
REFERENCES


