Guidance for Calculating Humanitarian Data-Collection Costs

COMMUNICATION CHANNELS AND INFRASTRUCTURE

DATA AND TECHNICAL PREPAREDNESS

DATA COLLECTION

DATA LITERACY AND TRAINING

DISASTER MAGNITUDE

LEGAL REQUIREMENTS

ORGANIZATIONAL READINESS

PHYSICAL ACCESS

POLICIES (INTERNATIONAL, NATIONAL AND LOCAL)

PRIVACY REQUIREMENTS

SECURITY REQUIREMENTS

DISASTER BACKGROUND

	Labor Costs	Operational Costs	Total Costs
TOTALS	$	$	$

April 2021

DH Network
Digital Humanitarian Network

Guidance for Calculating Humanitarian Data-Collection Costs

By

Alex Muller (ajcmuller@gmail.com)
Tilburg School of Economics and Management (TiSEM),
MSc Information Management Tilburg University

Andrej Verity (verity@un.org | @andrejverity)
United Nations Office for the Coordination of Humanitarian Affairs (OCHA)

Kenny Meesters (k.j.m.g.meesters@tilburguniversity.edu | @k_meesters)
Tilburg University

Design

Lorena Velázquez (velazquez@reliefweb.int | kunoichidesign.com)
United Nations Office for the Coordination of Humanitarian Affairs (OCHA)

Table of Contents

Key messages

- Data in the humanitarian sector is often collected without a discussion about corresponding costs. Data-collection costs have increased due to the expectation that data simply exists and costs are limited. Therefore, a discussion on the cost-efficiency of data-collection projects in the humanitarian sector is essential.

- The development of a framework to calculate humanitarian data-collection costs uncovered various factors (7 constructs, 25 attributes) that affect the costs of humanitarian data-collection projects.

- To address the factors that influence humanitarian data-collection costs, data collection is split into primary and secondary data based on the cost drivers. For primary data collection the main cost drivers are operational costs and labour hours. For secondary data collection the main cost drivers are labour hours.

- Long-term cost reductions can be achieved consistently through three factors: the level of preparation (through training and data preparedness), the level of project planning and the efficiency of inter-organizational data sharing.

- Data security and privacy are often overlooked. Given the high risks associated with the type of data often being captured, security and privacy must be at the forefront of all projects, thus increasing the cost estimate.

- Having and leveraging humanitarian data for decision-making is critical. This framework does not question its value, but it does provide a simple tool to determine if the collection costs could be optimized or reduced while delivering the same impactful result.

Interviewees

■ ■ ■

Interviews were conducted with 19 professionals in the fields of humanitarian response, humanitarian data management and information management. Of these interviews, 11 were conducted to develop the framework, 3 to validate the framework and 5 to conduct a case study as further validation.

Name	Organization	Position
Lars Peter Nissen	ACAPS	Director
Yan Wang	Delft University of Technology	Data Steward Coordinator
Yaelle Link	IFRC	Information Management Officer
Guido Pizzini	Independent Humanitarian Consultant	Consultant
Juan Duarte	MapAction	Technical Director
Ana Maria Pereira	OCHA	Information Management Officer
James Steel	OCHA	Information Management Officer
John Marinos	OCHA	Information Management Officer
Godfrey Takavarasha	OCHA Centre for Humanitarian Data	Information Management Officer
Javier Teran Castro	OCHA Centre for Humanitarian Data	Data Partnership Team Lead
Shahrooz Badkoubei	OCHA Centre for Humanitarian Data	Programme Management Officer
Patrice Chataigner	Okular Analytics	Information Management Officer
Gisli Olafsson	One Acre Fund	Strategy Technology Officer
Valeria Villan	Standby Task Force	Data Management Volunteer
David Whitford	Statistics Without Borders	Engagement Director
Yves Kim Creach	The Operations Partnership	Associate Director
Kenny Meesters	Tilburg University	Professor
Anonymized	Tilburg University	Assistant Professor
Leonie Tax	UNHCR	Information Management Officer

Introduction

Over the past decade, the amount of data collected by the humanitarian sector has increased exponentially. [1] This data revolution comes with the promise to increase the (cost-) efficiency of humanitarian aid. [2][3] However, humanitarian data is highly complex by nature, since it is time dependent and requires accuracy and quality. [4][5] Furthermore, the data is often used to make life-saving decisions, putting stress and pressure on both the data provider and the decision maker. [6] Yet organizations are not prepared to either estimate or articulate the cost of a data-collection exercise. This gap results in a lack of transparency with management, donor organizations and Governments. [7][8]

This paper develops a framework to help calculate humanitarian data-collection costs. Through this framework, each factor that influences the data-collection costs can be understood and calculated individually. This approach leads to an indicative cost estimate for a data-collection project. By leveraging the results, organizations can better understand opportunities they may not have considered, such as adjusting the collection plan, introducing efficiencies and improving partnerships. Essentially, they can answer the questions of whether and how to conduct the data-collection project.

1 Muller, A. (2021). "Towards a Framework for Humanitarian Data Collection Costs."

2 Read, R., Taithe, B., & Mac Ginty, R. (2016). Data hubris? Humanitarian information systems and the mirage of technology. Third World Quarterly, 37(8), 1314-1331

3 United Nations Office for the Coordination of Humanitarian Affairs (OCHA) (2016, May). "Building data responsibility into humanitarian action". OCHA Policy and Studies Series (018)

4 Van de Walle, B., & Turoff, M. (2008). Decision support for emergency situations. In Handbook on decision support systems 2 (pp. 39-63). Springer, Berlin, Heidelberg.

5 Comes, T., Vybornova, O., & Van de Walle, B. (2015, March). Bringing Structure to the Disaster Data Typhoon: An Analysis of Decision-Makers' Information Needs in the Response to Haiyan. In AAAI Spring Symposia.

6 Knox Clarke, P., & Campbell, L. (2020). Decision-making at the sharp end: a survey of literature related to decision-making in humanitarian contexts. Journal of International Humanitarian Action, 5, 1-14.

7 Puett, C. (2019). Assessing the cost-effectiveness of interventions within a humanitarian organisation. Disasters, 43(3), 575-590.

8 Besiou, M., & Van Wassenhove, L. N. (2020). Humanitarian operations: A world of opportunity for relevant and impactful research. INFORMS Journal on Computing: Manufacturing & Service Operations Management, 22(1), 135-145.

Estimating Humanitarian Data-Collection Costs

■ ■ ■

Research by Muller [9] has established a framework to estimate humanitarian data-collection costs. The framework consists of 7 main constructs, which comprise 25 attributes and are connected through 9 relationships.

The Seven Constructs

01 Level of preparation

The level of preparation refers to the amount of preparedness in place before a disaster. It is determined from a combination of individual, organizational and data perspectives.
Attributes: Training; Data preparation; Awareness

02 Individual's capabilities

The individual's capabilities address the capabilities present among the individuals conducting the data-collection project.
Attributes: Data literacy; Experience in data collection; Network in the humanitarian sector; Communication skills

03 Organizational capabilities

The organizational capabilities address the capabilities present among the organizations that collaborate to conduct the data-collection project.
Attributes: Organizational structure; Experience in data management; Willingness to share; Data structures and policy

04 Disaster level of difficulty

The disaster level of difficulty addresses the difficulty that a certain type of disaster imposes on the data-collection process.
Attributes: Speed of the onset; Length of forewarning; Scale & magnitude of the impact; Frequency of the event

05 Level of planning

The level of planning concerns the quality of the planning for data collection. This differentiates from the level of preparation through the focus on the planning of the actual data-collection project.
Attributes: Sampling design; Data-collection planning; Quality requirements and measurements

---

9   Muller, A. (2021). "Towards a Framework for Humanitarian Data Collection Costs."

06 Data level of difficulty

The data level of difficulty addresses the difficulties caused by certain data requirements and characteristics.

Attributes: *Data granularity; Data availability; Data quality; Available data structures*

07 Humanitarian data-collection costs

The humanitarian data-collection costs are the dependent variable of the framework, which is influenced by the six prior constructs. This concerns the estimation of the data-collection project's costs. Due to the scope of this research, it does not address costs associated with the communication, transfer, analysis, interpretation, dissemination and destruction of data.

Attributes: *Primary data-collection costs; Secondary data-collection costs; Data-cleaning and processing costs*
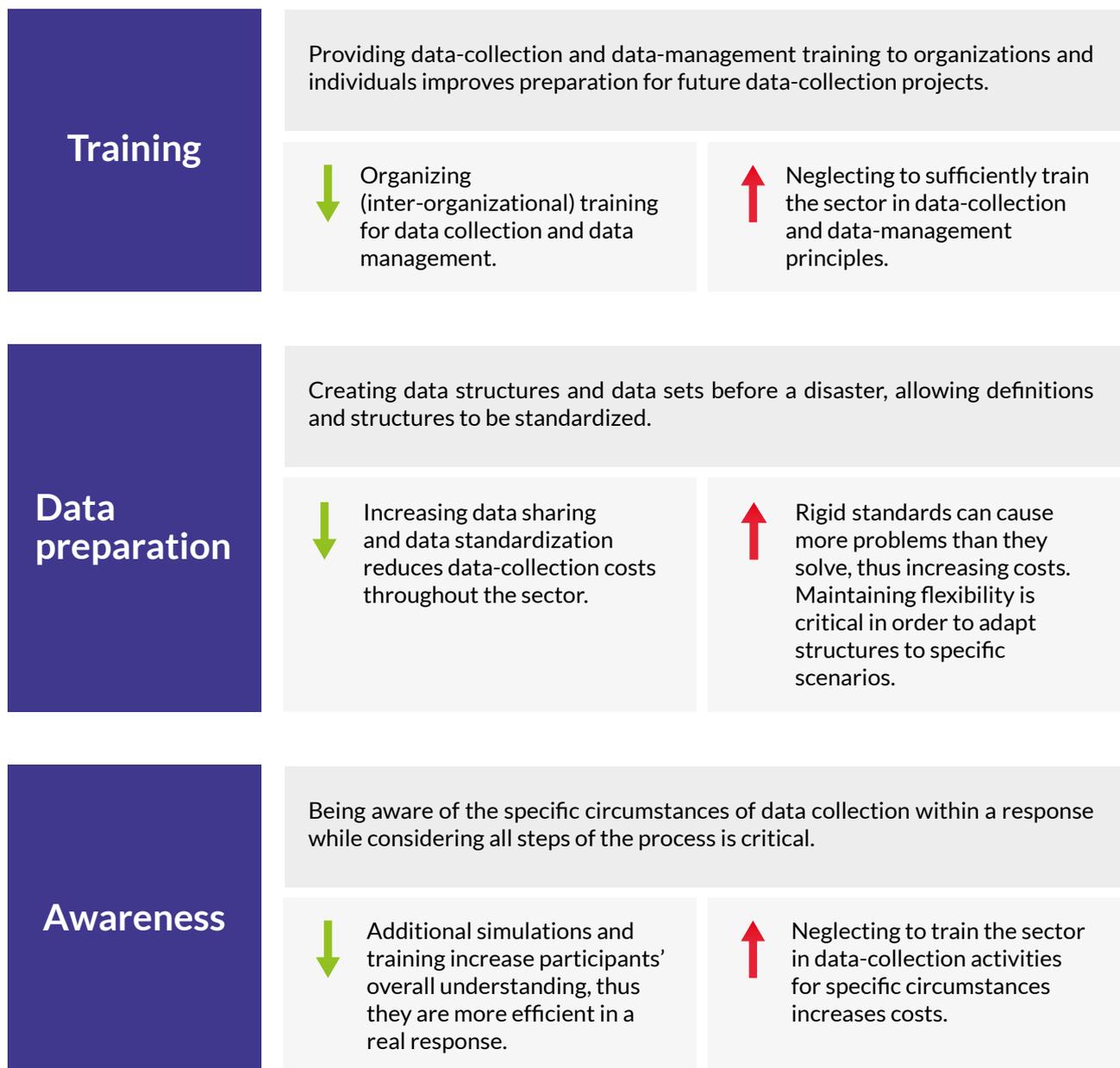
Attributes and how to influence them

Level of Preparation; **Attributes**

The level of preparation refers to the amount of preparedness in place before a disaster. It is determined from a combination of individual, organizational and data perspectives.

Training	Providing data-collection and data-management training to organizations and individuals improves preparation for future data-collection projects.	
	⬇ Organizing (inter-organizational) training for data collection and data management.	⬆ Neglecting to sufficiently train the sector in data-collection and data-management principles.

Data preparation	Creating data structures and data sets before a disaster, allowing definitions and structures to be standardized.	
	⬇ Increasing data sharing and data standardization reduces data-collection costs throughout the sector.	⬆ Rigid standards can cause more problems than they solve, thus increasing costs. Maintaining flexibility is critical in order to adapt structures to specific scenarios.

Awareness	Being aware of the specific circumstances of data collection within a response while considering all steps of the process is critical.	
	⬇ Additional simulations and training increase participants' overall understanding, thus they are more efficient in a real response.	⬆ Neglecting to train the sector in data-collection activities for specific circumstances increases costs.

Individual's Capabilities; Attributes

The individual's capabilities address the capabilities present among the individuals conducting the data-collection project.

Data literacy

Higher data literacy improves the efficiency of data-collection projects and reduces any on-site training.

⬇ Providing data-literacy training.

⬆ No training leads to mistakes, inefficiencies and/or on-site training.

Network in the humanitarian sector

An in-place network of data-related individuals and organizations will lead to a more open culture and increase data sharing.

⬇ A network helps to increase transparency and decrease redundancy in data collection.

⬆ No pre-existing network increases costs, as establishing such a network during a response requires extra time and resources.

Communication

Individuals' communication skills increase the value of their network, lead to a more transparent and direct culture, and increase inter-organizational collaboration.

⬇ Good communication skills ease collaboration.

⬆ Inability to clearly communicate and share data increases data-collection costs.

Experience in data collection

Experience in data collection concerns the knowledge and experience acquired during prior data-collection projects in the humanitarian sector (simulation and emergencies).

⬇ More experience in data collection under humanitarian circumstances increases efficiency of data collection.

⬆ Less experience decreases efficiency resulting in poorer quality or slower results.

Organizational Capabilities; **Attributes**

The organizational capabilities address the capabilities present among the organizations that collaborate to conduct the data-collection project.

Organization structure

An organization's structure (agile vs. bureaucratic) influences the efficiency to operate in environments of rapid change and extreme (time) pressure.

↓ Transforming the organization to become agile increases the ability to operate during crisis situations.

↑ Bureaucratic cultures within organizations hamper the operational ability during rapidly changing situations.

Experience in data management

Experience in data management increases the efficiency in collecting, storing, processing and cleaning data for the purpose of data sharing and decision-making.

↓ Including trained or experienced individuals in data management will reduce overall costs.

↑ Inexperience will lead to higher costs due to inefficiency.

Willingness to share

The willingness to share data impacts the entire sector's efficiency to respond to a humanitarian crisis and conduct data collection. If there is no communication or overview of which data has been collected, redundancy in data collection rises, decreasing the efficiency of the response.

↓ Increasing the willingness to share data reduces costs.

↑ Lack of willingness to share data increases costs.

Data structures and policies

Data structures and policies increase the efficiency of data sharing, thereby reducing data-collection redundancy throughout the humanitarian sector.

↓ Increasing data structures and policies reduces costs of data cleaning and processing.

↑ Rigid data structures can lead to an inability to adjust to context requirements.

Disaster Level of Difficulty; Attributes

The data level of difficulty addresses the difficulties caused by certain data requirements and characteristics.

Speed of the onset

A sudden-onset emergency generally increases the difficulty of primary data collection. Some can be foreseen (e.g. major storm) and others cannot (e.g. major earthquake).

↓ An increased speed of onset increases primary data-collection costs.

↑ An increased speed of onset decreases the level of planning.

Length of the forewarning

The amount of forewarning determines the preparation of data. Shorter forewarning prevents Governments, organizations and locals from protecting and preparing themselves.

↓ An increase in the length of forewarning reduces primary data-collection costs.

↑ Longer forewarning increases primary data-collection costs.

Scale & magnitude of the impact

A higher scale and magnitude of the impact affects resources and communication, which leads to difficulties for primary data collection.

↓ If the scale and magnitude of the impact are lower, less operational costs are required for the primary data collection.

↑ A higher scale and magnitude of the impact affects more resources, increasing the primary data-collection costs.

Frequency of the impact

A higher frequency of an event in a certain region can lead to the emergence of data structures and procedures to conduct the data collection in collaboration with the community.

↓ If the frequency of an event in a certain region increases, this generally enhances the response capabilities.

↑ If organizations do not apply lessons learned from each instance of an event, cost will continue to be higher than necessary.

Level of Planning; Attributes

The level of planning concerns the quality of the planning for data collection. This differentiates from the level of preparation through the focus on the planning of the actual data-collection project.

Sampling design

It is important to understand the techniques by which an assessment sample is created (e.g. to deal with the inaccessibility of certain regions), as their consequences may affect how data can or cannot be used.

⬇ Adjusting the sample design to a situation reduces costs and improves results.

⬆ An incorrect sample design could result in unusable data for important decisions.

Data-collection planning

The preparation of resources and personnel required for a data-collection project.

⬇ Increasing the planning beforehand will smoothen the data-collection process.

⬆ Does the data need to be perfect? People often aim for perfection, thus creating high cost and low success. What is good enough?

Quality requirements and measurement

Clearly define the required data quality, granularity and accuracy before the data collection. This results in the collection of appropriate data (i.e. not too much and not too little).

⬇ Matching the requirements will reduce data-collection redundancy.

⬆ If the quality and granularity collected are higher than needed, unnecessary costs have been incurred.

Data Level of Difficulty; Attributes

The disaster level of difficulty addresses the difficulty that a certain type of disaster imposes on the data-collection process.

Data granularity

The level of detail obtained within a data set. Depending on the data's goal, a certain level of detail is required. This is higher for actionable data and lower for strategic data.

↓ Collecting the appropriate amount of data avoids unnecessary work and resources.

↑ If the level of granularity is lower than required, critical data is missing thus generating bigger problems (and future costs).

Data availability

The data already collected by other organizations or third parties. If relevant data can be accessed or is openly available, an organization can avoid the data-collection process.

↓ Increasing transparency and data sharing reduces data-collection costs.

↑ Not searching for available data can lead to redundant data collection.

Data quality

The timeliness, accuracy, completeness and consistency of data. To prevent reconducting data-collection projects, quality needs to fit the purpose (rather higher quality than lower).

↓ Increasing data quality can increase the value of the data for analysis.

↑ Missing data may force the entire project to be reconducted. Yet aiming for perfection may be too much.

Available data structures

If data structures are defined and available, the costs of data cleaning and processing can be reduced.

↓ Higher standardization enables easier integration of data and reduces costs.

↑ Flexibility may be required, but this can increase costs significantly and even beyond its perceived value.

Humanitarian Data-Collection Cost; **Attributes**

The humanitarian data-collection costs are the dependent variable of the framework, which is influenced by the six prior constructs. This concerns the estimation of the data-collection project's costs.

Primary data-collection

Primary data-collection costs are related to creating new data. These costs are mainly influenced by the disaster's level of difficulty.

⬇ The higher the volume and detail required in the primary data, the higher the collection costs. If volume and detail can be reduced to an appropriate level, the costs can be minimized.

⬆ Limited or poor preparedness (organizational and individual) can increase primary data costs.

Secondary data-collection

Secondary data-collection costs are related to the collecting and acquiring previously produced/collected data.

⬇ Better planning and in-place technology and processes can help reduce costs.

⬆ Trying to collect too much for the real needs can result in data overload and limited value.

Data-cleaning & data-processing costs

Work required to clean and combine data to enable data analysis on the data set(s).

⬇ Higher volumes and more complicated data increase the costs of data processing and cleaning. Costs can be reduced by reducing the data complexity required, increasing the data standards used and implementing standard tools.

⬆ Poor organization planning and low data capabilities will increase costs.
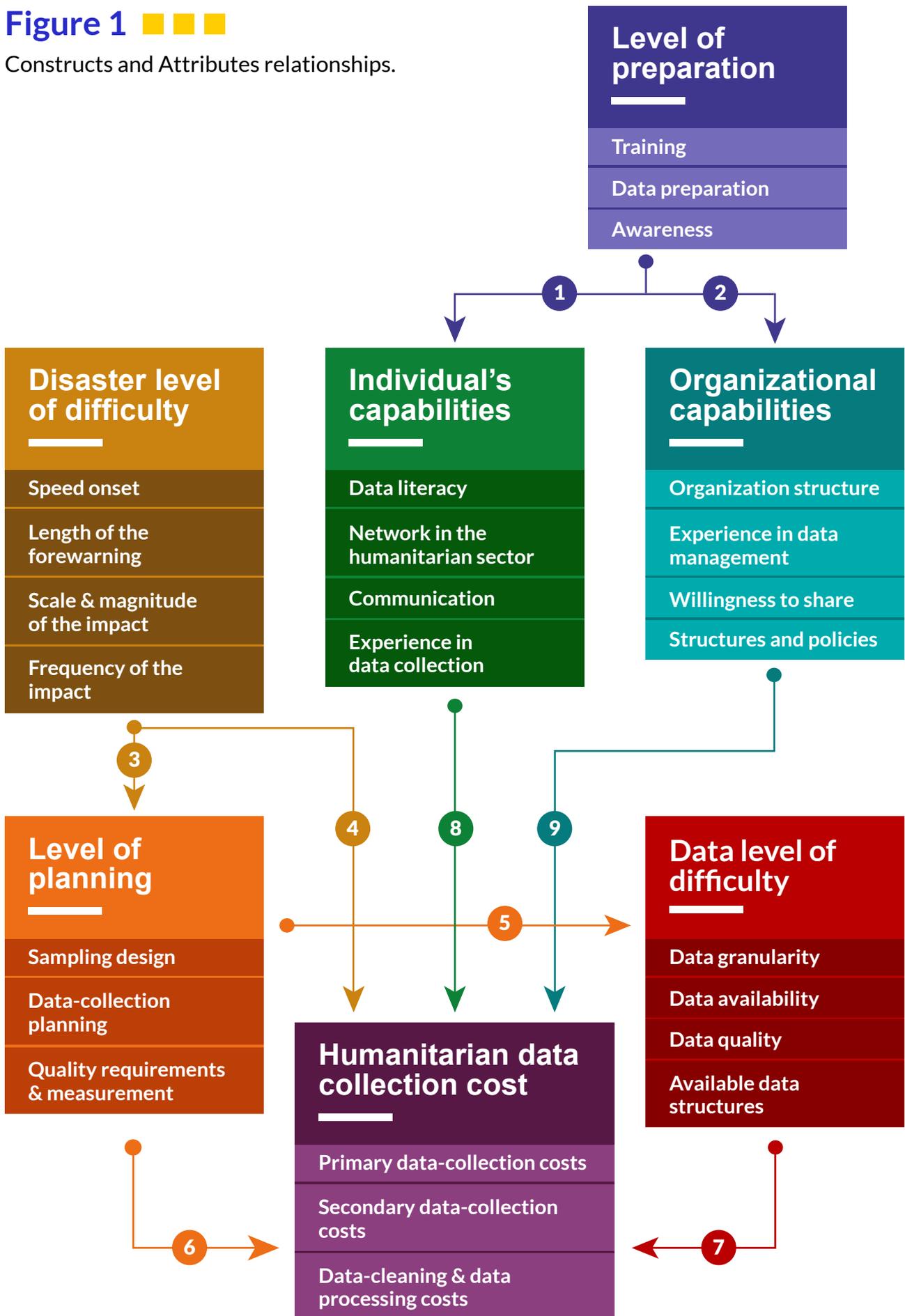
Relationships: Cause and Effect

The table below outlines the relationships between constructs - how one can effect the other. A visual representation can be seen in **Figure 1**.

Influencing Construct	Effected Construct	Cause	Effect
1. Level of preparation	Individual's capabilities	Training on data-collection activities and awareness of circumstances.	Improves the in-place networks as well as the community's communication and data-collection skills.
2. Level of preparation	Organizational capabilities	Data preparation and training on effective data sharing.	Increases available data (structures) and data-sharing efficiency.
3. Disaster level of difficulty	Level of planning	Challenging characteristics of a disaster (e.g. rapid onset, political challenges).	Impacts possibilities to conduct data-collection planning.
4. Disaster level of difficulty	Primary data-collection costs	Challenging characteristics of a disaster (e.g. geographical size and impact).	Impacts infrastructure and communication networks.
5. Level of planning	Data level of difficulty	Data-collection planning, setting quality requirements and deciding on a sampling design.	Enhanced insight into the required data quality, data granularity and available data.
6. Level of planning	Humanitarian data-collection costs	Data-collection planning, setting quality requirements and deciding on a sampling design.	Structured approach to data collection, mitigating difficulties (e.g. access restrictions).
7. Data level of difficulty	Humanitarian data-collection costs	Alignment of data requirements with the goal of the data-collection project.	Reduces unnecessary efforts for high data quality and granularity.
8. Individual's capabilities	Humanitarian data-collection costs	Increased data literacy, communication skills, experience and in-place networks.	Reduces on-site training thus improving data-collection efficiency.
9. Organizational capabilities	Humanitarian data-collection costs	Improved willingness to share information, organizational and data structures and policies.	Increases efficiency of data collection and data sharing.

Figure 1 ■ ■ ■

Constructs and Attributes relationships.

Level of preparation

- Training
- Data preparation
- Awareness

Disaster level of difficulty

- Speed onset
- Length of the forewarning
- Scale & magnitude of the impact
- Frequency of the impact

Individual's capabilities

- Data literacy
- Network in the humanitarian sector
- Communication
- Experience in data collection

Organizational capabilities

- Organization structure
- Experience in data management
- Willingness to share
- Structures and policies

Level of planning

- Sampling design
- Data-collection planning
- Quality requirements & measurement

Humanitarian data collection cost

- Primary data-collection costs
- Secondary data-collection costs
- Data-cleaning & data processing costs

Data level of difficulty

- Data granularity
- Data availability
- Data quality
- Available data structures

Rapid Calculation ■ ■ ■

Decisions about collecting data and its related costs often have to be made quickly. By leveraging the above constructs and ideally understanding their respective attributes (below), one can quickly develop a rough estimate. Working through the below questionnaire and then assigning rough estimates will give you a good estimate of the data-collection costs. From there, one can consider the construct relationships (below), thereby understanding how you may be able to influence the data-collection costs and ensure they are within your financial expectations.

Question	Labor Costs	Operational Costs	Total Costs
01 Is this an infrequent disaster for the region with a high scope and magnitude?			
02 Is the disaster a (rapid) onset with a low level of forewarning?			
03 Is accessibility to the affected regions by car or foot limited (due to the disaster)?			
04 To what extent can sampling design solve the issues related to inaccessibility of regions?			
05 Have data preparation and training in data collection and awareness occured for this type of disaster?			
06 To what extent can data be shared through the current existing networks?			
07 To what extent are data literacy, communication and experience in data collection already in place?			
08 To what extent is there a willingness to share data among the involved organizations?			
09 To what extent are data structures and policies in place within your organization to deal with this data collection?			
10 To what extent can data collection planning be conducted for this data collection project?			
11 To what degree do the data collection goals add complexity (e.g. granularity & quality)?			
12 To what extent do the data availability and data structures available mitigate this difficulty?			
13 Are there any additional difficulties to primary data collection? (eg. security, etc.)			
14 Are there any additional difficulties to secondary data collection?			
15 Are there any additional difficulties to processing and cleaning of the data?			
16 Are there any additional difficulties related to data security and privacy?			

Example ▪ ▪ ▪

Disaster Background

! **Before answering the questions below, it is important to take the effects of specific circumstances into account. Examples of such circumstances are listed below:**

- **Is the disaster type common in the region?** ➡ Are previous data collection plans and protocols available? Are evaluation reports and lessons learned available to organizations and local communities?

- **Is sufficient planning hampered by the sudden onset nature of the disaster or by a low level of forewarning?** ➡ Is it possible to establish the data collection planning, quality requirements and sampling design?

- **The level of effort to respond to the given disaster?** ➡ What kind of disaster does it concern? (e.g. major earthquake requiring significant needs assessment activities)

Fictitious disaster: A massive earthquake strikes an area generally not prone to large earthquakes severely damaging all infrastructure. Communication is not available in the early hours/days and movement is restricted due to infrastructure damage and ongoing fighting between the government and rebels. Collecting personal information is extremely sensitive as it could be perceived by waring factions as people aligning with the other side.

Customized questions	Labor Costs	Operational Costs	Total Costs
Cost to access disaster affected area (e.g. no land access so planes/helicopters required) [1]	€ 50.000,00	€ 200.000,00	€ 250.000,00
What is the goal of the data collection project? (e.g. strategic or actionable) What corresponding levels of data quality and granularity are involved? [2]	€ 50.000,00	€ 25.000,00	€ 75.000,00
Is onsight training of personnel in data collection required? [3]	€ 20.000,00	€ 20.000,00	€ 40.000,00
Has data preparation taken place? Or is it required to "start from skretch"? [4]	€ 20.000,00	€ 10.000,00	€ 30.000,00
Is an in-place network of data-related individuals and organizations available, or is this to be established during the response? [5]		€ 25.000,00	€ 25.000,00
What effort is required to acquire/receive information from other organizations? [6]		€ 10.000,00	€ 10.000,00
Is the data collection project subject to additional costs due to the specific circumstances? (e.g. security or others) [7]	€ 5.000,00	€ 20.000,00	€ 25.000,00
What are the expected additional costs to confirm to sufficient data security and data privacy levels? [8]	€ 1.000,00	€ 10.000,00	€ 11.000,00
TOTALS	**€ 146.000,00**	**€ 320.000,00**	**€ 466.000,00**

Comments

1. Helicopters are required for transportation, leading to costs for helicopters and pilots.
2. Goal of the data collection project is to acquire actionable insights from the collected data.
3. Personnel will require onsight training
4. No data preparation has taken place, required to start data structures from skretch.
5. No in-place network was available to structure communication and data sharing.
6. Unfortunately, hardly any data is available to be shared, thus not a lot of costs are concerned with acquiring/receiving data.
7. Security is required due to circumstances in the affected country and regions.
8. Additional costs are involved to ensure sufficient levels of data security and privacy due to sensitive information.

Reduce costs. Increase data impact.

■ ■ ■

Data-collection preparation

Preparation is one of the most important factors to influence data-collection costs, since organizations can manipulate it without external factors impacting the outcome. Increasing an organization's preparation efforts will improve individual and organizational capabilities, thereby reducing project costs. Therefore, it is strongly recommended to increase data preparation for disasters, increase (inter-) organizational training in data collection and data management and increase the understanding of the unique aspects within a collection project. [10|11]

Data sharing

Reducing the costs of humanitarian data-collection projects requires improving the efficiency of inter-organizational collaboration. Organizations often collect data without asking if it has already been collected. As mentioned in OCHA's report *The State of Open Humanitarian Data 2021*, [12] the Humanitarian Data Exchange has greatly increased data availability within the sector.

Implementing standard data structures will increase preparation and collaboration between organizations. As open data becomes the norm among humanitarian organizations, data structures will be a key factor to help facilitate data exchange and usage. This change, in turn, decreases the redundancy in data collection and reduces overall collection costs.

Wider data-related concerns

■ ■ ■

Data analysis capacity

Data analysis was discussed during the majority of interviews. There is an obvious mismatch between the capacity to collect and the capacity to analyse data in the humanitarian sector.

"Currently, roughly 20 per cent of the collected data is analysed. There is no check on the balance in terms of production and analysis, compared to the capacity to analyse. We collect too much data. A dashboard, for example, is not data analysis. It shows results but it does not inform on the findings from the results. Without such a conclusion, the data is not actionable."

Yves Kim Creach,
The Operations Partnership, Associate Director.

All interviewees agreed that data collection provides tremendous value to humanitarian response. Yet the potential value is not always achieved due to this capacity mismatch. To optimize the value of the collected data, any project must consider increasing its data analysis capacity.

Data privacy and security

Data privacy and security are often neglected in the humanitarian sector. The interviewees note that these have improved over time but still require additional attention to achieve sufficient levels.

"We have a long way to go with the security and privacy of data. Ninety per cent of humanitarian work happens in Excel or Google sheets. For some organizations, private data is shared publicly through Google sheets daily. People haven't been taught that they need to be cautious about how they share data."

Gisli Olafsson,
One Acre Fund, Strategy Technology Officer.

It is vital to protect the privacy of the vulnerable people and communities from whom data is collected. This requirement increases costs. The Inter-Agency Standing Committee (IASC) Operational Guidance on Data Responsibility [13] provides a road map and structure for organizations to implement higher data responsibilities throughout their activities.

Further improving the calculation

■ ■ ■

Combining the framework and the rapid cost-assessment questionnaires enables decision makers to estimate the cost of a data-collection project. Moreover, it outlines the main areas where improvements or adjustments can reduce costs. To improve the accuracy of the cost calculation, organizations can expand the calculation to include the entire collection cycle and/or develop a scientifically sound mathematical model.

Entire data-collection cycle

The scope of this paper only addresses planning, data collection, data storage, and data processing and cleaning within the information management cycle. Therefore, it excludes aspects relating to data transfer, analysis, communication and dissemination, feedback and evaluation, and data retention and destruction. These areas should be considered for a more complete picture of an organization's data costs.

A more mathematical approach

The framework and rapid cost-assessment questionnaire above provide an indicative estimate. They do not *calculate* a truly accurate cost. To increase the accuracy, a mathematical formula would need to be used. According to the research by Muller [14], the COCOMO II [15] and Time-Driven Activity-Based Costing [16] methodologies would be strong candidates. Organizations would need to select a model and invest resources to appropriately adopt the model for their unique situation.

---

13 Centre for Humanitarian Data (2021). "IASC Operational Guidance On Data Responsibility In Humanitarian Action." https://centre.humdata.org/iasc-operational-guidance-on-data-responsibility-in-humanitarian-action/

14 Muller, A. (2021). "Towards a Framework for Humanitarian Data Collection Costs."

15 Boehm, B., B. Clark, E. Horowitz, C. Westland, R. Madachy, and R. Selby (1995), "Cost Models for Future Software Life-cycle Processes: COCOMO II," Annals of Software Engineering 1, Special Volume on Software Process and Product Measurement, J.D. Arthur and S.M. Henry, Eds., 45–60.

16 Kaplan, R. S., & Anderson, S. R. (2007). "Time-driven activity-based costing: a simpler and more powerful path to higher profits." *Harvard Business Press.*