

A Framework for Strengthening Data Ecosystems to Serve Humanitarian Purposes

Elise Haak
Delft University of Technology
The Netherlands
haakelise@gmail.com

Jolien Ubacht
Delft University of Technology
The Netherlands
j.ubacht@tudelft.nl

Marc Van den Homberg
The Netherlands Red Cross
The Netherlands
marcjchr@gmail.com

Scott Cunningham
Delft University of Technology
The Netherlands
s.cunningham@tudelft.nl

Bartel Van den Walle
Delft University of Technology
The Netherlands
b.a.vandewalle-1@tudelft.nl

ABSTRACT

The incidence of natural disasters worldwide is increasing. As a result, a growing number of people is in need of humanitarian support, for which limited resources are available. This requires an effective and efficient prioritization of the most vulnerable people in the preparedness phase, and the most affected people in the response phase of humanitarian action. Data-driven models have the potential to support this prioritization process. However, the applications of these models in a country requires a certain level of data preparedness. To achieve this level of data preparedness on a large scale we need to know how to facilitate, stimulate and coordinate data-sharing between humanitarian actors. We use a data ecosystem perspective to develop success criteria for establishing a “humanitarian data ecosystem”. We first present the development of a general framework with data ecosystem governance success criteria based on a systematic literature review. Subsequently, the applicability of this framework in the humanitarian sector is assessed through a case study on the “Community Risk Assessment and Prioritization toolbox” developed by the Netherlands Red Cross. The empirical evidence led to the adaption the framework to the specific criteria that need to be addressed when aiming to establish a successful humanitarian data ecosystem.

CSS CONCEPTS

• **Information systems** → **Information systems applications; data warehouses; data analytics;** • **Applied computing** → **Operations research; forecasting;** • **Social and professional topics** → **Professional topics; management of computing and information systems;**

KEYWORDS

humanitarian sector, data ecosystem, data preparedness, governance, framework

ACM Reference Format

E. Haak, J. Ubacht, M. van den Homberg, B. van der Walle, S. Cunningham. 2018. *A Framework for Strengthening Data Ecosystems to Serve Humanitarian Purposes*. In Proceedings of 19th Annual International Conference on Digital Government Research (dg.o’18), Anneke Zuiderwijk and Charles C. Hinnant (Eds.). ACM, New York, NY, USA, 9 pages.

1 DATA FOR THE HUMANITARIAN SECTOR

The humanitarian sector worldwide is under pressure: there is an increasing demand for humanitarian aid and emergency response services, as both the number and length of humanitarian crises grow [1, 19, 33]. One of the causes of this increase is the rising number of disasters triggered by natural hazards, due to climate change, increased urbanization and population growth. This results in a fast-growing number of people in need [12]. Due to scarcity of budget and resources for disaster response, humanitarian actors cannot reach all affected persons. The prioritization of vulnerable and affected people is therefore at the core of all humanitarian interventions, and equally important in both the preparedness (pre-disaster) and the response phase (post-disaster) of humanitarian action [30]. Obtaining reliable and objective information for decision-making related to this prioritization process is considered very challenging [29]. Contributing factors are for example aid workers operating under stress and time pressure and being presented with uncertain information, which could lead to decision-making biases [7], and the complex organizational humanitarian setting in which information is scattered across many different organizations and sectors [20].

The rapidly changing information environment provides ways to deal with these challenges. The fast increase in the availability of data leads to a shift towards more evidence-based humanitarian decision-making, offering humanitarian actors ways to become more effective and efficient. Paper-based processes are increasingly digitized and trends around crowdsourcing, social media, collaborative digital spaces and mobile services lead to a significant increase in data availability. This also affects the humanitarian sector at large. Humanitarian donors for example, push the organizations they sponsor to open and share their data. In addition, major global agreements that recently came into effect, such as the Paris Climate Agreement, the Sustainable Development Goals and the Sendai

Framework for Disaster Risk Reduction, stimulate countries to collect data on different indicators that are potentially valuable for humanitarian organizations [6, 30]. As Sabou [27, par. 3] describes: “[H]umanitarian organizations now seek better ways to collect, store and use digital data to improve their collective responses to large scale crises” The United Nations term this trend “*the data revolution*” [13].

As a result, a growing number of studies is dedicated to mathematical models and data-analytic techniques that support the humanitarian decision-making processes in the preparation for or response to the consequences of a disaster [21]. Data-driven priority indices have grown popular for effectively and efficiently identifying communities most vulnerable for or affected by disasters and numerous data and machine learning tools have been developed to aid prioritization [5].

However, humanitarian actors miss timely, reliable and sufficient granular data and often lack the skills and tools to analyze the data needed for a transparent and structured prioritization [30]. Data is collected, stored and used by an uncoordinated, diverse group of actors with varying degrees of expertise. This lack of alignment causes a range of inefficiencies [6]. In addition, the collection, collation and analysis of data often only starts once a disaster hits, and to be able to apply these prioritization models, a certain level of ‘data preparedness’ is required [30]. Data preparedness is defined by Raymond & Al Achkar [25, p. 3] as “*the ability of organizations to be ready to responsibly and effectively deploy data tools before a disaster strikes*”. Van den Homberg, Visser & Van der Veen [30] go beyond only the aspect of being ready to deploy tools and include the pre-staging of data. They 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- like data on population density and vulnerability, (types of) roads, house construction material, sanitation and existing hospitals) and to develop capacities to collect data on affected communities and areas once a disaster hits to ensure a timely, efficient, and effective response. Their framework for Data Preparedness consists of five components, being ‘Data Sets’, ‘Data Services and Tooling’, ‘Data Literacy’, ‘Data Governance’ and ‘Networked Organizations for Data’. Consequently, there is a need to understand how to facilitate, stimulate and coordinate data-sharing between humanitarian actors to increase the level of data preparedness in a country and hence be able to apply data-driven models that can help to come to a better-informed prioritization of humanitarian aid. We address this knowledge gap by taking a “data ecosystem” approach, thereby referring to “*the people and technologies collecting, handling and using the data and the interactions between them*” [22, p. 557]. By applying this approach to the humanitarian sector, we developed a systematic approach for humanitarian organizations to develop humanitarian data ecosystems that optimize the availability of data for prioritization and decision making.

The ‘Data Services’ component of the above mentioned framework, relating to the required IT infrastructures to facilitate data preparedness, and the possible architectures of these infrastructures, will not be considered in this study as we focus on the upper layers of the OSI model for information systems. We do recognize that excluding the issue of the availability and design of the underlying technical information architecture for storing, communicating

and analyzing the data itself needs to be considered when actually setting up a data ecosystem. The information architecture will depend on local circumstances and readily available information systems such as spatial data platforms, open street map, etc.

This paper is structured as follows. First, we elaborate on the data ecosystem approach in section 2. We then present our research design in section 3. The outcomes of a systematic literature review to develop the generic framework for establishing a data ecosystem is presented in section 4. In section 5, we present the evaluation of the framework by means of expert interviews. In section 6 the framework is adapted to fit with the challenges of developing data ecosystems within the humanitarian sector. We conclude with a discussion of the framework and offer future research to test the framework in a diversity of countries to further enhance its practical use.

2 DATA ECOSYSTEM APPROACH

Academics in information intensive, socio-technical contexts have applied the ecosystems perspective to get an idea of the diverse interrelationships between data users, data providers, data itself, institutions and material infrastructure [10], as the perspective can help to deal with complexity [11]. According to Harrison, Pardo & Cook [10, p. 905], the approach can be used to outline existing conditions and develop desired conditions, stating that “*its users often aim to provoke new thinking about the conditions and requirements necessary to actively cultivate development of an ecosystem to achieve a set of specific and desirable goals*”. Data ecosystems can be used “*as a means for decision-making and planning*” [35, p.18], to locate the “*relative positions of the actors in the ecosystem (data providers, sources, resources and users)*” [31, p. 72], and to “*facilitate access to sharing and using data*” [32, p. 286].

Berens et al. already propose the use of a data ecosystem perspective on the use of digital data in the humanitarian sector: “[T]hat of a complex data ecosystem comprised of a variety of actors that are touched by flows of digital data due to data sharing and data-related service delivery, and hence become part of a phenomenon that stretches beyond their organizational boundaries” [6, p. 5]. The perspective allows for a cross-organizational understanding of data use, rather than only assessing internal data life cycles [6, 26]. Raymond et al. [26] describe the ‘humanitarian data ecosystem’ as the network of humanitarian organizations, affiliated and affected communities who are producing, collecting and analyzing digital data. We argue that it is essential that a humanitarian data ecosystem also comprises actors other than humanitarian organizations: also governmental and private sector organizations must be included, given that they hold key data. In section 6 we present the roles and responsibilities of stakeholders in a humanitarian data ecosystem in detail.

With the rise in data availability from a multitude of sources and a growing sophistication in data modeling, the data ecosystem approach becomes increasingly important to optimize the actual use of the data to address challenges in the humanitarian sector. However, only limited use is made of it as of yet. Based on a literature search in the literature databases Scopus, Web of Science and Google Scholar for ‘humanitarian data ecosystem’ or ‘humanitarian’ and ‘data ecosystem’ did not yield any results. No academic

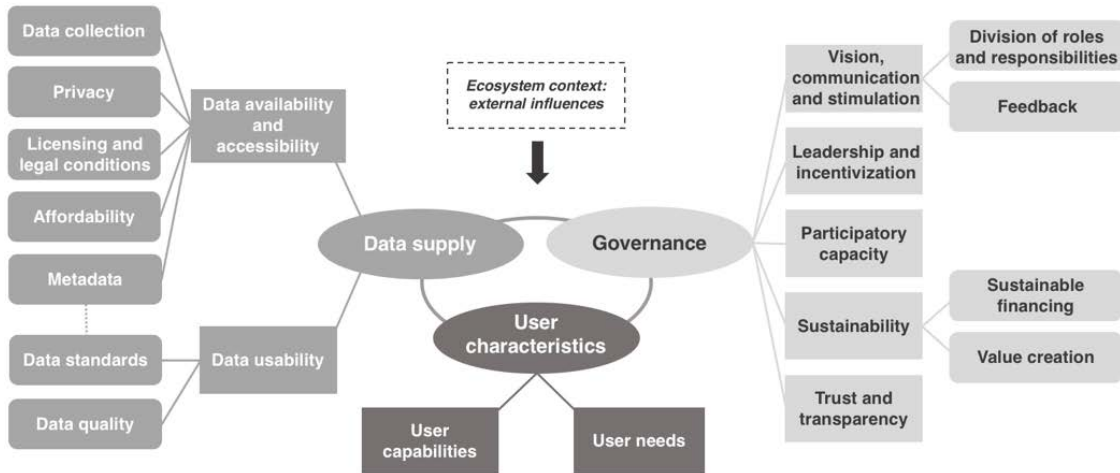


Figure 1: General framework of criteria for a successful data ecosystem

literature so far (May 2017) addresses the humanitarian data ecosystem, or the success criteria for the establishment of a humanitarian data ecosystem. The terms can be found in a limited number of non-academic publications, but in none of these the development of a successful humanitarian data ecosystem is addressed. Berens et al. [6] and Raymond et al. [26] for example, discuss the humanitarian data ecosystem from the angle of the responsible use of data and argue that the data ecosystem perspective can help to address the issue of a lack of centralized governance that results in a risk of harmful data use in the humanitarian sector. Sabou shares this view, claiming that *“the existing ‘humanitarian–digital data’–ecosystem is essentially a collection of ungoverned pilot programs, and a new lens is needed to understand how we can meet the increasing demand for humanitarian aid in disaster onsets with relevant and ethical humanitarian innovation.”* [27, par. 7]. Our literature review presents an academic knowledge gap to which we contribute by developing a framework for humanitarian organizations to develop such data ecosystems. In the next section, we present our research design towards this objective.

3 RESEARCH DESIGN

Our research design consisted of three phases. In the first phase, we used a literature review to retrieve categories of aspects that need to be considered to set up a successful data ecosystem in general. With these categories, we developed a framework for the success criteria for the development of a data ecosystem. This generic framework was evaluated with 17 academic researchers. In the second phase, we tested this generic framework by means of a case study in order to adapt it to the specific application in the humanitarian sector. The case study is the Community Risk Assessment and prioritization toolbox, developed by the Netherlands Red Cross. In the third phase, we assessed the practical relevance of the framework by means of interviews with eight humanitarian data experts. In the following sections, we go into more details on these research phases, to show the intermediate findings towards our final framework.

4 PHASE 1 - GENERIC DESIGN FRAMEWORK FOR DATA ECOSYSTEMS

To start with, we performed a systematic literature review, in which the outcomes from publications addressing the establishment of data ecosystems in general were combined. This allowed us to retrieve a first set of important aspects to consider when aiming to set up a successful data ecosystem.

Relevant literature was found by means of a thorough search for journal and conference articles, books, reports and other informative documents. Scopus, Web of Science, Google Scholar and Google Search were consulted as online databases. We used the following (combinations of) keywords: ‘Data ecosystem’; ‘Open data ecosystem’; ‘Government ecosystem’; ‘Criteria’ + ‘data ecosystem’; ‘Element’ + ‘data ecosystem’; ‘Successful’ + ‘data ecosystem’; ‘Establish’ + ‘data ecosystem’; ‘Data infrastructure’.

The search of relevant articles was an iterative process. By inserting (combinations of) these keywords, a multitude of articles was found. The found articles were judged on relevance by reading the titles and summaries, and by scanning the text: did they indeed address criteria for developing a successful data ecosystem? When an initial selection was made, an additional search for literature took place based on the references used in these selected articles, which resulted in an increase of articles with potential. The new articles were scanned in the same manner, and by doing so we narrowed down the set of articles to 17 publications, which are listed in Table 1.

Next, we manually coded all criteria that were presented in the selected articles, which led to the creation of the overview as presented in table 2. The literature review yielded a diverse set of aspects that needs to be considered when establishing a successful data ecosystem. We found that all aspects could be grouped into one of the categories identified by Welle Donker & Van Loenen [32], in their study on how to assess the success of the open data ecosystem. We therefore used this same categorization to organize the data ecosystem success criteria resulting from our literature review. Three different categories are distinguished:

Author(s)	Year	Focus	Objective	Domain of Author(s)	Ref.
Attard, Orlandi & Auer	2016	Economic data ecosystem	To project their vision of generating a new Economic Data Ecosystem that has the Web of Data as its core.	Enterprise Information Systems	[3]
Barthélemy	2016	Open data ecosystem	To provide an overview of the Belgian open data ecosystem	School of Management: MSc in Business Engineering	[4]
Davies	2012	Open data infrastructures and ecosystem	To highlight some of the interventions that may be necessary to support realization of impact from open data initiatives.	Open Data & Open Government	[8]
Dawes, Vidiásova & Parkhimovich	2016	Open government data ecosystem	To develop a preliminary ecosystem model for planning and designing Open Government Data programs	E-governance / Open Government	[9]
Harrison, Pardo & Cook	2012	Open government ecosystem	To create a research and development agenda with questions essential to the development of Open Government Ecosystems.	IT in Government / Open Government	[10]
Heimstädt, Saunderson & Heath	2014	Open data ecosystem	To identify a set of structural business ecosystem properties.	Open (Government) Data	[11]
Immonen, Palviainen & Ovaska	2014	Open data based business ecosystem	To define the requirements of an open data based business ecosystem (an open data ecosystem from the business viewpoint)	Service / Software Engineering	[14]
Jetzek	2017	Open data ecosystem	To explore the possibilities for sustainable value generation in the Open Data Ecosystem	IT Management: Big Data & Open Data	[16]
Lee	2014	Open data ecosystem	To specify a series of specific elements critical for building an Open Data Ecosystem	Linked & Open Data	[17]
Macharis & Crompvoets	2014	Spatial data infrastructure	To evaluate development scenarios for the spatial data infrastructure for Flanders	Supply Chain Management / Public Governance / Spatial Data Infrastructures	[18]
Parsons et al.	2011	Science data ecosystem	To success several short- and long-term strategies to facilitate a socio-technical evolution in the overall science data ecosystem	Science Data	[22]
Pollock	2011	Open data ecosystem	To stress the importance of data cycles with feedback loops when building open data ecosystems	Open Knowledge	[23]
Ponte	2015	Open data ecosystem	To provide an overview of the issues to be addressed when enabling an open data ecosystem	Organization Science & ICT	[24]
Van Schalkwyk, Willmers & McNaughton	2016	Open data ecosystem	To consider the supply, demand and use of open data, as well as the roles of intermediaries, using an ecosystem approach	Open data / Open ICT Ecosystems & Scholarly Communication	[31]
Welle Donker & Van Loenen	2017	Open data ecosystem	To develop an open data assessment framework based on three output indicators as conditions for a successful open data ecosystem	Open Data / Information Infrastructures	[32]
Wiener et al.	2016	Open data ecosystem	To enable an open data ecosystem for the neurosciences	Neuroscience & Data-sharing	[34]
Zuiderwijk, Janssen & Davis	2014	Open data ecosystem	To provide an overview of essential elements of open data ecosystems for enabling easy publication and use of open data	Open Data / ICT & Governance	[35]

Table 1: List of articles reviewed (objectives and methodologies have been retrieved from the articles)

- (1) Data supply, relating to the provision of data as open data,
- (2) Governance, being the framework of policies, processes and instruments to realize common goals in the interaction between entities (and facilitating the data supply), and
- (3) User characteristics [32].

An overview of the articles in which each criterion appeared, can be found in Table 2. Later in this section, all criteria are described in full.

We evaluated the findings from our literature review in a validation session with a group of open data researchers. During this session, the completeness of the literature list (as presented in Table 1), was assessed, as well as the completeness and correctness of the set of data ecosystem governance criteria. All participants were provided with a hand-out of both lists. The outcomes were elaborated on in a presentation, after which an open discussion took place in which the participants provided feedback. Taking this feedback into account, the findings from the literature review and the validation session led to the framework as visualized in Figure 1. In the following paragraphs we explain the criteria that are used in the framework.

4.1 Data Supply (DS) Criteria

The Data Supply (DS^o) criteria are described as follows:

DS1 Data availability and accessibility: It should be clear for actors in a data ecosystem how to find data, where to find it and how to access it.

DS1a Data collection: It should be clear to actors in the data ecosystem how data can be collected.

DS1b Privacy: It should be ensured that data protection laws are followed.

DS1c Licensing and legal conditions: Another important aspect to enhance the accessibility of data in a data ecosystem is the type of licensing associated with the data, which is necessary to ensure the legal foundation for the potential (re)use of data

DS1d Affordability: The data shared in the data ecosystem should be affordable.

DS1e Metadata: The existence of appropriate metadata can help to improve the availability and accessibility of data.

DS2 Data usability: High quality of the data shared in a data ecosystem should be ensured to enhance its usability:

Category	Criterion	Description	Source(s)
Data Supply	Data availability and accessibility	It should be clear for actors in a data ecosystem how to find data, where to find it and how to access it	[3, 9, 10, 14, 17, 18, 22, 31, 32, 34, 35]
	Data collection	It should be clear to actors in the data ecosystem how data can be collected	Added after expert validation
	Privacy	It should be ensured that data protection laws are followed	[9, 14, 17, 34]
	Licensing and legal conditions	Another important aspect to enhance the accessibility of data in a data ecosystem is the type of licensing associated with the data, which is necessary to ensure the legal grounding for the potential (re)use of data	[4, 9, 14, 17, 18, 22, 31, 32, 35]
	Affordability	The data shared in the data ecosystem should be affordable.	Added after expert validation.
	Metadata	The existence of appropriate metadata can help to improve the availability and accessibility of the data	[8–10, 17, 18, 22, 32, 34, 35]
	Data usability: Data quality	High quality of the data shared in a data ecosystem should be ensured to enhance its usability	[3, 4, 8–10, 14, 17, 18, 22, 31, 32, 35]
	Data standards	An important aspect that determines the quality of data is the presence of standards to facilitate data interoperability	[4, 9, 10, 14, 17, 18, 32, 34]
Data Governance	Vision, communication and stimulation	A collaborative, interactive environment should be established and cooperation between stakeholders should be stimulated	[3, 4, 9, 14, 32, 34]
	Division of roles and responsibilities	There should be a clear division of the roles and responsibilities of the actors in an ecosystem	[4, 8–10, 14, 17, 31, 35]
	Feedback	The data ecosystem should include feedback mechanisms to enable data users to provide feedback to data providers	[4, 9, 10, 14, 17, 35]
	Leadership and incentivization	Activities in an ecosystem should be stimulated, incentivized and coordinated (either top-down or bottom-up, depending on the context)	[4, 8–10, 14, 17, 32, 34]
	Participatory capacity	Data supply should match data demand in an ecosystem; public bodies require certain capacities to be able to participate in an ecosystem	[3, 9, 17, 32, 34]
	Sustainability: financing and value creation	For a data ecosystem to become sustainable, sustainable financing should be arranged and the value should be generated for the ecosystem stakeholders	[8, 10, 11, 16, 17, 22, 24, 31, 32, 34]
	Trust and transparency	Data providers and data users should mutually trust and be transparent to each other.	Added after expert validation.
User Characteristics	User capabilities	The capabilities of the data users in an ecosystem should be considered	[4, 8–10, 17, 32, 35]
	User needs	The needs of the data users in an ecosystem should be considered	[9, 10, 14, 17, 32]
External Context	-	The external context of the data ecosystem should be considered	[4, 9, 10, 18, 22, 31, 35]

Table 2: Criteria for a successful data ecosystem found in the systematic literature review, complemented with insights from expert.

DS2a Data standards: An important aspect that determines the quality of data is the presence of standards to facilitate data interoperability.

DS2b Data quality: The data shared in the data ecosystem should be of high quality

4.2 Governance (G) Criteria

The Governance (Gⁿ) criteria are described as follows:

G1 Vision, communication and stimulation: A collaborative, interactive environment should be established and cooperation between stakeholders should be stimulated. A fragmented approach should be avoided.

G1a Division of roles and responsibilities: All actors in the data ecosystem should be identified, and there should be a clear division of the roles and responsibilities of the actors in a data ecosystem.

G1b Feedback: The data ecosystem should include feedback mechanisms to enable data users to provide feedback to data providers.

G2 Leadership and incentivization: Activities in a data ecosystem should be stimulated, incentivized and coordinated by a problem owner, either top-down or bottom-up. Incentivization can take place by for example make data sharers feel intrinsically rewarded, or by lowering the barrier to entry for data sharing.

G3 Participatory capacity: Data supply should match data demand in a data ecosystem: public bodies require certain capacities to be able to participate in a data ecosystem, like:

- Technical knowledge on certain systems and technologies involved
- Data management knowledge, on how to ensure high data quality
- Operational knowledge, on how to incorporate data activities into current practices

G4 Sustainability: For a data ecosystem to become sustainable, sustainable financing should be arranged and value should be generated for the ecosystem stakeholders

G4a Financing: Arrange for sustainable financing

G4b Value creation: Value generation for the ecosystem stakeholders

G5 Trust and transparency: Data providers and data users should mutually trust and be transparent to each other

4.3 User Characteristics (UC)

UC1 User capabilities: The capabilities of the data users in an ecosystem should be considered.

UC2 User needs: The needs of the data users in an ecosystem should be considered.

4.4 External Context (EC)

One last aspect related to data ecosystems coming forward in the literature is the importance of considering the external context of an ecosystem. This is not exactly a criterion for a successful data ecosystem; it is merely an overarching factor that should be considered when analyzing a data ecosystem. The functioning of a

data ecosystem also depends on for example the local culture, the political system and historical influences. These aspects all together form the institutional conditions in which the ecosystem is or needs to be embedded, that influence how actors in the ecosystem function and how different ecosystem elements are arranged. Hence:

External context: The external context of the data ecosystem should be considered.

As can be seen in Table 1, 14 of the 16 articles focus on open data ecosystems and many of the authors have a background in open (government) data. To assess whether the criteria that hold for the successful establishment of open data ecosystems specifically can also be translated into the humanitarian data ecosystem, we used a case study in the humanitarian sector in the next phase of our research.

5 PHASE 2 - PRACTICAL VALIDATION

In our literature review we discovered that the ecosystem approach was mainly applied in the research domain of open data. The literature yielded an initial set of criteria, but also raised the issue whether these are fully applicable to the humanitarian sector, in which much of the data is not shared openly (and sometimes not even stored digitally). Therefore, we empirically validated the generic framework of success criteria for the development of a data ecosystem in the humanitarian sector.

For this empirical validation, we tested the framework by means of the ‘Community Risk Assessment and Prioritization toolbox’, developed by the data team of the Netherlands Red Cross. This is an alternative data-driven solution that preemptively gathers and combines relevant data on three risk dimensions—Lack of Coping Capacity, Vulnerability and Hazard & Exposure, according to the INFORM Risk Index [15]—to provide a detailed risk assessment for areas and communities in a country. This risk assessment supports a faster identification of priority areas for humanitarian intervention related to natural disasters. The output is an easy-to-understand dashboard in which colors on a map visualize the risk or predicted damage for a specific area. Currently, this toolbox is under development for up to 10 countries (status January 2018) [2]. To be able to develop this model for a country, a humanitarian data ecosystem needs to be created in which data from different administrative levels can be collected, collated and users are stimulated to make use of the model.

As potential users of the humanitarian data ecosystem framework, we asked seven interviewees involved in the development of the Community Risk Assessment and Prioritization toolbox to assess whether additional criteria could be identified that were relevant specifically for the humanitarian sector.

Secondly, to check for the practical relevance of the framework criteria on a regional/country-specific level, interviews with eight humanitarian data experts were held. These data experts represented five different regions (Greater Horn of Africa¹, Latin America and Caribbean, Malawi, the Philippines and the Southern Africa Development Community²), to assess a variety of data landscapes and to be able to distinguish between country-specific and generic

answers. They all had knowledge on the broader data landscape in their country, and experience with collecting data similar (or identical) to the input data for the Community Risk Assessment and Prioritization toolbox on low administrative levels, which is necessary for the toolbox to generate accurate outcomes. In the following section, we present the final framework for developing a data ecosystem for humanitarian purposes.

6 PHASE 3 - FRAMEWORK FOR HUMANITARIAN DATA ECOSYSTEMS

The practical validation showed that the generic framework, which was predominantly based on literature from the open data research domain, needed to be adapted in several aspects to make it suitable for the context of the humanitarian data ecosystem. Firstly, the addition of a staged approach towards sets of criteria was proposed by the developers of the Community Risk Assessment and Prioritization toolbox. Secondly, the importance of the (development of the) governance part of the framework was recognized. And thirdly, the empirical evidence as put forward by the expert interviews led to an adaptation of the (classes of) governance criteria. We will elaborate on these adaptations in the following paragraphs.

First, the developers of the Community Risk Assessment and Prioritization toolbox indicated that the ‘data supply’ and ‘user characteristics’ criteria both relate to the design and development of the toolbox. In contrast, the ‘governance’ criteria relate to the creation of an optimal context for the development of the toolbox. Having practical experience with the development of the toolbox, they evaluated the relevance of the ‘data supply’ and ‘user characteristics’ criteria and confirmed that all related criteria were relevant in the context of the humanitarian data ecosystem. However, they also stressed that some of the elements will receive a stronger focus in a later stage of the data ecosystem, stating that the humanitarian data ecosystem around the Community Risk Assessment and Prioritization toolbox is still in a very early phase of development. In the current stage, the focus is mainly on the data supply criteria. These are directly related to data collection and collation, which is essential to be able to start developing the Community Risk Assessment and Prioritization toolbox in the first place. To illustrate this, the criterion of data standards for example, is indeed ‘nice to have’ and makes the data collection and analysis easier, but it is also possible to get the data ecosystem running without data standards in place. In the early stage of the humanitarian data ecosystem development, also the user characteristics require extra attention. If the outcomes of the toolbox do not match the user capabilities or information needs of humanitarian decision-makers, the desired effect will not be achieved.

The second result of the empirical validation was the labeling of the governance part of the framework as important (being a major challenge when developing a data ecosystem). This is indeed an issue with the humanitarian data ecosystem around the Community Risk Assessment and Prioritization toolbox: there are no, or very few, governance structures in place yet that can stimulate the data ecosystem development. Many fragmented organizations are involved that are inclined to pursue their own objectives, while there

¹The Greater Horn of Africa comprises eight countries: Djibouti, Eritrea, Ethiopia, Kenya, Somalia, South Sudan, Sudan and Uganda.

²The Southern Africa Development Community (SADC) has 16 member states: Angola, Botswana, Comoros, Democratic Republic of the Congo, Lesotho, Madagascar, Malawi,

Mauritius, Mozambique, Namibia, Seychelles, South Africa, Swaziland, Tanzania, Zambia and Zimbabwe.

is a lack of central governance to coordinate and align all these individual activities. Berens et al. [6] refer to this situation as the ‘governance disparity’, which they describe as follows: “*In humanitarian work, the multitude of independent organizations, government actors, multinational initiatives, individuals, private sector recourse providers, and digital platforms constitute a dynamic ecosystem with no clear leader or dominant force*” [26, p.1].

Thirdly, the interviews with eight humanitarian data experts representing five different regions, and one having global experience, yielded an overview of the main practical implications of the theoretical governance criteria in the current humanitarian data ecosystem. Assuming that a humanitarian data ecosystem spans one country, and that the governance is taken care of within this country, the data experts offered the following implications that need to be included when setting up a humanitarian data ecosystem within a country:

- (1) **Vision, communication and stimulation:** To create a sense of responsibility and ownership, to have all ecosystem stakeholders understand the value and importance of data-sharing and to address the issue of fragmentation, it is important that the party in the leadership role actively involves all other relevant stakeholders in the establishment of the data ecosystem.
- (a) **Division of roles and responsibilities:** In the specific context of the Community Risk Assessment and Prioritization toolbox, there are four different roles to be identified and assigned, of which the first two have a leading role (and hence also relate to DG2 – Leadership and incentivization in the framework):
 - *The initiator/coordinator*, which is the leading agency, most likely a United Nations agency, in setting up the humanitarian data ecosystem. In the toolbox context, this party should initiate, coordinate and support the development of the toolbox in a country, bring all relevant actors together and incentivize them to facilitate data-sharing and to use the toolbox.
 - *The local lead*, which is a national body that becomes the local manager after the initiator/coordinator has set out the initial data ecosystem, and is responsible for locally promoting participation. This role is important to ensure the sustainability of the ecosystem.
 - *Data providers* should be identified and are responsible for sharing their data that is considered relevant.
 - *Toolbox users*, which are humanitarian or government agencies who should use the toolbox (i.e. the data processed into understandable information) and incorporate the outputs in their operations and decision-making.
- (b) **Feedback:** The possibility to provide feedback to data providers is considered useful, but irrelevant in the current phase of the humanitarian data ecosystem: collecting relevant primary data and collating existing data is already sufficiently challenging. The interviewees rather look for alternatives than putting effort into the development of feedback mechanisms. It is relevant, but in a later, more developed stage of the humanitarian data ecosystem. The same holds for transparency from data providers on how

they got their data—‘transparency’ is considered to be overlapping with ‘feedback’.

- (2) **Leadership and incentivization:** The data experts suggested to split up this element, as both aspects can be assessed individually. ‘Leadership’ was already addressed with the division of roles and ‘responsibilities’ (indicating an overlap) and incentivization is considered very important, and something that relates to all identified roles in the data ecosystem. One of the most suggested approaches to incentivize parties to participate in the data ecosystem was by meeting the vision, ‘communication and stimulation’ criterion: to involve stakeholders from the start.
- (3) **Participatory capacity:** In addition to the importance of incentivization, the experts also labeled ‘participatory capacity’ as very important, and something that needs to be addressed and built for every single role in the ecosystem. This criterion is essential in the humanitarian data ecosystem, as stakeholders generally lack technical and information management capacities, especially at lower administrative levels.
- (4) **Sustainability (financing and value creation):** Sustainable financing should be the responsibility of the initiator/coordinator of the data ecosystem. The criterion of ‘value creation’ in the general framework is considered to be overlapping with incentivization as parties can be incentivized to participate by creating value for them, hence the experts consider ‘value creation’ as a criterion of ‘incentivization’.
- (5) **Trust and transparency:** The experts suggested to merge transparency with ‘feedback’ as they relate to each other. Trust between actors in the data ecosystem is indeed considered important. However, it is not a criterion for a successful humanitarian data ecosystem, but rather something that follows from the creation of the vision, ‘communication and stimulation’ criterion.

Based on the empirical validation, the governance part of the framework was changed by adapting the (classification of) the governance criteria. The changes are indicated by a bold border in Figure 2.

By means of the empirical validation we adapted the generic framework based on the criteria of developing ecosystems for open data research to fit with the context of the humanitarian sector. In the following section, we discuss and reflect on this deliverable and offer future research topics.

7 DISCUSSION AND CONCLUSION

This study has taken a ‘data ecosystem’ approach to create a framework to guide the facilitation, stimulation and coordination of data collection, collation and data sharing between humanitarian actors to increase the level of data preparedness in a country. By doing so, this study is the first academic study to address the ‘humanitarian data ecosystem’. A framework of criteria for the successful development of a data ecosystem has been developed based on a systematic literature review and was validated with a panel of open data experts. Subsequently, this framework has been tested on its application and relevance in the humanitarian sector by means of a case study. The need to test this stemmed from the fact that the



Figure 2: Governance criteria after empirical validation

ecosystem approach has so far mainly been applied in the research domain of open data. However, the issue in the humanitarian sector is that much data is not shared openly, and sometimes not even stored digitally. The question to be answered was therefore: should these same criteria be considered when establishing a data ecosystem in the complex, fragmented humanitarian context around the Community Risk Assessment and Prioritization toolbox developed by the Netherlands Red Cross?

We found that rather than classifying the humanitarian data ecosystem as a distinct type of data ecosystem, it is a data ecosystem that is in an emerging phase of development whereby it links to other "sub-data ecosystems or in some cases data collaboratives", that collect, handle and use data on specific risk indicators (such as in relation to the health sector) - with a relatively long retention period - or crisis data with a short retention period. To stimulate its development, the special point of attention in the humanitarian data ecosystem appeared to be the governance part of the framework. Therefore, eight respondents representing five different regions with varying data landscapes were asked about how they perceived the relevance and application of the different governance criteria in their region. It was hereby assumed that a humanitarian data ecosystem spans one country, and that the governance is taken care of within this country. The interviewees also specified the practical implications of the criteria for the humanitarian data ecosystem specifically 'Value creation' and 'trust' were found to be overlapping with other criteria and hence omitted, but all the other criteria were considered very relevant. A remark must be made for 'transparency and feedback' though; according to the interviewees these only become relevant in a later data ecosystem stadium, due to the current focus on enhancing, facilitating and speeding-up data-sharing practices.

By means of the interviews, it was also found that many of the governance criteria closely interrelate with 'data supply' and 'user characteristics' criteria: they are all interdependent. The division of roles and responsibilities for example, should also include the identification of data providers - relating to the data supply part of the framework, and toolbox users - relating to the user characteristics part of the framework. Moreover, the 'leadership' is also a

part of the division of roles and 'responsibilities', and likewise 'incentivization' and 'participatory capacity' hold for every identified role.

Approaching data-sharing in the humanitarian sector with a data ecosystem perspective provides insight into the elements to be addressed and helps to provide structure. However, the interrelations between the criteria are a lot more complex than visualized in Figure 1, which indicates that the humanitarian data ecosystem might be too complex to capture in a static framework. Additionally, some of the criteria are considered to be of secondary importance: they only become relevant in a late stage of the data ecosystem. Susha, Janssen & Verhulst [28] try to capture the more dynamic elements through introducing the concept of Data Collaboratives, a cross-sector (and public-private) collaboration initiative aiming at data collection, sharing, or processing for the purpose of addressing a societal challenge.

To summarize, the humanitarian data ecosystem is emerging and currently in its initial development phase. There is a strong focus on 'data supply' and on - though less but still present - 'user characteristics', but the governance part requires attention to create an optimal context to facilitate meeting the data supply and user characteristics criteria.

Future research should extend this study by also testing the theoretical framework of data ecosystem success criteria to a set of other case studies, as an extra validation. This would both provide insight into the generic applicability of the framework, and in the specificity of the findings for the humanitarian data ecosystem. Other case studies can also provide insight into the evolution of the maturity of other data ecosystems, which is something the immature humanitarian data ecosystem can learn from. Moreover, future research should address the identification and establishment of the IT infrastructures, including geospatial data-sharing platforms, required to support the development of the humanitarian data ecosystem. The architectural design of these infrastructures should match the local context, and the capabilities of the stakeholders in the data ecosystem. We also recommend future research into potential undesired consequences of setting up a humanitarian data ecosystem. The selection of actors to be included or excluded from the data ecosystem can lead to shifts in tasks and, consequently, power balances amongst stakeholders, which can lead to power abuse. Additionally, as the data will also include personal data, compliance with personal data protection laws and regulation need to be carefully assessed. In general, unethical use of data (be it personal or other types of data) needs to be carefully assessed. Lastly, the application in practice of the developed governance structure should be evaluated. For this purpose, actual use of the framework in a diversity of countries can yield even more empirical evidence for either confirmation or refinement of (the components and their criteria in) the framework. This empirical approach will also enable the inclusion of the viewpoints of other actors than the data experts to which our study was limited.

ACKNOWLEDGMENTS

We would like to thank the contributors to this study: all interviewees, Jannis Visser for developing the Community Risk Assessment

and Prioritization toolbox, and the Princess Margriet Fund for financing the initial toolbox development.

REFERENCES

- [1] 510 Global. 2016. Personalised Humanitarian Aid Through Big Data: an initiative of the Dutch Red Cross. (2016). <http://510.global> (accessed: March 21, 2017).
- [2] 510 Global. 2017. Community Risk Assessment and Prioritization toolbox. (2017). <https://dashboard.510.global/> (accessed: March 21, 2017).
- [3] Judie Attard, Fabrizio Orlandi, and Sören Auer. 2016. Data Value Networks: Enabling a New Data Ecosystem. In *Proc. IEEE/WIC/ACM International Conference on Web Intelligence (WI)*. 453–456.
- [4] Florian Barthélemy. 2016. "The Belgian open data ecosystem and innovation through open data. (2016).
- [5] A Benini. 2015. The use of Data Envelopment Analysis to calculate priority scores in needs assessments. *Active Learning Network for Accountability and Performance in Humanitarian Action (ALNAP)*, London, UK (2015).
- [6] J Berens, N Raymond, G Shomshon, S Verhulst, and L Bernholz. 2015. The Humanitarian Data Ecosystem: the Case for Collective Responsibility. (2015). https://pacscenter.stanford.edu/wpcontent/uploads/2017/08/humanitarian_data_ecosystem.pdf (accessed: September 19, 2017).
- [7] Tina Comes. 2016. Cognitive biases in humanitarian sensemaking and decision-making lessons from field research. In *Proc. IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)*. 56–62.
- [8] Tim Davies. 2011. Open Data: infrastructures and ecosystems. *Open Data Research* (2011).
- [9] Sharon S Dawes, Lyudmila Vidiyasova, and Olga Parkhimovich. 2016. Planning and designing open government data programs: An ecosystem approach. *Government Information Quarterly* 33, 1 (2016), 15–27.
- [10] Teresa M Harrison, Theresa A Pardo, and Meghan Cook. 2012. Creating open government ecosystems: A research and development agenda. *Future Internet* 4, 4 (2012), 900–928.
- [11] Maximilian Heimstädt, Fredric Saunderson, and Tom Heath. 2014. Conceptualizing Open Data ecosystems: A timeline analysis of Open Data development in the UK. In *Proc. Conference for E-Democracy and Open Government*. 245.
- [12] Dorothea Hilhorst. 2013. *Disaster, conflict and society in crises: everyday politics of crisis response*. Routledge.
- [13] IEAG, United Nations. 2014. A World that Counts—Mobilising the Data Revolution for Sustainable Development. (2014). <http://www.undatarevolution.org/report> (accessed: January 24, 2018).
- [14] Anne Immonen, Marko Palviainen, and Eila Ovaska. 2014. Requirements of an open data based business ecosystem. *IEEE access* 2 (2014), 88–103.
- [15] INFORM. 2016. Methodology. (2016). <http://www.inform-index.org/InDepth/Methodology> (accessed: June 24, 2017).
- [16] Thorhildur Jetzek. 2017. Innovation in the Open Data Ecosystem: Exploring the Role of Real Options Thinking and Multi-sided Platforms for Sustainable Value Generation through Open Data. In *Analytics, Innovation, and Excellence-Driven Enterprise Sustainability*. Springer, 137–168.
- [17] Deirdre Lee. 2014. Building an open data ecosystem: an Irish experience. In *Proc. International Conference on Theory and Practice of Electronic Governance*. ACM, 351–360.
- [18] Cathy Macharis and Joep Crompvoets. 2014. A stakeholder-based assessment framework applied to evaluate development scenarios for the spatial data infrastructure for Flanders. *Computers, Environment and Urban Systems* 46 (2014), 45–56.
- [19] Asmat Monaghan and Mark Lycett. 2013. Big data and humanitarian supply networks: Can Big Data give voice to the voiceless?. In *Proc. IEEE Global Humanitarian Technology Conference (GHTC)*. 432–437.
- [20] S O'Brien. 2017. This is how we build a stronger, data-driven humanitarian sector. (2017). <https://www.weforum.org/agenda/2017/01/this-is-how-we-build-a-stronger-data-driver-humanitarian%20sector> (accessed: July 3, 2017).
- [21] MT Ortuño, P Cristóbal, JM Ferrer, FJ Martín-Campo, S Muñoz, G Tirado, and B Vitoriano. 2013. Decision aid models and systems for humanitarian logistics. A survey. In *Decision aid models for disaster management and emergencies*. Springer, 17–44.
- [22] Mark A Parsons, Øystein Godøy, Ellsworth LeDrew, Taco F De Bruin, Bruno Danis, Scott Tomlinson, and David Carlson. 2011. A conceptual framework for managing very diverse data for complex, interdisciplinary science. *Journal of Information Science* 37, 6 (2011), 555–569.
- [23] Rufus Pollock. 2011. Building the (open) data ecosystem. *Open Knowledge Foundation Blog* 31 (2011).
- [24] Diego Ponte. 2015. Enabling an Open Data Ecosystem.. In *ECIS*.
- [25] N Raymond and Z Al Achkar. 2016. Data preparedness: connecting data, decision-making and humanitarian response. (2016). http://hhi.harvard.edu/sites/default/Harvard_Humanitarian_Initiative_signal_program_on_human_security_and_technology_standards_and_ethics_series_01, (accessed: July 4, 2017).
- [26] N Raymond, Z Al Achkar, S Verhulst, and J Berens. 2016. Building data responsibility into humanitarian action. (2016). http://www.unocha.org/sites/dms/Documents/TB18_Data%20Responsibility_Online.pdf OCHA Policy and Studies Series 018, (accessed: July 4, 2017).
- [27] J Sabou. 2016. The Humanitarian "Data" Ecosystem: Intersection of Innovation & Disaster Relief. (2016). <http://www.leidensafetyandsecurityblog.nl/articles/the-humanitarian-data-ecosystem-intersection-of-innovation-disaster-relief> (accessed: July 3, 2017).
- [28] Iryna Susha, Marijn Janssen, and Stefaan Verhulst. 2017. Data collaboratives as a new frontier of cross-sector partnerships in the age of open data: taxonomy development. (2017).
- [29] Bartel Van de Walle and Tina Comes. 2014. Risk accelerators in disasters: Insights from the Typhoon Haiyan Response on Humanitarian Information Management and Decision Support. In *Proc. International Conference on Advanced Information Systems Engineering*. Springer, 12–23.
- [30] Marc van den Homberg, Jannis Visser, and Maarten van der Veen. 2017. Unpacking Data Preparedness from a humanitarian decision making perspective: toward an assessment framework at subnational level. In *Proc. Information Systems for Crisis Response and Management (ISCRAM) Conference*.
- [31] Francois Van Schalkwyk, Michelle Willmers, and Maurice McNaughton. 2016. Viscous open data: The roles of intermediaries in an open data ecosystem. *Information Technology for Development* 22, sup1 (2016), 68–83.
- [32] Frederika Welle Donker and Bastiaan van Loenen. 2017. How to assess the success of the open data ecosystem? *International Journal of Digital Earth* 10, 3 (2017), 284–306.
- [33] K Whipkey and A Verity. 2015. Guidance for incorporating Big Data into Humanitarian Operations. (2015). <http://digitalhumanitarians.com/resource/incorporating-big-data-humanitarian-operations> (accessed: April 5, 2017).
- [34] Martin Wiener, Friedrich T Sommer, Zachary G Ives, Russell A Poldrack, and Brian Litt. 2016. Enabling an open data ecosystem for the Neurosciences. *Neuron* 92, 3 (2016), 617–621.
- [35] Anneke Zuiderwijk, Marijn Janssen, and Chris Davis. 2014. Innovation with open data: Essential elements of open data ecosystems. *Information Polity* 19, 1, 2 (2014), 17–33.