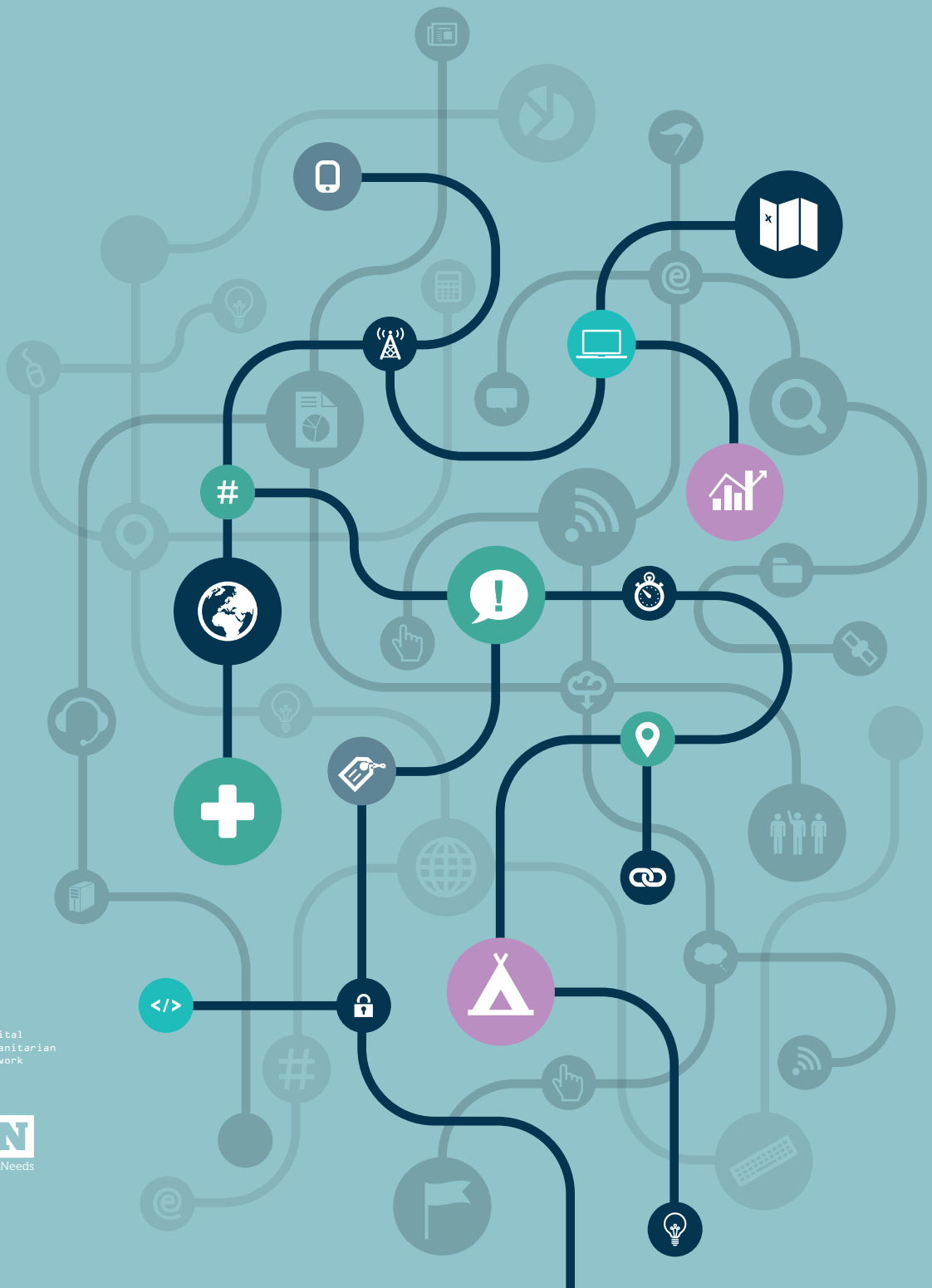
Guidance for Incorporating

BIG DATA INTO HUMANITARIAN OPERATIONS



digital humanitarian network

DM-N
Decision Makers Needs

Licensing Information

Guidance for Incorporating

BIG DATA INTO HUMANITARIAN OPERATIONS

by

Katie Whipkey (katiewhipkey@gmail.com)
MSPPM '16 Carnegie Mellon University Heinz College

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

Design

Ignacio G. Rebollo (igrebollo.arq@gmail.com)
M.A. '15 CICE - Professional School of New Technologies
B.S. '12 IE University - School of Architecture

With the support of UN-OCHA, this document was created as part of the Digital Humanitarian Network's community of interest on Decision Makers Needs.

DM-N Decision Makers Needs

DH digital humanitarian network

We would like to acknowledge and thank the individuals from the humanitarian organisations, V&TCs, and academic community who provided their experiential knowledge, insight, and advice through interviews and correspondence on the drafts of this document.

Those contributors include:

Hayyaan Ahmad
B.S. '15 Lahore University of Management Sciences

Helen Campbell
Assessments Capacities Project

Tim Clark
Humanitarian UAV Network

Robert Colombo
United Nations Office for the Coordination of Humanitarian Affairs

Martina Comes
Universitetet I Agder Centre for Integrated Emergency Management

Patrick Gordon
United Nations Office for the Coordination of Humanitarian Affairs

Patrick Meier
Qatar Computing Research Institute

Jeffrey Villaveces
United Nations Office for the Coordination of Humanitarian Affairs

This guidance is intended to assist information or data focal points in humanitarian organisations understand the variety of, the categories of, and possible approaches to integrating big data into their organisations. Prior to this guidance, the authors often heard either promises from data organisations like "*big data will solve your big problems*" or disparaging comments from humanitarians such as "*I have no idea what big data actually is*" or "*I have no idea how to use or integrate it*". This guidance aims to help bridge the gap between these two extremes.

-The Authors,
September 2015

CONTENTS

EXECUTIVE SUMMARY 1

PART ONE | **THE BASICS FOR UNDERSTANDING BIG DATA**

A. What is Big Data 3
B. A Brief History of Big Data 3
C. Types of Big Data 4

PART TWO | **INCLUDING BIG DATA IN HUMANITARIAN RESPONSE**

A. The Benefits 7
B. The Challenges 10

PART THREE | **GUIDANCE FOR INCORPORATING BIG DATA INTO HUMANITARIAN RESPONSE**

A. Internal Considerations 14
B. Policy Considerations 15
C. Humanitarian Big Data Wheel 16
D. Acquisition of Big Data 18
E. Using "Big Data" and "Traditional Data" 19
F. Contextual Considerations 19
G. Timeline and Lifespan of Big Data 21
H. Conclusion 21

THE ANNEX

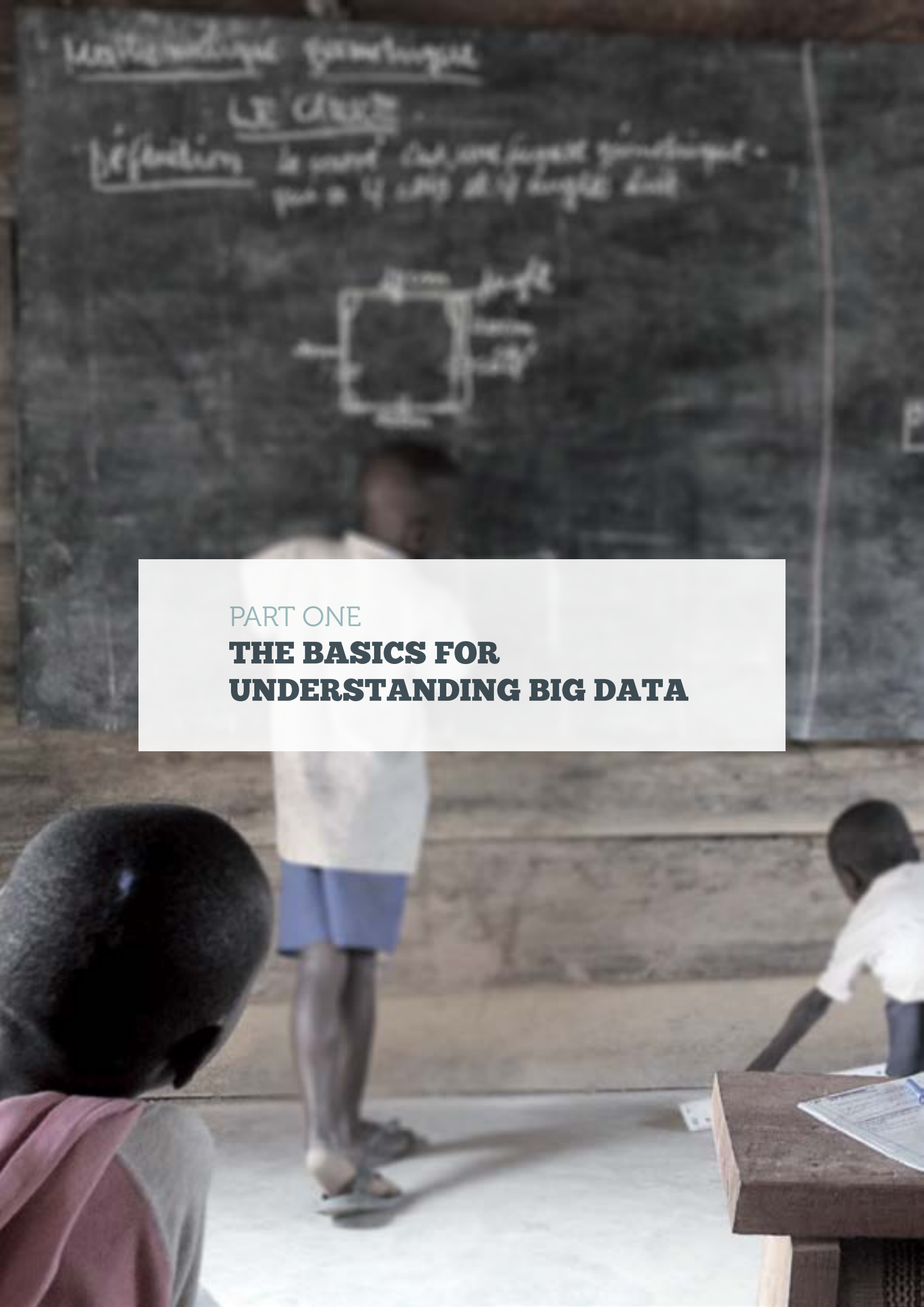
A. Questions to Consider 23
B. Checklist of Risks 27
C. Open Data Sources 29
D. Matrix of Products (Example) 30
E. Decision Makers Needs 31
F. References 32

EXECUTIVE SUMMARY

Big data in humanitarian response is becoming more prominent and important. We are at the onset of a data tsunami, a "big data revolution" as some would say, as data is being produced at an exponential rate.[1] [2] The time and resource constraints of humanitarian organisations and responders arebecoming increasingly stretched with the growing number and length of humanitarian crises.[3] This reality, in combination with the vulnerability of the affected populations that are often the subjects of the data, makes many organisations resist incorporating big data practices into response. Despite these risks, big data is a growing wealth of critical information and its use is rapidly rising in popularity. However, it is not clear how field managers can incorporate this new and big data into operations. As a result, it has been difficult for many humanitarian organisations and decision makers to effectively utilise big data in humanitarian service delivery.

Humanitarian organisations should begin the process now of understanding what big data is, what its analysis can yield answers to, and how and when to use it. This document has been created to provide a broad overview of big data usage in humanitarian organisations and general guidance on how organisations can incorporate it into operations. It describes big data and its role within the humanitarian sector, offers a categorisation of the large variety of big data types, highlights benefits and risks of incorporating big data into response, identifies policy and ethical considerations for the organisation, and provides example materials organisations can use when starting the process of incorporating big data. The goal is to create dialogue and generate structure in the conversation among decision makers, data scientists, and volunteers and technical communities.

This is intended for humanitarian response organisations across the globe that are looking for new and innovative ways to combat difficulties crises. It is aimed at organisations that do not know how to use big data at all and/or those that are looking to use it more strategically. It can also be useful to development agencies, policy makers, or other international organisations interested in implementing big data analysis into their business model. The document is not designed to be a manual for the detailed intricacies of data science. It should be viewed from the decision maker's level to understand how to work with data scientists.



PART ONE

THE BASICS FOR UNDERSTANDING BIG DATA

WHAT IS BIG DATA?

The definition of "big data" is subjective and fluid, as researchers have yet to agree upon a singular concept and definition. In general terms, big data refers to large and/or complex datasets that go beyond the capabilities of people and traditional, desktop software to capture, store, manage, and analyze in their entirety.[2] A size or amount has not been assigned to quantify what qualifies as big data because the capabilities of technology are rapidly changing over time and capacities vary across sectors and locations.[2] Also, due to the differing concerns of scientists, technological entrepreneurs, researchers, data analysts, and practitioners, the definition of big data varies.[4] In practical terms and for the purpose of humanitarian response, big data can be understood as a large amount of data coming from any source(s) that can be collected and analysed to inform decision making.

The Four V's

One of the most widely accepted ways to describe big data is the "Three V's" (volume, variety, and velocity), though the "Four V's" have been developed to more accurately reflect the value of big data. The International Data Corporation wrote a report in 2011 that defined the concept in their effort to characterize big data. The four V's represent "Volume (great volume), Variety (various modalities), Velocity (rapid generation), and Value (huge value but very low density)".[4] These characteristics of big data build our understanding that it is very large amounts of data gathered from many sources at a high speed that have the capacity to be extremely useful to its end users.

Traditional Data

Traditional (or "classical, offline, or small data") data mining is the counterpart to real-time, large-scale data mining, which is generally synonymous with big data. Traditional data mining is learning from historical data by collecting, studying, deducing results, or making decisions based upon the information collected from the past.[5] It often contains information that cannot be obtained by big data such as historical information or personal surveying.

BRIEF HISTORY OF BIG DATA

The first time the term "big data" appeared was in a NASA publication in 1997. It was described as the problem of data not fitting in the core processing memory or external disks.[6] Over ten years later in 2008, the term was popularized by computer scientists from the Computing Community Consortium when they predicted that

"big data computing" would transform processing across the spectrum from military to business.[7] However, the concept of big data goes back much earlier as the first attempt at defining the idea was in 1941 when the Oxford English Dictionary defined "information explosion".[8] Many believe that since big data still does not have a consistent and agreed upon definition by scholars, the term may fade out of existence but the concept will remain as it changes along with new technologies that emerge.

> "Approximately 90% of the world's data in 2014 was created since 2012."

Today, using big data is becoming increasingly commonplace. Approximately 90% of the world's data in 2014 was created since 2012 and it is estimated to be growing at a rate of 2.5 quintillion bytes per day, or approximately 1.7 million billion bytes per minute.[9] [10] [11] It is used across nearly every industry, including humanitarian agencies. It's a highly valuable resource and has a large impact economically due to its potential to increase effectiveness and efficiency.

One of the earliest uses of big data in humanitarianism was in 2007 when Ushahidi used crowdsourcing to map instances of violence in Kenya after the elections.[12] The 2010 earthquake in Haiti was the first major use of big data when the public sector partnered with Digicel, Haiti's largest mobile phone company, to compare people's movements pre- and post-earthquake. This enabled responders to locate where most people fled and deliver medical and relief supplies more efficiently.[1] [21] Since then, the use of big data in humanitarian response has been growing rapidly from being integrated into traditional response to the formation of digital volunteer communities and data organisations and labs.

TYPES OF BIG DATA

Big data can be understood through two main streams: **streaming and batch**.[13] Streaming data is real-time, constantly coming in and refreshing itself. It enables us to collect information about what people like, consume, think, feel, where they go, and what they do, directly from the source. It is difficult to analyse because it is unpredictable and difficult to manipulate. Examples include social media and closed circuit security cameras. Generally, streaming data is converted into batch data once it is received. Batch data, on the other hand, is data that is collected in a set or group, one batch at a time. It can be historic or current, but it is understood to be static once received. Government records are examples of batch data. Some data sources can be either streaming or batch depending on how and when they are collected. For example, historic internet search records can be collected as batch data, but any searches that are performed from the moment that an organisation decides to start collecting that data can be processed as streaming.

The chart below has divided big data into eight categories. Some of the categories are related,

such as user-generated and social media, but they are designed to serve as a catalyst for expanding ideas about different types of big data that can be used. The chart is not designed to be completely comprehensive and organisations will not be using every type of data listed. It can be used to help organisations determine how to use these types of data separately and/or together to develop the most appropriate analysis and response.

User-Generated	Social Media	Captured*	Business Systems
• *Blogs and Comments*	• *Facebook*	• *Banner Ad Clicks*	• *Business Identification Number*
• *Email*	• *Google+*	• *Bluetooth Connections*	• *Driver's License*
• *Messaging (WhatsApp, Skype, etc.)*	• *Instagram*	• *Customer Loyalty Cards*	• *News Articles*
• *Mobile App Data*	• *LinkedIn*	• *Customer Surveys*	• *Passport Tracking*
• *Personal Documents*	• *Pinterest*	• *Focus Groups*	• *Radio Transmissions*
• *Photo Sharing (Flickr, etc.)*	• *Private/Custom Networks*	• *Internet Searches*	• *Rental Records*
• *Rating Systems*	• *Skype*	• *Mobile Phone Tracking*	• *School Records*
• *SMS*	• *Tumblr*	• *Security Cameras*	• *Tax Information*
• *User-Generated Maps*	• *Twitter*	• *UAVs*	• *TV Transmissions*
	• *Vine*	• *Vehicle GPS*	• *Trade Records*
	• *WeChat*	• *Web Logs*	• *Transportation Logs*
		• *Website Click-Throughs*	• *Voting Records*

Transactional	Sensors	Biological	Public Records
• *Banking Records*	• *Home*	• *Dental Records*	• *Birth Certificates*
• *Commercial Trans.*	• *Mobile Phones*	• *DNA Testing*	• *Census Data*
• *Credit Card History*	• *Pollution*	• *Facial Recognition*	• *Death Certificates*
• *Credit Scores*	• *Satellites*	• *Fingerprinting*	• *Land Titles*
• *E-Commerce*	• *Scientific*	• *Medical History*	• *Marriage Certificates*
• *Mobile Cash Trans.*	• *Traffic*		• *Military Records*
• *Stock Trading*	• *Weather*		• *Police Records*
• *Swipe Card Aid Transactions*			• *Vehicle Registration*

**Captured data refers to data that users can choose to elect into our out of, whereas business systems data is data that collected as a required practice without a choice by the user.*

PART TWO

INCLUDING BIG DATA IN HUMANITARIAN RESPONSE

Big data can be an essential tool in humanitarian response for organisations and governments because it allows responders to develop insights into humanitarian trends. By utilising large datasets such as telephone records, sensors, social media, and others in combination with traditional datasets, big data can enable deeper analysis into the humanitarian system.[14] It can help identify when and where people are in need of relief or are in danger. Resources can be allocated more efficiently and effectively with this knowledge in hand.[12] In order to use big data in humanitarian response, it first must be understood that there is no such thing as humanitarian data; rather, there are humanitarian questions that big data can answer. Three such humanitarian questions are the following:[14]

1. Where are the affected populations; where did they originate and where are they going?
2. What do the affected populations need?
3. What are the gaps that need addressed?

These broad humanitarian questions can help determine the appropriate indicators. Once indicators are identified, specific proxies can be selected. Organisations can then decide how big data can be best utilised to give value to the indicators and proxies identified.

The table on the following page illustrates questions to consider, examples of big data sources that can be used to answer the questions, and an example of how big data has been used in the past to address the question. When an organisation uses the table, it should be modified and used through the lens of the crisis itself. Considerations should be specified for the country context and type of disaster and sources should reflect information available and relevant for the context.

THE BENEFITS

Utilising big data in humanitarian response has many benefits for the effectiveness and efficiency of delivering services to the affected population. It brings in more real time information, which builds new insights into the decision making process. It can also be used to shape messages, make adjustments to plans, and measure impact more quickly.

Real Time Information

Organisations gain access to more real time data by incorporating a big data plan into its operations. Big data includes information from sources that report in real time, which is otherwise unattainable from traditional sources. This results in a higher quality of information produced because it is based on real-time analytics.[5] It eliminates the need to store historical data for an extended period of time in order to analyse it at a later date.[5] Organisations can use real time technologies that generate data as well, such as new social media applications or UAVs, to send live videos of cases that require immediate attention. This also has a policy implication as policy decisions are historically shaped on data that is two-to-three years old, so these decisions can now be made with current data instead.[11]

Question	Considerations	Sources of (Big) Data
Where are the impacted populations?	- Where have populations moved? - What proportion have moved and at what rate? - Why are they moving?	Email, SMS, messaging apps, photo sharing, video sharing, social media, user-generated maps, UAVs, mobile phone tracking, mobile apps, vehicle GPS, security cameras, news articles, passport tracking, radio and television transmissions, commercial transactions, credit card history, traffic sensors, satellites, facial recognition, census data.

[Example 1] *The 2010 Haiti earthquake was the first time Flowminder, in conjunction with Karolinska Institute and Columbia University, was launched. Flowminder received mobile data from Haitian mobile provider Digicel from 42 days before the earthquake through 341 days after. Using the mobile phone towers and SIM card tracking, researchers tracked movements of people as a result of the earthquake. It was predicted to be 85% accurate in predicting IDPs' locations.*[15]

What do the affected populations need?	- What needs are most urgent? - How quickly must the needs be met? - How can the affected population receive needed services? - How are needs changing over time?	Email, SMS, messaging apps, photo sharing, video sharing, social media, user-generated maps, blogs and comments, mobile apps, banner click ads, customer surveys, focus groups, UAVs, security cameras, weblogs, website click-throughs, news articles, radio and television transmissions, commercial transactions, credit card history, school records, banking records, home sensors, pollution sensors, weather sensors, medical history.

[Example 2] *The Digital Humanitarian Network activated to respond to Typhoon Pablo in the Philippines in 2012. As a result, the Standby Volunteer Task Force and Humanity Road created the first UN crisis map using social media data. The map gave a detailed situation report including factors such as housing, infrastructure, crop damage, population displacement, and more using metadata from Twitter. The maps were used by UN agencies and the Philippines government to address the needs of the affected population.*[16] [17]

What are the gaps that need addressed?	- Can the organisation meet the needs? - What additional resources are needed? - What other parties need to be involved?	Email, SMS, messaging apps, photo sharing, video sharing, social media, blogs and comments, user-generated maps, mobile apps, customer surveys, focus groups, UAVs, mobile phone tracking, security cameras, news articles, radio and television transmissions, school records, commercial transactions, traffic sensors, census data, transportation logs, satellites, home sensors, pollution sensors, scientific sensors, medical history, police records.

[Example 3] *The German Red Cross and the Red Cross Red Crescent Climate Centre created a financing mechanism utilising scientific and weather sensors. When extreme weather is forecasted, a warning is issued and funding is automatically released to enable quick and preventative humanitarian response. The Red Cross and participating communities have an agreed upon plan prior to the disaster and this plan can be put into place immediately upon the release of the funds, leading to a reduction in damage and suffering.*[18]

Informed Decision Making

Because big data can provide real time information from a variety of different platforms, organisations can make more informed decisions. It gives organisations the ability to make immediate adjustments and adaptations when the environment changes and the changes are detected in the data.[5] Data is often time, date, and geo-stamped so decisions can be based on how applicable they are at any moment.[4] Systems become more responsive and effective as the data informs action and action informs data.[5] The cyclical ability to refresh big data allows for a feedback mechanism to inform decision making, facilitates demand-driven assistance, and enables ongoing effectiveness evaluation. Additionally, the ability to source multiple information platforms for relevant information and get it into a common format increases reliability of information and enhances decision-making.

New Insights

Big data is by nature, big. This provides enormous opportunities and possibilities for new insights.[19]

> "Organisations can monitor sources, especially social media, to detect patterns and trends to foresee potential crises."

Having a large sample of data is more valuable than small samples that traditional data generally provides because it identifies patterns and correlations, which overpower any individual fluctuations and outliers. It can detect interactions, correlations, and irregularities amongst different variables.[19] It can also bring hidden patterns to attention that would not have been noticed otherwise and analysts and decision makers can gain a deeper understanding of them.[4 20 21] Additionally, it can assist in dis-

CASE STUDY | **PREVENTIVE USE OF BIG DATA IN HAITI** [22]

The cholera epidemic that began in 2010 in Haiti exemplified how big data could have been used to advance the humanitarian response. An analysis on news media reports through HealthMap, Twitter postings, and government-reported cases was performed of the first 100 days of the outbreak. Data scientists found that the number of officially reported cases varied in correlation with the amount of tweets and news reports. In fact, the informal sources (Twitter and HealthMap) were able to make the trend in volume available two weeks earlier. Although the data from informal sources was found to have a one day lag compared to official sources, the informal data is made publicly available in real-time as opposed to official data which is released after a delay. Using informal big data provides earlier insight into the evolution of an epidemic, which can have implications for disease control measures and ultimately save lives.

covering new models that explain how things could evolve in the future.[21] Big data facilitates the understanding of the human system at the systemic level.[19]

Preparedness

Big data can be used to help pre-empt humanitarian crises, particularly epidemics and escalating conflicts. Organisations can monitor sources, especially social media, to detect patterns and trends to foresee potential crises. Social media networks become a public forum where organisations can gain an internal perspective into the needs, thoughts, and movements of the local population. They can be used to determine how to modify service delivery or to gain insight as to how the population is responding to the relief that has already been delivered. Organisations can also use social media to spread awareness about a variety of topics, which will generate a new batch of big data. They can be used to warn people of impending dangers, notify them how and where to find services, and ask them for direct feedback. This is a part of the feedback loop of big data that organisations can use to be better prepared for incorporating big data analytics in the future.

THE CHALLENGES

Incorporating big data into humanitarian response poses a number of challenges aside from the vast amount of data to sort and compile. Humanitarians must take a calculated inventory of the risks involved through the lens of the crisis at hand to best determine if and how to use big data and what mitigating measures should be taken. The major challenge areas to consider include geography, users, validity, technology limitations, policy and ethics, and staff knowledge.

Geography

The places in the world that are the most at-risk for both natural disasters and armed conflict are often the most marginalised regions. As such, many people do not have easy access to mobile phones and the Internet, especially when an emergency strikes. This would limit the amount of post-crisis user-generated big data that can be reported and collected. Also, they may have a more vulnerable infrastructure that can be easily destroyed or interrupted. A disaster can limit the amount of resources available and that can be used, including damaging infrastructure and creating connectivity issues due to disrupted networks. Also, language can be a barrier, as not all staff will speak the varying languages in which the data may be produced. Finally, big data processing requires a robust availability of electrical energy, so the appropriate infrastructure must be in place as well as the people with electrical competencies who are ready and able to work.[4]

Users

The challenges posed by humanitarians processing the data include measurement error, decision makers considerations, selection and other sources of bias, user-errors, clarity of purpose, and lack of collaboration.[23] [24] It is critical to determine the problem that using big data will solve and how to collect information to find solutions before beginning the process of analysis. Analysis can be time consuming and expensive, especially during an emergency, so a clear

plan for thoughtful measurement, careful research design, and useful statistical techniques in place is necessary.[24] Additionally, users can produce errors in analysis due to the overwhelming nature of big data.[23] Decision makers in particular face risks since there are often unclear decision maker triggers, timelines, and response structures. Finally, the lack of coordination between organisations and amongst different sectors poses a risk.[4] There may be hesitancy to use tools between sectors, distrust of information provided by different groups or of the method by which data was filtered, duplication of efforts, and a lack of oversight across all organisations regarding what all are doing and need.[23] All of these factors may result in data not being used by the organisation, even if it was requested.[25]

Technology

There is a limit to big data's capabilities as technology continues to develop. Datasets are growing so technology must grow and keep up to date in response.[23] We are at the onset of the data revolution, with 90% of big data still unstructured, so many systems are yet to be developed and/or perfected.[10] Data scientists and programmers need to create adaptive models to be predictive and scalable, which is very complex. They also need to create filters that do not discard useful information, such as outliers, in case these outliers are actually useful information rather than faulty or bad data.[19 20] Data can come in many different formats as it is pulled from multiple sources.[23] Systems need to figure out how to determine what information is needed and establish a structure in a form that can be used and stored for analysis.[20] If the data is not in the correct format, the risk of it not actually being used by organisations is high.[25]

> "As the use of big data grows, the need for qualified professionals also rises."

Validity

Ensuring that data is valid before using it to inform actions is essential. There is a risk of error-laden data or data that misrepresents the situation. Some common errors to watch for include: data redundancy, inaccurate, faulty, and untrustworthy sources, biases, and outdated/expired information.[4 19 23] To mitigate this, data scientists must understand and model these sources of errors to develop data cleaning techniques.[20] Data should be monitored longitudinally and analysis should be replicable to ensure that patterns rather than trends are being identified.[19] As systems change, data collected can be inconsistent if it does not change along with the systems. A final danger is manipulated data through overloading of networks to make certain information come to the forefront.[19]

Policy and Ethics

One standard list of protocols and nomenclatures has yet to be developed for identifying information.[23] There is a risk of not adhering to established humanitarian ethics and principles because big data policy is not internationally understood yet. Organisations must uphold all

humanitarian principles as they navigate big data, paying special mind to data protection and privacy and beneficiary confidentiality in datasets.[4] Additionally, many governments, corporations, and organisations have not developed data policies, so humanitarians must navigate within undeveloped frameworks and face security issues with accessing data. Policy considerations will receive further elaboration later in the document.

Staff Knowledge

As the use of big data grows, the need for professionals who can manage the big data wheel effectively and responsibly also rises. The humanitarian data scientist role spans a number of disciplines including information management, humanitarian response, social sciences, statistics, programming, and data management.[26] Staff must have the hard skills required to acquire, store, transform, validate, model, and visualise the data along with the soft, managerial and social science skills to make accurate causal inferences.[24] Finding individuals that can do the data management and analytics as well as being effective and creative communicators and leaders is rare and it is not a common role that already exists within organisation.[21] Fostering stronger collaboration between information managers equipped with knowledge of the hard skills alongside trained social scientists and business analysts could be an effective mitigating measure.[24] Incorporating volunteers, Digital Humanitarians, or consultants is an alternative means of gaining the necessary knowledge to manage big data that may not be available internally. Semantic knowledge management can also come into play as the semantics of humanitarian situations can be a limiting factor in effective data management. Utilising the Humanitarian Exchange Language (HXL) and other vocabularies could help mitigate this risk.[27]

CASE STUDY | **EBOLA RESPONSE 2014-2015 AND STATISTICS WITHOUT BORDERS**[28]

During the Ebola crisis in West Africa from 2014-2015, Statistics Without Borders (SWB) deployed through the Digital Humanitarian Network at the request of UN OCHA and with NetHope to assist with the visualisation of the data related to the crisis response. With NetHope, SWB analysed telecommunication issues such as mobile connectivity strength, data transfer types and methods, and connectivity needs for staffing demands. Through UN-OCHA, SWB used big data to analyse location and sizes of burial teams, testing laboratories, and infected healthcare workers. SWB also tracked volume of food aid being received from the World Food Programme, locations of internet, radio, and voice communications services, and educational facility needs.

PART THREE

GUIDANCE FOR INCORPORATING BIG DATA INTO HUMANITARIAN RESPONSE

INTERNAL CONSIDERATIONS

Applying big data into humanitarian organisational operations requires a lot of consideration and planning before implementation. First and foremost, organisations must understand **staff capacity** to incorporate it into operations. It might be necessary to hire new staff or contractors or to determine if volunteers can be utilised. A combination of staff and volunteer data management could be the ideal situation to maximise impact and minimise costs. Volunteer groups such as the Digital Humanitarian Network and local tech hubs/labs could be potential volunteer partners. Organisations can also create partnerships with other organisations and companies that can assist in data management and analysis such as QCRI, UN Global Pulse, or private companies. Additionally, the risk checklist should be done in order to inform if the risks are too high or to put mitigating measures into place before the process begins.

Data should not be brought into operations simply for the sake of data. It should only be incorporated if it can be used to solve a problem and organisations should clearly identify the issue before putting a structure in place.

Oftentimes data is requested but never actually integrated into operations and decision making for a variety of reasons. **Having a structure in place can help to minimise this risk**. The structure should include the assignments of responsibilities, a decision-making and reporting structure, information sharing mechanisms, and an evaluation procedure. Responsibility for data validity, selecting data types to use, data tools, software and hardware to use, and data visualisation must be assigned. The decision-making process is important, which includes determining how big data can inform decision-making, selecting how and when to use big data, selecting the final interpretation of the data, making predictions, and taking actions.

> "Big data informs action and action informs big data."

Once the determination has been made to incorporate big data, then an **evaluation mechanism and feedback loop** should be put into place to monitor its effectiveness. As big data informs action and action informs big data, this evaluation process is an essential component of the response procedures. Organisations should **communicate** amongst each other and share information. Some information gathered will be more relevant to other organisations that are working in different thematic areas. This information can be shared between organisations, with necessary privacy measures considered, to assist in overall humanitarian response. Communication also puts a safeguard in place to prevent against the duplication of efforts. Finally, organisations should clearly document the use of big data in operations to ensure compliance with ethical standards and to inform decision-making in the future.

CASE STUDY | **NEPAL EARTHQUAKE 2015 AND MICROMAPPERS**[29] [30]

MicroMappers alongside the Standby Task Force and Qatar Computer Research Institute used big data to create maps of the damage created as a result of the April 25, 2015 earthquake in Nepal and the locations of where humanitarian groups were deploying. 2803 volunteers identified, collected, and processed 234,727 images and 55,044 Tweets about damage assessments, needs, and deployments in Nepal. Their maps were updated at least hourly during the most urgent and immediate time post-earthquake and they successfully created maps of 410 relevant images and 219 relevant texts.

Processing vast amounts of big data requires extensive **infrastructure**. If processing is done by the organisation itself, the amount of servers needed must be considered. Organisations may choose to outsource some of the processing to volunteers to save on infrastructure needs. Alternatively, data can be stored and processed on the cloud. Budget considerations will be critical for deciding how to establish the necessary infrastructure. Organisations may be able to set up partnerships with corporations such as Amazon or Google to make cloud computing more feasible.

POLICY CONSIDERATIONS

Perhaps one of the most difficult big data challenges to overcome is negotiating policy considerations. Because the big data revolution is still new and relatively uncharted, there are still many policy questions unanswered, but using what information is available and making informed decisions based on policy considerations is paramount. First and foremost, it is critical to know what data policies are already in place. The government and/or organisation could already have existing policies for data usage in general or specifically within aspects of humanitarian response. If so, organisations need to read the policies carefully and make sure to understand and comply with them consistently. If not, organisations should shape a data policy to put in place as a short-term solution at the onset of the crisis. This policy should be reevaluated consistently to make sure it is meeting the needs of the affected population and the organisation. Designing a data policy before an emergency or crisis occurs is highly recommended and taking policy issues into consideration early is paramount. Establishing partnerships with corporations and governments, especially in vulnerable locations prone to crisis, prior to the onset of an emergency is also best practice.

Privacy and ethics are important considerations in regard to data policies. Data often contains personal and confidential information. It is the responsibility of the organisation to make sure the private information is being cleaned as necessary from the data as soon as possible. Organisations need to ensure the privacy and

confidentiality of the affected population when using their personal data. This is the cornerstone of humanitarian ethics that should be upheld at all times. Once the data has been used, organisations must develop a policy for determining how long data should be kept and what to do with it following its use.

> "Many policies that discuss data considerations are already in place."

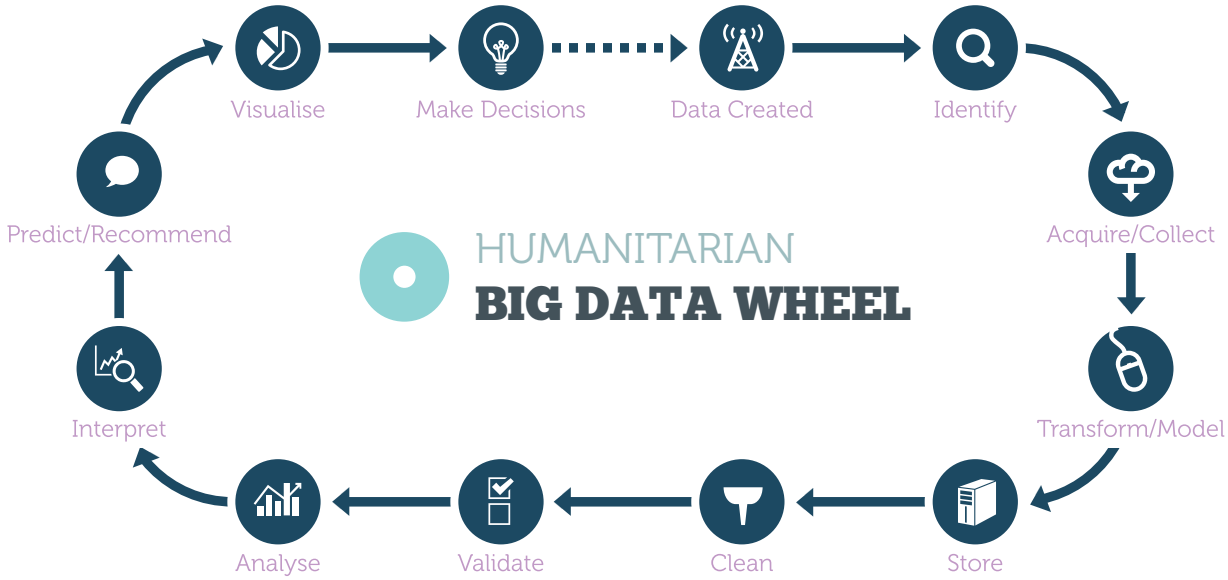
At the time of publication of this paper, in 2015, UN-OCHA had begun data security and policy discussions within the humanitarian sectors and partners. The goal of these discussions is to establish policies for humanitarian organisations to uphold when using data in humanitarian work. Many policies that discuss data considerations are already in place. The following are a list of useful resources to use when shaping data policy:

- **Organisation for Economic Co-operation and Development (OECD):** Mission to "promote policies that will improve the economic and social well-being of people around the world" and developed the Guidelines on the Protection of Privacy and Transborder Flows of Personal Data.
- **International Aid Transparency Initiative (IATI):** Global standard for availability of information on government and NGO spending; publishes the IATI Standard as guidelines for technical publishing.
- **Developing an Organisational Policy for Responsible Data:** Amy O'Donnell from Oxfam put together a page and forum for organisations to understand their process in developing a data policy and to ask questions while shaping their own.
- **Oxfam:** Developing a Responsible Data Policy and has created a robust Open Information Policy to guide how information is used and shared.
- **International Committee of the Red Cross:** The Professional Standards for Protection Work provides minimum standards for humanitarian and human rights organisations.

HUMANITARIAN BIG DATA WHEEL

In order to utilise big data in humanitarian response, there are multiple steps in the process that must be undertaken before being able to make decisions based on the information. Each step could involve different staff or volunteers and require a varying array of resources. For example, the people who validate data may be different than the people who visualise data or make decisions based on data. Organisations should take the time to think about each step in the data wheel and note the people involved along the way. Assigning responsibilities for each step can help ensure a smooth, thoughtful, and secure process. Because some data may be confidential, having a secure process in place is necessary for upholding ethical standards.

1. **Data Created:** Generated by the public, companies, machines, etc.
2. **Identify:** Determine which data sources should and could be used.
3. **Collect/Acquire:** Gather from open sources and/or obtain necessary permissions to acquire data from governments, companies, organisations, etc.
4. **Transform/Model:** Put (multiple sources of) data in proper format or structure for querying and analysis.*
5. **Store:** Determine appropriate and secure method of storing data.*
6. **Clean:** Ensure confidential information is removed from dataset.
7. **Validate:** Check for errors, including bias errors, and make corrections.
8. **Analyse:** Filter, search, and manipulate datasets to gain insight into situation and key questions.
9. **Interpret:** Describe findings from data for relevant staff and users of information.
10. **Predict/Recommend:** Discover patterns and trends to make informed predictions and recommendations.
11. **Visualise:** Create advanced visualisation, interactive dashboards, maps, infographics, etc.
12. **Make Decisions:** Use data in combination with human considerations and causal inferences to meet decision makers needs.



The steps of transforming/modeling and storing are interchangeable. The order depends on the type of data in the pipeline. Streaming data will be transformed/modeled first before storing and batch data will be stored first before transforming/modeling.

The decisions made then tend to have an impact the affected population, which is why the process is a wheel. Therefore, new data will be created that reflects the changes that result from decision-making. Monitoring the patterns in data after implementing changes will help inform decision makers on how their decisions are impacting the affected population. This can result in a more in depth understanding of the population and better decision making in the future.

ACQUISITION OF BIG DATA

Big data acquisition includes collection, transmission, and pre-processing. This phase includes a number of dangers and considerations to highlight including data sources and data redundancy.

There are numerous **open source** data resources available. These open source datasets allow for data scientists to very quickly and easily obtain and analyse the information available in order to allow decision makers to make the most informed decisions on the quickest timeframe. Open source data does not require permissions and has already been cleaned for confidentiality. There is a push to make more data open worldwide and numerous governments have already begun publishing datasets. See Annex III for a list of open data sources that can be used for data acquisition.

Oftentimes organisations will need to acquire data from sources that are not open source though. These sources usually include **governments and private corporations**. Gathering this data is more time consuming and process-laden as it often requires receiving special permissions, which take time and resources, and involve a lengthy process of making sure the data is cleaned and has privacy considerations in place. Organisations should begin preparedness measures to create partnerships with governments and corporations in advance of crises so data can be released immediately upon the onset of an emergency.[i] [12] Some preventative measures include establishing

i *Data Philanthropy, companies sharing their datasets for social good, is increasing in popularity. This is particularly helpful if media and telecommunications companies participate, which would allow for efficient media monitoring.*

CASE STUDY | **MOBILE DATA COLLECTION** [18]

The Open Data Kit (ODK) by the Red Cross enables faster and more efficient data collection by utilising mobile phones to collect data. Data is collected through built-in mobile cameras and GPS and is processed through a single cloud-based or local database. It is open source, multimedia rich, scalable, flexible, and easy-to-use. It enables faster and more accurate decision-making. It has been used to register people in internally displaced camps, in malaria prevention programs, for monitoring and evaluation, and more. It has been deployed in dozens of countries in Latin America, Africa, and Asia since 2010.

MOUs and SLAs with these entities. For example, the headquarters office of an organisation can negotiate agreements and protocols with mobile phone operators for call detail records (CDRs) as a preemptive measure for crises.

The collected datasets may sometimes include redundant or useless data, which unnecessarily increases storage space and affects the subsequent data analysis. For example, high redundancy is very common among datasets collected by sensors for environmental monitoring. Data compression technology can be applied to reduce the redundancy. Therefore, data pre-processing operations are indispensable to ensure efficient data storage and exploitation.[4]

USING BIG DATA & TRADITIONAL DATA

Using both types of data together could be the most beneficial means of information management. Historical data and offline data methods can be used to create initial models and templates for assessment then online and big data analysis can use these pre-existing frameworks.[5] Starting with the framework built from offline methods also helps maintain foundational issues of accuracy, reliability, transparency, and measurement.[19] It is critical to remember that big data is not a substitute, but rather a complement, to traditional data analysis.

> "Big data is not a substitute, but rather a complement, to traditional data analysis."

Organisations should determine the type of data to use on a case-by-case basis by undergoing a cost-benefit analysis. Using a combination of big data alongside traditional methods is likely the best way to gain a comprehensive understanding of the information available. Together they can improve on the way data is collected and analysed. By detecting innovative big data collection strategies while improving traditional data collection, a wider variety of sources can become available for use in humanitarian response.[19]

CONTEXTUAL CONSIDERATIONS

When initiating the use of big data into response activities, organisations should carefully analyse how the context of the implementation affects the process. Taking what information is already known into account and determining if it is accurate, valid, and useful before collecting and analysing new data should be done before the big data pipeline begins. There are a number of factors to consider including the type of emergency, geographic location, and the thematic focus area. These contextual considerations will assist in designing a framework for a big data plan.

The location of the emergency is significant because it determines the type of data available, what language it is in, and whether it can be easily accessed. Having the ability to readily translate data should be a priority consideration before the acquisition process. Social data is location dependent and organisations should evaluate what are the main social media and communications platforms being used regionally. Also, the energy infrastructure in the location of the crisis is an important factor to consider as it may be damaged by the crisis or be sporadic, which could lead to difficulties in data processing.

Some governments have accessible open data while others involve a lengthy application process and local corporations may have varying procedures for data sharing. This affects what data is available to use and how quickly it can be collected. Also, crises often affect the most marginalised populations and these people might not have the strongest communication methods. It becomes critical in these instances to determine if the data available accurately reflects the affected population.

> "Big data could provide answers to fundamental questions for all sectors."

Types of Emergencies*

Natural Disasters	Conflict and Violence	Man-Made	Biological	Complex
• *Avalanches* • *Cyclones* • *Droughts* • *Earthquakes* • *Floods* • *Hurricanes* • *Storms* • *Tornados* • *Tsunamis* • *Typhoons* • *Volc. Eruptions*	• *Armed Conflicts* • *Bombings* • *Civil War* • *International War* • *Terrorist Attacks*	• *Building Collapses* • *Chemical Explosions* • *Fires* • *Hazardous Material Spills* • *Industrial Accid.* • *Nuclear Accid* • *Plane Crashes* • *Train Crashes*	• *Biological Weapons* • *Epidemics* • *Infectious Disease* • *Plagues* • *Plant or Animal Contagion* • *Plant or Animal Infestation*	• *Climate Change* • *Displaced Populations* • *Food Insecurities* • *Human Trafficking* • *Refugees*

* *The list was based on emergencies identified by Glide and expanded to add additional man-made emergency types. This list is not comprehensive.*

TIMELINE AND LIFESPAN OF BIG DATA

Different types of information are needed at varying times in the humanitarian response timeframe. Some data that is useful in the early stages of response might not be useful later on and vice versa. Also, some data may be needed as much and as often as possible at the beginning but will only be needed once a day or once a week as time progresses. Decision makers' needs and the reporting structure within the organisation also influence the timing of big data management. These considerations, alongside how long datasets are valid and useful, will inform how frequently datasets will need to be refreshed or updated.

Organisations should be mindful of the resources needed to support data management and processing and have the capacity to fully support it through the duration of the data lifetime. Determining how long datasets should be stored and what should be done with those datasets once their useful life is up are additional important considerations.

CONCLUSION

As the big data revolution continues to gain momentum, humanitarian organisations can look to this new tool to solve persisting problems and gain new insights. Many have claimed that big data will provide answers to fundamental questions for all sectors, but the reality is that large datasets alone cannot solve complex humanitarian issues.[24] It can be another option to understand and analyse the crises at hand. Organisations should carefully weigh its benefits, risks, costs, and needed resources before implementing a big data strategy. It is a rapidly changing field due to the expansion of data creators/sources and data collection tools. The best means of staying on top of the changes is to have a foundation and strategy in place that takes the most consistent components into consideration such as policies, ethical standards, and decision-making practices. With these measures in place, alongside established MOUs and SLAs, organisations can make great strides in incorporating big data analytics into operations.

PART FOUR ANNEX

QUESTIONS TO CONSIDER

Using big data in humanitarian response is intended to support decision-making. The following is a list of questions to consider and decisions that need made in order to utilise big data. This list is not comprehensive, but it should provide a thorough starting point for organisations looking to incorporate big data analytics into operations.

Acquisition

- Have all risks been considered before obtaining data?

- What types of data would be useful?
  Are these types accessible in the country/region?

- Will data be obtained via an open source method or will it need to be negotiated?
  - Open Source
    - Are the sources valid and reliable?
    - How long will it take to receive the data?
  - Closed Source
    - What is involved in obtaining the data?
    - What permissions are needed?
    - Is it legal to obtain the data?
    - Should an MOU or SLR be negotiated?
    - How long will it take to acquire the data?
      - Will data still be relevant and needed by the time it is obtained?
    - Does that data cost anything to obtain?
      - If so, has a cost-benefit analysis been considered?

- Does the dataset include redundant data?

- Should/could multiple sources be used?

- Is the data in a useable format?
  - If not, can it be cleaned and transformed quickly enough and accurately?

Policy

- Is a data policy in place by the government?
    - If so, are all policies being complied with and understood?
    - If not, can a general a policy be put in place quickly?

- Is a data policy in place within the organisation?
    - If so, are all policies being complied with and understood?
    - If not, can a general a policy be put in place quickly?

- Is private information being cleaned as necessary from data?

- Can personal privacy and confidentiality of the affected population be guaranteed?

- What is being done with the data following the emergency?

Timeline

- How will data needs change over time?

- Are there necessary resources available for the duration of the data lifetime?

- How long are datasets valid?

- How long are datasets useful?

- How often does data need updated/refreshed?

- How frequently does data need to be reported?

- What is the reporting structure and how long does it take?

- How long should datasets be stored?

Context

- Why should big data be used?
- What are the gaps in the knowledge about the crisis?
- What is already known about the community?
  - Is it reliable and/or up to date?
- What type of data is useful for the specific type of emergency?
- What data is available in the region?
- What form of communication is most widely used in the community?
- What are the main social media platforms regionally?
- What language is the data in and can it be translated if necessary?
- Has the crisis impacted the type of data available?
- Can the energy infrastructure support the needs of processing big data?
- Is the data relevant to the thematic area?
- Does the data accurately reflect the affected population?

Internal

- What is the overall big data strategy?
- Is there staff capacity to incorporate big data into response?
- How can the organisation allow space for creativity in analysis while still being strategic?
- Can volunteers be utilised?
  - Is there staff capacity to manage volunteers?

- Should partnerships be established with other organisations/companies to assist in data management or analysis?
    - Which organisations are relevant?
    - What can the organisation afford?
    - Does an MOU or SLA need to be in place?
- Can duties be shared between staff or between staff and external people?
- How does big data inform decision-making?
- Who is ensuring the validity of the data?
- How is validity being ensured?
- How quickly can we generate answers to humanitarian questions using big data?
- Will the person who is providing resources also be the person asking for consent to data?
- What is the feedback loop for how big data is being used?
- Is there an evaluation mechanism for determining the effectiveness of big data usage?
- How can organisations share information together?
- What mechanisms are in place to safeguard against duplication of efforts?
- Who has responsibility for selecting when and how to use big data and which types?
- Has a risk checklist been performed?
- Are sufficient resources available including software and hardware?
- What are the different ways the data can be interpreted?
    - Who determines the final interpretation?
- How should data be visualised?
- How can data be used to make predictions?
    - What actions can be taken based off of the predictions?

CHECKLIST OF RISKS

The risk checklist can be utilised to identify risks that may be involved in using big data overall or with specific data sources. When a risk is identified, organisations should assess necessary mitigating measures or special considerations to take. The list is not comprehensive and organisations should make modifications and additions as necessary.[31]

Geographic

- Lack of access to mobile phones and internet by locals for reporting during crisis
- Availability of resources as a result of damage caused by crisis
- Connectivity issues; disrupted networks in disasters
- Availability of energy infrastructure and electricity competencies
- Difficult to interpret language of data

Users

- Being overwhelmed by the amount of types of big data available
- Assumption that big data can be a substitute for traditional data
- Hesitancy to use tools provided by other sectors
- Lack of staff time for processing data in real time
- Inability to decide the best means for storing information
- Not having a clear problem statement or mandate on how/why to use big data
- Too expensive for staff, equipment, electricity, etc. to incorporate big data into response
- No coordination of efforts between organisations, government, private sector, and disciplines
- Lack of trust between different people, organisations, government, and private sector on validity and confidentiality of data
- Unclear decision maker triggers and timelines
- Duplication of efforts between different organisations
- No oversight across all organisations of information gaps or what each organisation is doing

Technology Limitations

- Inability to search, aggregate, organize, and validate many types of data
- Outdated or unpopular technology being used
- Excessive energy consumption used by data processing system mainframes
- Difficulty sourcing resources and efforts by multiple stakeholders
- Difficulty scaling to expand with greater volume and complexity

- Difficulty structuring data in consistent and usable format for analysis
- Filtering out useful information because it is inconsistent or appears to be an error
- Information from varying sources having different formats
- Models are not adaptable, so cannot be predictive

Validity

- "Red team" attacks on systems where data-generating processes are manipulated by people
- Overloading of social media networks to ensure certain topics are trending or spread rumors
- Manipulation of internet and mobile applications by engineers that make research and data collection inconsistent
- Perception of trends as being patterns
- Inconsistencies in data representation that lead to inaccurate measurements
- Data redundancy and compressed data
- Difficulty identifying sources of information across multiple data types
- Unreliable data sources (ex. faulty sensors, biased opinions, outdated websites, etc.)
- Untrusted data platforms
- Expiration of data
- Language barriers

Policy and Ethics

- Lack of standard protocols and nomenclatures for identifying information
- No international framework for standardizing information, especially in social media
- Security issues with accessing certain data and receiving permissions to obtain
- Not maintaining necessary confidentiality in datasets
- Strict policies of governments that make data difficult to obtain and/or use
- Data protection and privacy challenges
- Power asymmetry; person providing resources is also asking for data consent
- Consent to data usage will be difficult to obtain

Staff Knowledge

- Lack of staff with knowledge necessary to facilitate the big data pipeline
- Staff specialising in certain skills without ability to be inter-disciplinary
- Unknowledgeable decision-makers about the capabilities of big data
- Weak collaborative ties between information managers and social scientists

OPEN DATA SOURCES

Datacatalogs.org is a project aiming to be the most comprehensive source of open data catalogs in the world. It is a searchable portal that provides links to all open data by governments and is curated by government representatives, international organisations, and NGOs.

Other Non-Governmental Open Source Data Platforms

Amazon Web Services	Human genetic information, NASA's satellite imagery and more.
DBPedia	Wikipedia data
Facebook Graph	Public information of Facebook users
Freebase	Community-compiled database about people, places, and things
Gapminder	World Health Organization and World Bank economic, medical, and social statistics
Google Books Ngrams	Searchable full text of millions of books
Google Trends	Search volume (as a proportion of total search) for any given term since 2004
HDX	Open platform for sharing and finding humanitarian data.
Likebutton	Overview of public Facebook data of what people "Like"
National Climatic Data Center	Environmental, meteorological, and climate data sets from the US National Climatic Data Center
New York Times	Searchable, indexed archive of news articles going back to 1851
The CIA World Factbook	History, population, economy, government, infrastructure, and military information
Topsy	Searchable database of public tweets going back to 2006
UNdata	Statistical data from Member States and UN Agencies

MATRIX OF PRODUCTS (Example)

Organisations can create a variety of matrices to assist in determining when and how to use big data and which types. Some potential matrix examples include: timelines, products, specific countries or regions, or certain disaster types. Matrices need to be modified to be relevant for the organisation, crisis, and topic at hand and each organisation must adjust accordingly. These matrices can serve a variety of purposes from helping analysts prioritise to decision-making to establishing MOUs with partners.

**This matrix is for illustration purposes only - each organisation should create a chart that reflects their own needs.*

Usefulness ranking 0 - 9 (not useful - very useful)

Products	*Cellular Data*	*Census Data*	*Phone GPS Tracking*	*News Feeds*	*Photo Sharing*	*Satellite imagery*	*Security Cameras*	*SMS*	*Social Media*	*UAVs*	*Video Sharing*
3W (Who does What Where)	1	1	1	7	3	1	1	1	7	1	3
Assesment											
Response Monitoring											
Datasets (COD/FOD)											
Contacts											
Funding											
...											

DECISION MAKERS NEEDS (Example)

The following matrix on decision makers needs was created by a group of organisations and individuals making up the Decision Makers Needs (DMN) Community. The DMN held a workshop on the information requirements for humanitarian responders and released a report on their findings.[32] The matrix utilises the results of the DMN workshop in the left column, arranged roughly corresponding to the response timeline with earlier information at the top.

To use this matrix, organisations should modify accordingly to the needs of the specific decision makers and arrange in an appropriate timeline. Next, data sources that can be used to support the decisions and their overall validity should be identified. Noting how soon and how frequently the data will need to be used and refreshed can be added to enhance management practices. Naming the responsible parties for each and identifying risks associated can help ensure data security and ethical principles being upheld and to help inform of any mitigating measures that will need to be taken.

Decision Makers Needs Chart (Example)

Decision Makers Needs	*Data Sources*	*Data Validity*	*Timeframe*	*Resources Needed*	*Risks*
[First Days]					
CONTEXT & SCOPE	(list data sources)	(% valid/reliable of each source)	(identify when information is needed and for how long)	(identify staff/volunteer needs)	(list risks involved with the data)
SCOPE OF EMERGENCY SITUATION	Example:				
Impact: damage to infrastructure, livelihoods, etc.	- SMS (partner with xyz telecommunications)	100%	- First days, refresh for first months	- 1 Staff	*Difficulty obtaining partnerships, delay in receiving data, locations not being geotagged on social media, falsification of photos or videos, weather disruptions with UAV, electricity disruptions*
	- Messaging apps (Whatsapp)	100%	- First days, refresh for first months	- 1 Staff	
	- Photo sharing (Twitter)	75%	- First days, refresh for first weeks	- Volunteers (DHN)	
	- Video sharing (YouTube)	65%	- First days, refresh for first weeks	- Volunteers (DHN)	
	- Social media (Twitter)	30%	- First days, refresh for first weeks	- Volunteers (DHN)	
	- UAVs (partner with xyz UAV group)	80%	- First days	- 1 Staff and external contractor (UAViators)	
	- Security cameras (partner with xyz security)	75%	- First days	- 1 Staff	

For the full table created in the Decision Makers Needs workshop, kindly see their report.

REFERENCES

1 Harvard Humanitarian Initiative, "Disaster Relief 2.0: The Future of Information Sharing in Humanitarian Emergencies," 2011, UN Foundation & Vodafone Foundation Technology Partnership.

2 Manyika, James et al, "Big data: The next frontier for innovation, competition, and productivity," 2011, McKinsey Global Institute.

3 OCHA and DARA, "Saving Lives Today and Tomorrow: Managing the Risk of Humanitarian Crises," 2014.

4 Chen, Min et al, "Big Data: A Survey," 2014, Mobile Networks and Applications, Vol. 19 Issue 2.

5 Paprotny, Alexander and Thess, Michael, "Realtime Data Mining: Self-Learning Techniques for Recommendation Engines," 2013.

6 Press, Gil, "12 Big Data Definitions: What's Yours?," 2014, Forbes.

7 Bryant, Randal et al, "Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society," 2008, Computing Community Consortium.

8 Press, Gil, "A Very Short History of Big Data," 2013, Forbes.

9 IBM, "What is big data?"

10 Kim, Gang-Hoon, et al., "Big-Data Applications in the Government Sector," 2014, Communications of the ACM, Vol. 57 No. 3.

11 MacLean, Dana, "Analysis: Potential, pitfalls of "big data" for humanitarians," 2013, IRIN.

12 Barton, Simon, "Big Data and Humanitarianism," 2014, Innovation Enterprise.

13 Leskovec, Jure, et al, "Mining of Massive Datasets," 2014.

14 Gordon, Patrick, personal communication, 21 July 2015.

15 Lendrich, Pauline, "Featured Tool: Flowminder Foundation," 2014, Emergency Journalism.

16 Meier, Patrick, "Summary: Digital Disaster Response to Philippine Typhoon," 2012, iRevolutions.

17 Meier, Patrick, "How the UN Used Social Media in Response to Typhoon Pablo (Updated)," 2012, iRevolutions.

18 The Red Cross and Red Crescent, "Principled Approach to Innovation," 2015.

19 Lazer, David et al, "The Parable of Google Flu: Traps in Big Data Analysis," 2014, Science.

20 Jagadish, H.V. et al, "Big Data and Its Technical Challenges," 2014, Communications of the ACM, Vol. 57 No. 7.

21 Yiu, Chris, "The Big Data Opportunity: Making governments faster, smarter and more personal," 2012, Policy Exchange.

22 Chunara, Rumi et al, "Social and News Media Enable Estimation of Epidemiological Patterns Early in the 2010 Haitian Cholera Outbreak," 2012, American Journal of Tropical Medicine and Hygiene, Vol.86 No. 1.

23 Homeland Security, "Using Social Media for Enhanced Situational Awareness and Decision Support," 2014.

24 Grimmer, John, "We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together," 2015.

25 Meier, Patrick, personal communication, 29 July 2015.

26 Verity, Andrej and Cruz, Hildemar, "Humanitarian Data Scientist - who and how?," 2014, verity think.

27 Clark, Tim, personal communication, 28 July 2015.

28 Digital Humanitarian Network, "Digital Humanitarians supporting the West Africa Ebola Response."

29 Meier, Patrick, "A Force for Good: How Digital Jedis are Responding to the Nepal Earthquake (Updated)," 2015, iRevolutions.

30 MicroMappers, "MicroMappers for Nepal Earthquake Response (April 2015)," 2015.

31 Leson, Heather, "Advanced Research and Technology from WGET: ICT Humanitarian Innovation Forum," 2015, Text on Techs.

32 Gralla, Erica et al, "Report from the workshop on Field-Based Decision Makers' Information Needs in Sudden Onset Disasters," October 2013.

digital humanitarian network

DM-N
Decision Makers Needs

Guidance for Incorporating

BIG DATA INTO HUMANITARIAN OPERATIONS