Guidance for Incorporating Big Data into Humanitarian Operations

Purpose: The intention of this document is to provide a broad overview of big data usage in humanitarian organisations as well as to provide general guidance on how humanitarian organisations can incorporate it into operations. The goal is to create dialogue and generate some structure in the conversation about big data. The document is designed to give guidance to organisations on why, when, and how to use big data and to give staff a resource to reference.

The primary audience for the paper is humanitarian response organisations across the globe who are looking for new and innovative ways to combat difficult 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 thoughtfully and strategically. It may 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.

What is needed: Please provide your comments, questions, etc. about the concept in general and specifically. We are looking to improve the idea and your input is invaluable. The purpose is to help understand big data, structure an explanation of it for organisations to understand what it would mean to incorporate big data into operations, and to analyse its impact on humanitarian response.

Provide feedback directly into the document where possible. If you would like to comment outside of the document via email or a Skype call, that would be appreciated as well.

Please provide any input no later than 1 September.

Outline:

INTRODUCTION

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. 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 all humanitarian problems.[1] The time and resource constraints of humanitarian organisations and responders is becoming increasingly stretched with the increasing number of length of humanitarian crises. This reality, in combination with the vulnerability of the affected population that would be the subjects of the data, makes many organisations resist incorporating big data practices into response. The risk of not incorporating it is huge due to its growing wealth of critical information and rapid rise in popularity. Therefore, it is essential for humanitarian organisations to begin the process now of understanding what big data is, what its analysis can yield answers to, and how and when to use it.

PART ONE: THE BASICS

I. 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 software and people 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.[3] Also, due to the differing concerns of scientists, technological enterprises, 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)".[5] 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.

---

[1] Grimmer, John, " We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together," 2015.
[2] Manyika, James et al, "Big data: The next frontier for innovation, competition, and productivity," 2011, McKinsey Global Institute.
[3] Manyika, James et al, "Big data: The next frontier for innovation, competition, and productivity," 2011, McKinsey Global Institute.
[4] Chen, Min et al, "Big Data: A Survey," 2014, Mobile Networks and Applications, Vol. 19 Issue 2.
[5] Chen, Min et al, "Big Data: A Survey," 2014, Mobile Networks and Applications, Vol. 19 Issue 2.

Traditional Data

Traditional (or "classical, offline, or small data") data mining is the counterpart to real-time data mining, which is generally synonymous with big data. Traditional data mining is learning from historical data by collecting, studying, and deducing results or decisions based upon the information collected from the past.[6] It often contains information that cannot be obtained by big data such as historical information or personal surveying.

II. 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.[7] The term was popularized over ten years later in 2008 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.[8] In reality, 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".[9] Many believe that since big data still does not have a consistent and agreed upon definition by scholars that the term may fade out of existence or may change entirely as new technologies emerge.

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.[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 Digicall, Haiti's largest mobile phone company, to compare people's movements pre- and post-earthquake. This enabled responders to find where most people fled to and from and deliver medical and relief supplies more efficiently.[13] 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 to newly formed data organisations and labs.

III. Types of Big Data

Big data can be understood through two main streams: **streaming and batch**. Streaming data is constantly coming in and refreshing itself. It enables us to collect information about what people like, consume, think

[6] Paprotny, Alexander and Thess, Michael, "Realtime Data Mining: Self-Learning Techniques for Recommendation Engines," 2013.
[7] Press, Gil, "12 Big Data Definitions: What's Yours?," 2014, Forbes.
[8] Bryant, Randal et al, "Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society," 2008, Computing Community Consortium.
[9] Press, Gil, "A Very Short History of Big Data," 2013, Forbes.
[10] IBM, "What is big data?" and Kim, Gang-Hoon, Trimi, Silvana, and Chung, Ji-Hyong, "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] Barton, Simon, "Big Data and Humanitarianism," 2014, Innovation Enterprise.

about, feel, and where they go and what they do direct from the source. It is difficult to analyse because it is unpredictable and difficult to manipulate. Examples include social media and 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. It is collected one batch at a time. Government records are examples of batch data.

User-Generated	**Social Media**	**Captured**[14]	**Business Systems**
Blogs and Comments	Facebook	Banner Ad Clicks	Business Identification Number
Email	Google+	Customer Loyalty Cards	Driver's License
Messaging (WhatApp, Skype, etc.)	Instagram	Customer Surveys	News Articles
Mobile apps	LinkedIn	Focus Groups	Passport Tracking
Personal Documents	Pinterest	Internet Searches	Radio Transmissions
Photo Sharing (Flickr, etc.)	Skype	Mobile Phone Tracking	Rental Records
Rating Systems	Tumblr	Security Cameras	School Records
SMS	Twitter	UAVs	Tax Information
User-Generated Maps	Vine	Vehicle GPS	Television Transmissions
	Private/custom social media networks	Web Logs	Trade Records
		Website Click-Throughs	Transportation Logs
			Voting Records
Transactional	**Sensors**	**Biological**	**Public Records**
Banking Records	Home	Dental Records	Birth Certificates
Commerical Transactions	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 Transactions	Traffic		Military Records
Stock Trading	Weather		Police Records
Swipe Card Aid Transactions			Vehicle Registration

[14] 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 more in combination with traditional datasets, big data can enable deeper analysis into the humanitarian system.[15] 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.[16] 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. The major humanitarian questions are:

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

The following table illustrates further 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 to address the question. The table should be 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.

Question	Considerations	Sources of (big) data	Example
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	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.[17]
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	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	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, and crop damage,

[15] Gordon, Patrick, personal communication, 21 July 2015.
[16] Barton, Simon, "Big Data and Humanitarianism," 2014, Innovation Enterprise.
[17] Lendrich, Pauline, "Featured Tool: Flowminder Foundation," 2014, Emergency Journalism.

	time?	transmissions, commercial transactions, credit card history, school records, banking records, home sensors, pollution sensors, weather sensors, medical history	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.[18] [19]
Where 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	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 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.[20]

I. 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[21] It eliminates the need to store historical data for an extended period of time in order to analyze it at a later date[22] 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

[18] Meier, Patrick, "Summary: Digital Disaster Response to Philippine Typhoon," 2012, iRevolutions.
[19] Meier, Patrick, "How the UN Used Social Media in Response to Typhoon Pablo (Updated)," 2012, iRevolutions.
[20] The Red Cross and Red Crescent, "Principled Approach to Innovation," 2015.
[21] Paprotny, Alexander and Thess, Michael, "Realtime Data Mining: Self-Learning Techniques for Recommendation Engines," 2013.
[22] Paprotny, Alexander and Thess, Michael, "Realtime Data Mining: Self-Learning Techniques for Recommendation Engines," 2013.

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.[23]

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 adaptation when the environment changes and the changes are detected in the data.[24] Data is often time, date, and space stamped so decisions can be based on how applicable they are at any moment.[25] Systems become more responsive and effective as the data informs action and action informs data.[26] 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.[27] 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.[28] 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.[29] [30] [31] Additionally, it can assist in discovering new models that explain how things could evolve in the future.[32] Big data facilitates the understanding of human system at the systemic level.[33]

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 networks become a public forum where organisations can gain an internal perspective into the needs, thoughts, and movements of the local population. It 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

[23] MacLean, Dana, "Analysis: Potential, pitfalls of "big data" for humanitarians," 2013, IRIN.
[24] Paprotny, Alexander and Thess, Michael, "Realtime Data Mining: Self-Learning Techniques for Recommendation Engines," 2013.
[25] Chen, Min et al, "Big Data: A Survey," 2014, Mobile Networks and Applications, Vol. 19 Issue 2.
[26] Paprotny, Alexander and Thess, Michael, "Realtime Data Mining: Self-Learning Techniques for Recommendation Engines," 2013.
[27] Lazer, David et al, "The Parable of Google Flu: Traps in Big Data Analysis," 2014, Science.
[28] Lazer, David et al, "The Parable of Google Flu: Traps in Big Data Analysis," 2014, Science.
[29] Jagadish, H.V. et al, "Big Data and Its Technical Challenges," 2014, Communications of the ACM, Vol. 57 No. 7.
[30] Chen, Min et al, "Big Data: A Survey," 2014, Mobile Networks and Applications, Vol. 19 Issue 2.
[31] Yiu, Chris, "The Big Data Opportunity: Making governments faster, smarter and more personal," 2012, Policy Exchange.
[32] Yiu, Chris, "The Big Data Opportunity: Making governments faster, smarter and more personal," 2012, Policy Exchange.
[33] Lazer, David et al, "The Parable of Google Flu: Traps in Big Data Analysis," 2014, Science.

topics. It can be used to warn people of impending dangers, notify them how and where to find services, and ask them for direct feedback.

Case Study: Preventative Use of Big Data in Haiti

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 cases reported officially 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.[34]

II. 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 geographic, 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 at often the most marginalised regions. As such, many people do not have easy access to mobile phones and Internet, especially when an emergency strikes. This would limit amount of post-crisis user-generated big data that can be reported and collected. Also, they may have more vulnerable infrastructure that can be more easily destroyed or interrupted. A disaster can limit the amount of resources available and can be used, including damaged infrastructure and connectivity issues due to disrupted networks. Also, language can be a barrier as not all staff will speak the varying languages that the data may use. Finally, big data processing requires a robust availability of energy, so the appropriate energy infrastructure must be in place and there needs to be people with electricity competencies ready to work.[35]

Users

The challenges posed by humanitarians processing the data include selections, measurement error, decision makers considerations, other sources of bias, user-errors, clarity of purpose, and collaboration.[36] [37] It is critical to determine the problem that using big data will solve and how to collect information to find

---

[34] 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.
[35] Chen, Min et al, "Big Data: A Survey," 2014, Mobile Networks and Applications, Vol. 19 Issue 2.
[36] Homeland Security, "Using Social Media for Enhanced Situational Awareness and Decision Support," 2014.
[37] Grimmer, John, " We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together," 2015.

solutions before beginning the process of analysis. Analysis can be time consuming and expensive, especially during an emergency, so a clear plan for thoughtful measurment, careful research design, and useful statistical techniques in place is necessary.[38] Additionally, users can produce errors in analysis due to the overwhelming nature of big data.[39] 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.[40] There may be hesitancy to use tools between sectors, distrust of information provided by different groups, duplication of efforts, and a lack of oversight across all organisation regarding what all are doing and need.[41]

Technology Limitations

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.[42] We are at the onset of the data revolution, with 90% of big data still unstructured, so many systems are still yet to be developed and/or perfected.[43] 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.[44] [45] Data can come in many different formats as it is pulled from multiple sources.[46] 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.[47]

Validity

Ensuring that data is valid before using to inform actions is essential. There is a risk of error-laden data or data that misrepresents that situation. Some common errors to watch for include data redundancy, inaccurate, faulty, and untrustworthy sources, faulty sources, biases, and outdated/expired information.[48] [49] [50] To mitigate this, data scientists must understand and model these sources of errors to develop data

[38] Grimmer, John, " We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together," 2015.
[39] Homeland Security, "Using Social Media for Enhanced Situational Awareness and Decision Support," 2014.
[40] Chen, Min et al, "Big Data: A Survey," 2014, Mobile Networks and Applications, Vol. 19 Issue 2.
[41] Homeland Security, "Using Social Media for Enhanced Situational Awareness and Decision Support," 2014.
[42] Homeland Security, "Using Social Media for Enhanced Situational Awareness and Decision Support," 2014.
[43] Kim, Gang-Hoon, Trimi, Silvana, and Chung, Ji-Hyong, "Big-Data Applications in the Government Sector," 2014, Communications of the ACM, Vol. 57 No. 3.
[44] Lazer, David et al, "The Parable of Google Flu: Traps in Big Data Analysis," 2014, Science.
[45] Jagadish, H.V. et al, "Big Data and Its Technical Challenges," 2014, Communications of the ACM, Vol. 57 No. 7.
[46] Homeland Security, "Using Social Media for Enhanced Situational Awareness and Decision Support," 2014.
[47] Jagadish, H.V. et al, "Big Data and Its Technical Challenges," 2014, Communications of the ACM, Vol. 57 No. 7.
[48] Chen, Min et al, "Big Data: A Survey," 2014, Mobile Networks and Applications, Vol. 19 Issue 2.
[49] Lazer, David et al, "The Parable of Google Flu: Traps in Big Data Analysis," 2014, Science.
[50] Homeland Security, "Using Social Media for Enhanced Situational Awareness and Decision Support," 2014.

cleaning techniques.[51] Data should be monitored longitudinally and analysis should be replicable to ensure that patterns rather than trends are being identified.[52] 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.[53]

Policy and Ethics

One standard list of protocols and nomenclatures has yet to be developed for identifying information.[54] 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.[55] 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.

Staff Knowledge

As the use of big data grows, the need for professionals who can manage the big data pipeline 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.[56] Staff must have the hard skills required to acquire, store, transform, validate, model, and visualise the data along with the soft skills to make accurate causal inferences.[57] 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.[58] 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.[59] Incorporating volunteer and technical communities (V&TC) is an alternative means of gaining the necessary knowledge to manage big data.

Case Study: Ebola Response 2014-2015 and Statistics Without Borders

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

[51] Jagadish, H.V. et al, "Big Data and Its Technical Challenges," 2014, Communications of the ACM, Vol. 57 No. 7.
[52] Lazer, David et al, "The Parable of Google Flu: Traps in Big Data Analysis," 2014, Science.
[53] Lazer, David et al, "The Parable of Google Flu: Traps in Big Data Analysis," 2014, Science.
[54] Homeland Security, "Using Social Media for Enhanced Situational Awareness and Decision Support," 2014.
[55] Chen, Min et al, "Big Data: A Survey," 2014, Mobile Networks and Applications, Vol. 19 Issue 2.
[56] Verity, Andrej and Cruz, Hildemar, "Humanitarian Data Scientist - who and how?," 2014, verity think.
[57] Grimmer, John, " We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together," 2015.
[58] Yiu, Chris, "The Big Data Opportunity: Making governments faster, smarter and more personal," 2012, Policy Exchange.
[59] Grimmer, John, " We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together," 2015.

volume of food aid being received from the World Food Programme, locations of internet, radio, and voice communications services, and educational facility needs.[60]

[60] Digital Humanitarian Network, "Digital Humanitarians supporting the West Africa Ebola Response."

PART THREE: GUIDANCE FOR INCORPORATING BIG DATA INTO HUMANITARIAN RESPONSE

I. 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. The structure should include the assignments of responsibilities, decision making and reporting structure, information sharing mechanisms, and 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 including 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.

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 communication amongst each other and share information. Some information will be needed depending on the thematic area while other information may not, so sharing this information can assist in overall response. Communication also puts a safeguard in place to prevent against the duplication of efforts.

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, though this option is more expensive than others. Organisations may be able to set up partnerships with corporations such as Amazon or Google to make cloud computing more feasible.

Case Study: Nepal Earthquake 2015 and MicroMappers

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 and they successfully created maps of 410 relevant images and 219 relevant texts.[61] [62]

II. Humanitarian Big Data Pipeline

1. **Data Created:** Public, companies, machines, etc. generate types of big data
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:** 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.

III. Acquisition of Big Data

As the third phase of the humanitarian big data system, big data acquisition includes data collection, data transmission, and data 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

[61] Meier, Patrick, "A Force for Good: How Digital Jedis are Responding to the Nepal Earthquake (Updated)," 2015, iRevolutions.
[62] MicroMappers, "MicroMappers for Nepal Earthquake Response (April 2015)," 2015.

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 are 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. 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.[63]

The collected datasets may sometimes include much 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 environment 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.[64]

Mobile Data Collection

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.[65]

IV. Using “Big Data” and “Traditional Data”

Using both types of data together is 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.[66] Starting with the framework built from offline methods also helps maintain foundational issues of accuracy, reliability, transparency, and measurement.[67] It’s critical to remember the big data is not a substitute, but rather is a complement, to traditional data analysis.

Using a combination of big data alongside traditional methods is the best way to gain a comprehensive understanding of the information available. Together they can improve on the way data is collected and

[63] Barton, Simon, “Big Data and Humanitarianism,” 2014, Innovation Enterprise.
[64] Chen, Min et al, “Big Data: A Survey,” 2014, Mobile Networks and Applications, Vol. 19 Issue 2.
[65] The Red Cross and Red Crescent, “Principled Approach to Innovation,” 2015.
[66] Paprotny, Alexander and Thess, Michael, “Realtime Data Mining: Self-Learning Techniques for Recommendation Engines,” 2013.
[67] Lazer, David et al, “The Parable of Google Flu: Traps in Big Data Analysis,” 2014, Science.

analyzed and a wider variety of sources can become available through detecting innovative big data collection and ways of improving traditional data collection.[68]

V. 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 recommended. 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 the secondary 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. Humanitarian ethics 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 data policies 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.

VI. Context

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

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

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 the main social media and communications platforms are 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.

Types of Emergencies

Natural Disasters	**Conflict and Violence**	**Man-Made**	**Biological**	**Complex**
Avalanches	Armed Conflicts	Building Collapses	Biological Weapons	Displaced Populations
Cyclones	Bombings	Chemical Explosions	Epidemics	Food Insecurities
Droughts	Civil War	Fires	Infectious Disease	Human Trafficking
Earthquakes	International War	Hazardous Material Spills	Plagues	Refugees
Floods	Terrorist Attacks	Industrial Accidents		Combination of Others
Hurricanes		Nuclear Accidents		
Storms		Plane Crashes		
Tornados		Train Crashes		
Tsunamis				
Typhoons				
Volcanic Eruptions				

VII. Timeline and Lifespan of Big Data

Different types of information is 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.

PART FOUR: ANNEX

I. Questions to Consider

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?
        - 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 easily 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 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 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 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 useage?
- 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?

II. Checklist of Risks

Geographic

- ❏ Lack of access to mobile phones and Internet by locals for report 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 and 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 useable 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 of 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
- ❏ Unknowledgable decision-makers about the capabilities of big data
- ❏ Weak collaborative ties between information managers and social scientists

III. 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

Source	Data Available
Amazon Web Services	Human genetic information and NASA's database of satellite imagery of Earth
DBPedia	Wikipedia data
Facebook Graph	Public information of Facebook users
Freebase	Community-compiled database of structured data about people, places and things, with over 45 million entries
Gapminder	World Health Organization and World Bank economic, medical and social statistics data
Google Books Ngrams	Search and analyze the full text of any of the millions of books
Google Trends	Search volume (as a proportion of total search) for any given term, since 2004
Likebutton	Overview of public Facebook data of what people "Like" at the moment
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

IV. 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 examples include for timelines, products, specific countries or regions, or with certain disaster types. Matrices need to be modified to be relevant to both 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**	**Mobile 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	3	7	1	1	1	9	1	7
Assessment											
Response Monitoring											
Datasets (COD/FOD)											
Contacts											
Funding											
Humanitarian Access											
Humanitarian Needs Overview (HNO)											
Maps											
Dashboard											
Most Affected Areas Matrix and Graphs											
Prioritization Ranking											
Severity Estimate Ranking											
Infrastructure Damage											
Displacement Tracking											
Analysis Services											
Situation											
Funding											
Geospatial											
Gap											
Needs											
Strategic Plans											

V. Decision Makers Needs

Decision Makers Needs (first days)	Data Sources	Data Validity	Timeframe	Resources Needed	Risks
CONTEXT AND SCOPE	(list data sources)	(percentage 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:*	*Example:*	*Example:*	*Example:*	*Example:*
Impact: damage to infrastructure, livelihoods, etc.	*- SMS (partner with xyz telecommunications)* *- Messaging apps (Whatsapp)* *- Photo sharing (Twitter)* *- Video sharing (YouTube)* *- Social media (Twitter)* *- UAVs (partner with xyz UAV group)* *- Security cameras (partner with xyz security)*	*100%* *100%* *75%* *65%* *30%* *80%* *75%*	*- First days, refresh for first months* *- First days, refresh for first months* *- First days, refresh for first weeks* *- First days, refresh for first weeks* *- First days, refresh for first weeks* *- First days* *- First days*	*- 1 Staff* *- 1 Staff* *- Volunteers (DHN)* *- Volunteers (DHN)* *- Volunteers (DHN)* *- 1 Staff and external contractor (UAViators)* *- 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*
Geographic areas affected					
Assistance requirements					
AFFECTED POPULATIONS					
Number of affected, locations					
Status of affected: dispaced, vulnerable, etc.					
CONTEXT					
Local socio-economic, political context					
Local environment, weather, livelihoods					
Local community capacity, coping mechanisms					
PUBLIC AND MEDIA PERCEPTION					
Public perception, awareness, attention					
Media perception					
Political will, donor will					
HUMANITARIAN NEEDS					
NEEDS					
Number in need					
Types of needs (health, shelter, water, etc.)					
Locations of needs					
Needs of sub-groups: displaced, vulnerable					
PRIORITIES					

Geographic priorities					
Priorities across sector					
Within-sector priorities					

RESPONDER REQUIREMENTS					
Basic infrastructure for responders					
Security, access					
(first weeks)					
CAPACITY AND RESPONSE PLANNING					
OTHER ACTORS' CAPACITY AND RESPONSE					
(incl. gov't, military, local community, commercial aid agencies)					
Responses of other actors (who, what, where, etc.)					
Capacity of other actors (skills, equipment, scale, etc.)					
INTERNAL CAPACITY AND RESPONSE					
Internal response plan					
Internal capacity, structure					
AVAILABLE RESOURCES: financial, personnel, stocks, technical					
OPERATIONAL SITUATION					
SECURITY					
Current threats					
Future threats and risks					
ACCESS					
Limits to access					
Logistics capacity and structure					
MONITORING					
Issues					
Trends					
Accomplishments					
MEASURING AND OUTPUTS					

Measurable indicators for output					
Standards					
COORDINATION AND INSTITUTIONAL STRUCTURES					
COORDINATION OF THE RESPONSE					
External coordination (with other actors, various levels)					
Internal coordination (with other parts of the org.)					
RELEVANT LAWS AND POLICIES					
External coordination (with other actors, various levels)					
Internal coordination (with other parts of the org.)					
(first months)					
LOOKING FORWARD					
RECOVERY AND RECONSTRUCTION					
National development strategies					
Needs and plans for recovery					
PREPAREDNESS					
Information to collect before crisis					
OVERALL EVALUATION OF DATA STRATEGY					
META INFORMATION					
Information available	(information learned from data)				
Sources of information	(all data sources used)				
Accuracy, validity and information	(overall validity/reliability)				
Agreement on needs	(data needs specifically)				
Extent of assessments	(evaluations performed)				
Actions to improve access to information	(feedback loop and evaluations)				
Preparedness information	(what can be done in the future to be better prepared to use data)				

VI. References

- Barton, Simon, "Big Data and Humanitarianism," 2014, Innovation Enterprise.
- Bryant, Randal et al, "Big-Data Computing: Creating revolutionary breakthroughs in commerce, science, and society," 2008, Computing Community Consortium.
- Chen, Min et al, "Big Data: A Survey," 2014, Mobile Networks and Applications, Vol. 19 Issue 2.
- 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.
- Gordon, Patrick, personal communication, 21 July 2015.
- Grimmer, John, " We Are All Social Scientists Now: How Big Data, Machine Learning, and Causal Inference Work Together," 2015.
- Homeland Security, "Using Social Media for Enhanced Situational Awareness and Decision Support," 2014.
- IBM, "What is big data?"
- Jagadish, H.V. et al, "Big Data and Its Technical Challenges," 2014, Communications of the ACM, Vol. 57 No. 7.
- Kim, Gang-Hoon, Trimi, Silvana, and Chung, Ji-Hyong, "Big-Data Applications in the Government Sector," 2014, Communications of the ACM, Vol. 57 No. 3.
- Lazer, David et al, "The Parable of Google Flu: Traps in Big Data Analysis," 2014, Science.
- Lendrich, Pauline, "Featured Tool: Flowminder Foundation," 2014, Emergency Journalism.
- Leson, Heather, "Advanced Research and Technology from WGET: ICT Humanitarian Innovation Forum," 2015, Text on Techs.
- MacLean, Dana, "Analysis: Potential, pitfalls of "big data" for humanitarians," 2013, IRIN.
- Manyika, James et al, "Big data: The next frontier for innovation, competition, and productivity," 2011, McKinsey Global Institute.
- Marz, Nathan and Warren, James, "Big Data: Principles and best practices of scalable realtime data systems," 2015.
- Meier, Patrick, "A Force for Good: How Digital Jedis are Responding to the Nepal Earthquake (Updated)," 2015, iRevolutions.
- Meier, Patrick, "How the UN Used Social Media in Response to Typhoon Pablo (Updated)," 2012, iRevolutions.
- Meier, Patrick, "Summary: Digital Disaster Response to Philippine Typhoon," 2012, iRevolutions.
- MicroMappers, "MicroMappers for Nepal Earthquake Response (April 2015)," 2015.
- Paprotny, Alexander and Thess, Michael, "Realtime Data Mining: Self-Learning Techniques for Recommendation Engines," 2013.
- Press, Gil, "12 Big Data Definitions: What's Yours?," 2014, Forbes.
- Press, Gil, "A Very Short History of Big Data," 2013, Forbes.
- The Red Cross and Red Crescent, "Principled Approach to Innovation," 2015.
- Verity, Andrej, "Humanitarian Data Scientist - who and how?," 2014, verity think.
- Yiu, Chris, "The Big Data Opportunity: Making governments faster, smarter and more personal," 2012, Policy Exchange.