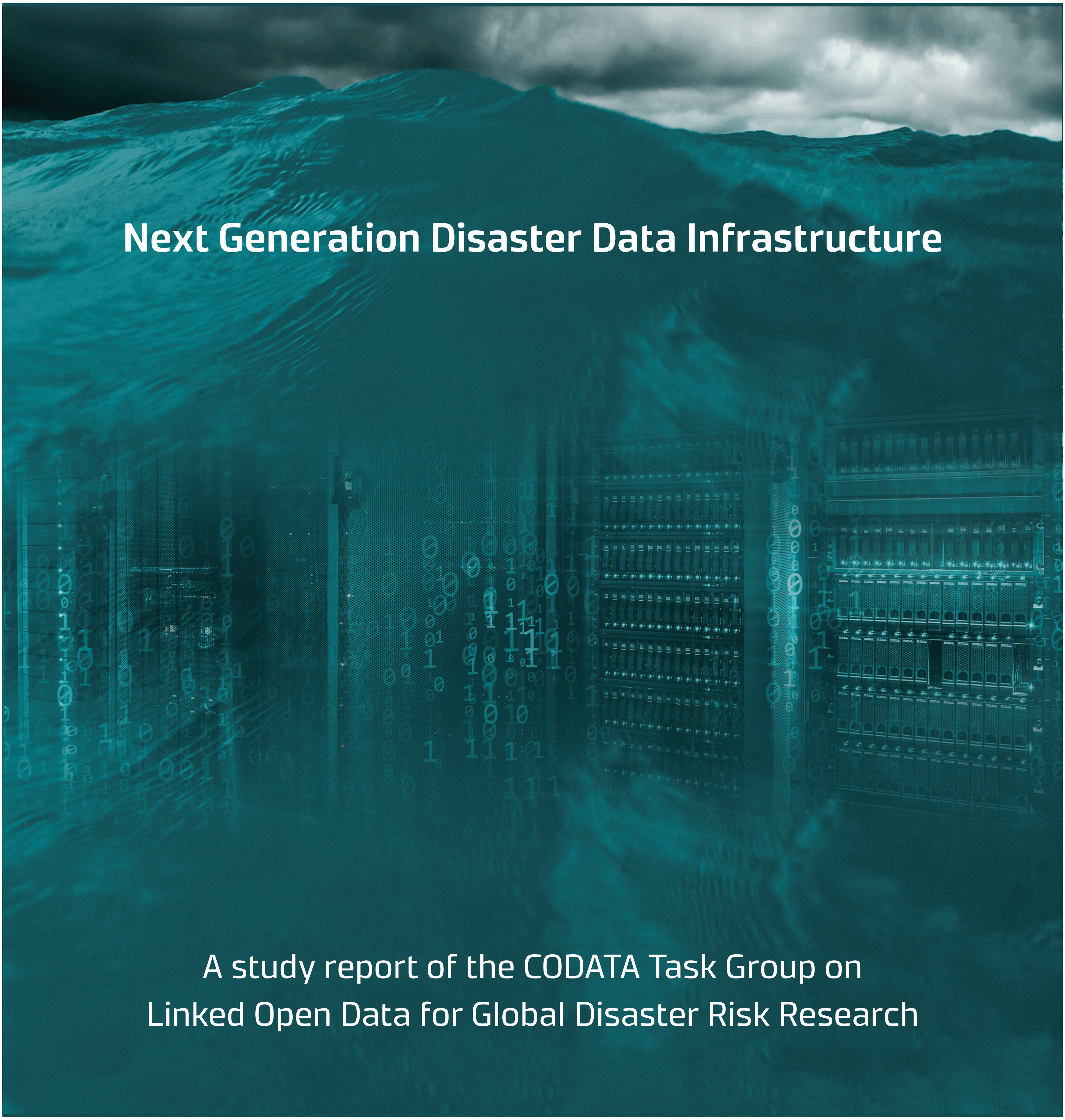
CODATA COMMITTEE ON DATA INTERNATIONAL SCIENCE COUNCIL

IRDR Integrated Research on Disaster Risk



Next Generation Disaster Data Infrastructure

A study report of the CODATA Task Group on Linked Open Data for Global Disaster Risk Research

Supported by

中国科学院 CHINESE ACADEMY OF SCIENCES

Tonkin+Taylor

Study Panel

Lead Authors in alphabetical order

Bapon Fakhruddin, Tonkin & Taylor International, New Zealand

Edward Chu, CODATA Task Group on Linked Open Data for Global Disaster Risk Research

Guoqing Li, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, China

Contributing Authors in alphabetical order

Amit Paul, BNA Technology Consulting LTD., India

David Wyllie, Tonkin & Taylor International, New Zealand

Durga Ragupathy, Tonkin & Taylor International, New Zealand

Jibo Xie, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, China

Kun-Lin Lu, CODATA Task Group on Linked Open Data for Global Disaster Risk Research

Kunal Sharma, National Disaster Management Authority, India

I-Ching Chen, CODATA Task Group on Linked Open Data for Global Disaster Risk Research

Nina Frolova, Institute of Environmental Geosciences of Russian Academy of Sciences (SC IEGRAS)

Nuha Eltinay, Arab Urban Development Institute, Kingdom of Saudi Arabia

Saurabh Dalal, National Disaster Management Authority, India

Raj Prasanna, Joint Centre for Disaster Research, Massey University

Richard Reinen-Hamill, Tonkin & Taylor International, New Zealand

Tengfei Yang, Institute of Remote Sensing and Digital Earth, Chinese Academy of Sciences, China

Vimikthi Jayawardene, The University of Queensland, Australia

Acknowledgments

This work is supported by the Committee on Data for Science and Technology (CODATA) of International Science Council (ISC), Integrated Research on Disaster Risk (IRDR); National Key R&D Program of China (2016YFE0122600) and Tonkin & Taylor International Ltd, New Zealand.

LODGD (2019), Next Generation Disaster Data Infrastructure -White Paper. CODATA Task Group, Linked Open Data for Global Disaster Risk Research (LODGD). September 2019. Paris

Cover page and design: Julian Turner
Language: Lindy Andrew & Lorraine Barnes

Table of Contents

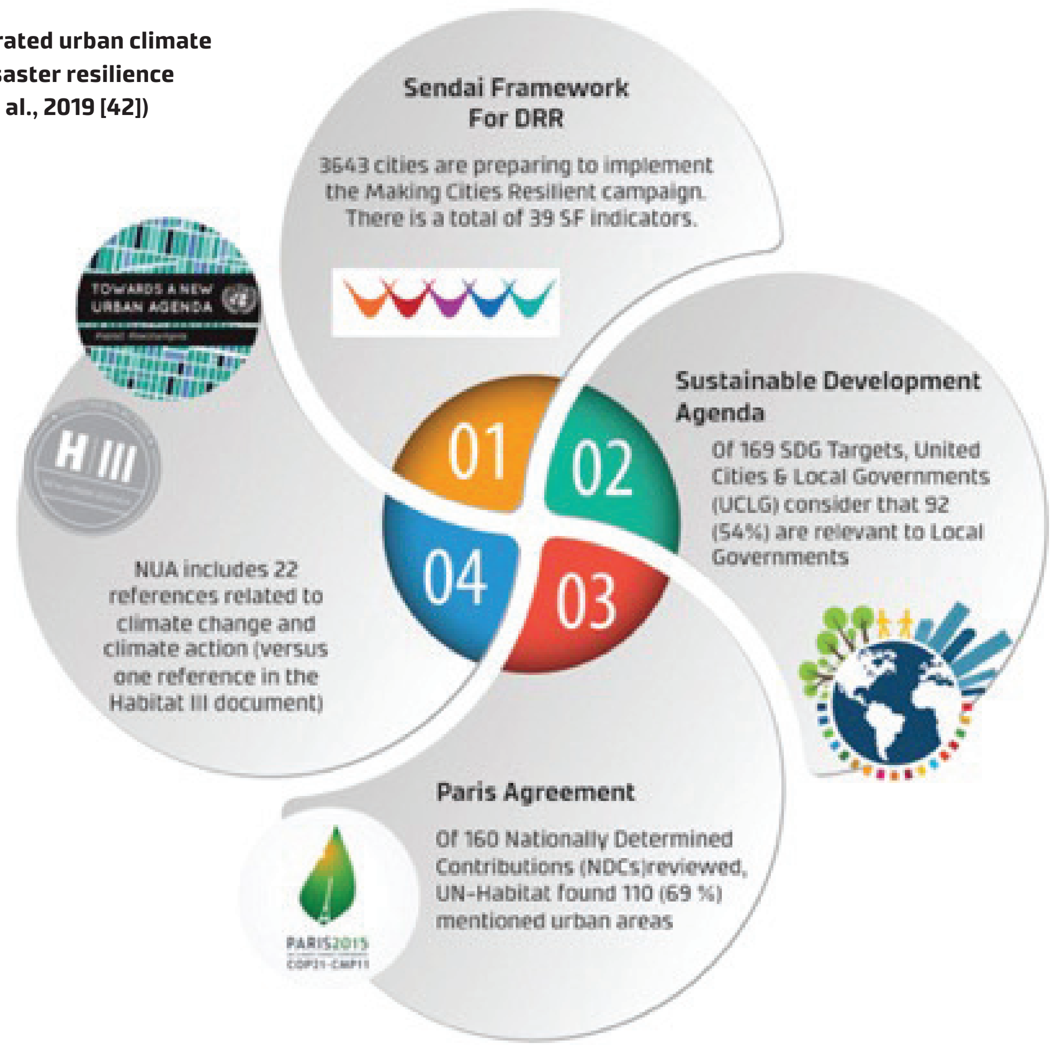
1. Introduction

This white paper proposes the next generation of disaster data infrastructure, which includes both novel and the most essential information systems and services that a country or a region can depend on to successfully gather, process and display disaster data to reduce the impact of natural hazards.

As a result of climate change and global warming, the frequency and severity of extreme weather has been increasing all around the world. According to the Sendai Framework for Disaster Risk Reduction, between 2005 and 2015 over 700,000 people lost their lives, more than 1.4 million were injured and approximately 23 million people were made homeless as a result of disasters. The severity of disasters is expected to surpass those of the past within the foreseeable future. In addition, the huge amount of disaster data collected from different sources could easily overwhelm and impair disaster risk reduction related applications, systems and their hardware platforms, especially in the case of large-scale disasters.

To build resilience and reduce losses and damages, the Sendai Framework prioritizes actions in the following four areas: (1) understanding disaster risk (2) strengthening disaster risk governance to manage disaster risk (3) investing in disaster risk reduction for resilience (4) enhancing disaster preparedness for effective response and to “Build Back Better” in recovery, rehabilitation and reconstruction. In particular, the framework emphasizes that governments should strengthen the utilization of media, including social media, traditional media, big data and mobile phone networks, to support nationwide disaster management and damage reduction. The availability of public access to multi-hazard early warning systems, disaster risk information and assessments should substantially increase by 2030. To assist in this dissemination to the public, governments should consider the needs of different categories of users and bespoke data dissemination to enhance disaster preparedness for effective response. In addition, satellite and in situ information, including geographic information systems (GIS), are needed to be fully utilized to enhance disaster analysis tools and to support real time access to reliable disaster data.

Figure 1: Integrated urban climate change and disaster resilience (Fakhruddin et al., 2019 [42])



It has been more than four years since United Nations member states adopted the Sendai Framework for Disaster Risk Reduction 2015-2030 on March 2015, at the Third United Nations World Conference on Disaster Risk Reduction in Japan (UNDRR 2015). Three other UN landmark agreements linking with the Sendai Framework were made in 2015 and 2016, including the Sustainable Development Goals (SDGs) (United Nations 2015), the Paris Climate Agreement (UNFCC 2015), and the Habitat III New Urban Agenda (United Nations Habitat III 2016) .

The Sendai Framework has strong links with the SDGs. The SDGs help end poverty, inequality, climate change and build resilience to disasters. Measuring and following progress in achieving these targets is fundamental in order to enable identification of priority areas where member states should focus their resources. Furthermore, the adoption of the New Urban Agenda Development Framework would be beneficial in completing the integration and provide a pathway for integrated climate change and disaster resilience improvements.

The safety of a country's economy and its citizens is reliant upon functioning critical infrastructure systems. An infrastructure system is defined as "a network of independent, mostly privately-owned, human-made systems and processes that function collaboratively and synergistically to produce and distribute a continuous flow of essential goods and services". Disasters amount to "the interdependent cascade of failure triggered by an extreme event that is exacerbated by inadequate planning and ill-informed individual or organizational actions" (Comfort, 2005) [33]. There are many methods available based on data availability both to better understand the criticality of any infrastructure system, as well as to support decision making and planning. For the modelling and simulations of interdependencies across critical infrastructure, these include: empirical approaches, agent-based approaches, system dynamic approaches, economic theory based approaches and network based approaches (Ouyang, 2014) [34]).

Based on the above discussion and the targets of the Sendai Framework, this white paper proposes the next generation of disaster data infrastructure natural hazards. Fundamental requirements of disaster data infrastructure include (1) effective multi-source big disaster data collection (2) efficient big disaster data fusion, exchange, and query, (3) strict big disaster data quality control and standard construction (4) real time big data analysis and decision making and (5) user-friendly big data visualization.

The rest of the paper is organized as follows: First, several future scenarios of disaster management are developed based on existing disaster management systems and communication technology. Second, fundamental requirements of next generation disaster data infrastructure inspired by the proposed scenarios are discussed. Following that, research questions and issues are highlighted. Finally, policy recommendations and conclusions are provided at the end of the paper.

2. Future scenarios and applications

2.1 Active hazard warning and emergency response system for living environments

With the rapid development of hazards early warning systems technology and sophisticated network infrastructure, many countries have already adopted standard Multi-Hazards Early Warning Systems (MHEWS) (WMO, 2019 [43] to inform the public that a natural hazard has either occurred or will happen. For example, after an earthquake strikes in the U.S., the official agency uses the Public Warning System (PWS) to broadcast CAP (Common Alerting Protocol) warning messages to inform people in the affected areas. This is done through different media channels, such as radio, television, short message service (SMS), smart phones, the internet or electronic signs. In addition, XML (Extensible Markup Language)-formatted CAP messages are sent to Active Emergency Response Systems (AERS) to automatically start the process of disaster risk reduction, such as stopping elevators at the closest floor, cutting off the gas, opening doors and windows, slowing down high-speed trains and putting factory machines into protection mode to avoid possible damages. The CAP standard is being leveraged in an increasing number of countries worldwide to greatly improve emergency alerting, so that everyone in harm's way receives timely and appropriate messages that enable them to act to protect their lives and livelihoods. The rapid alert notification system (RANS) using CAP is now widely used for rapid onset hazard information dissemination. RANS is a process that rapidly and widely communicates hazard warning information to the public from National Warning Service Providers. A RANS requires the establishment of emergency communication procedures and protocols to meet a country's unique attributes and culture, government structure and available technologies, mechanisms and infrastructure for emergency communication. RANS also ensures the implementation of the International Telecommunications Union (ITU) guidelines on "National Emergency Telecommunication Plans" (NETP). These guidelines should be adhered to in order to carry out best practice for efficient RANS (SPREP, 2019 [44]).

In the future, AERS will play an increasingly important role in disaster prevention and become ubiquitous in our living environment. AERS, with its simple functionality, will also support people with decision making. Customized warning messages will also be sent to different recipients based on their identities, spatial locations and the emergency levels of the disaster to assist people to be better prepared for natural hazards. Relevant services, such as health care and transportation, will also be integrated with AERS to support mass crowd evacuations and emergency medical services. Advances in big data, in particular space-based data, space technologies and corresponding data science can be used to improve all components of early warning and disaster response systems. Fusing different big data sources offers great potential to increase timeliness and granularity of a data-driven MHEWS. Space-based data and space technologies can be used to monitor transboundary hazards and the use of satellite telecommunications, enabling the monitoring of hazards in remote places. It is recommended that those involved in developing and managing EWS, whether international organizations or national and local organizations, develop a coherent data/digital strategy - a digital roadmap of how to include big data into the different MHEWS components, as well as into their internal processes.

Currently evacuation systems for people inside buildings only provide static information (i.e. evacuation maps, fire equipment locations and emergency contacts). In the future, AERS will provide them with dynamic evacuation instructions, real time disaster information and the progress of rescue operations, so that people can safely leave danger areas or find a safe place to stay. AERS will provide on-scene commanders with dynamic information relating to victims, such as their identities, spatial locations and physiological status, as well as the current status of the disaster. Visualization technology will be used to highlight severely affected areas and the status of both victims and responders. Intelligent decision making services will also be applied to support rescue operations and health care resource management. In addition, for first responders, AERS will provide not only victim information, but also indoor navigation capability and real time disaster information.

2.2 Crowdsourcing supported disaster information system

With the advance of social networks, research efforts are also focusing on using social media for disaster management. The key reason is that social networks could provide not only rich information but also near real time response. Taking the New Zealand Kaikōura Earthquake in 2016 as an example, social media provided a large number of photos with geographic information in a short period of time. This assisted official agencies in efficient focusing of their rescue actions. In the future, social networks will play a more important role in disaster management. Advanced technology, such as machine learning, big data analysis and image processing will be further investigated and developed so as to accurately classify disaster information. Data fusion is necessary to have a comprehensive view of threatened areas. The disaster information system should be able to classify the huge information collected from both social media and sensor networks to provide corresponding personnel with classification results.

The results of the classification would be filtered based on the user's role and responsibility. Taking a typhoon disaster as an example, the fire department may require detailed information of injured people, while other service providers may require different information sets. In addition, it could also provide useful information for people near disaster areas to take appropriate responses.

2.3 Disaster data quality assurance and control

Active Emergency Response Systems (AERS) are designed to perform a range of safety related tasks. Liu and Chu (2015) [1] explained that AERS are made feasible due to the advancement of four major technology domains; (1) advances in sensor and analysis technologies (2) emergence of Common Alerting Protocol (CAP), an XML-based data format for exchanging public warnings (3) development of platforms that integrate multiple communication channels enabled to receive CAP messages like Integrated Public Alert and Warning System (IPAWS)-OPEN (FEMA, 2017) [2] and (4) development of building information models and associated digital data exchange standards. Therefore, it is evident that an AERS is primarily a gigantic network of data models designed to perform real time data collection transmission and processing for decision making. Hence the success of an AERS heavily depends on its quality of data and information.

The quality of information provided by sensor data is therefore a critical concern in an emergency situation. For example, let us consider a medical sensing scenario. Medical sensors can be deployed on a patient's body to monitor health related parameters. This data is collected via wireless personal area network so that doctors can monitor the patient's health status in real time. The environmental sensors at a crisis site, such as smoke sensors, can detect fire in a building and also work with camera sensors to help determine a route so that patients can be rescued in a timely manner. The examples above show that the usefulness of sensors in case of emergency response are extremely diverse in practice and require timeliness, prioritization and sensing defect tolerance. Sachidananda et al. (2010) [3] argue that currently the quality of sensor data is usually addressed in isolation, by focusing on discrete data processing operations such as raw data collection, in-network processing (compression aggregation), information transport and sink operations for decision making. Furthermore, Qin et al. (2013) [4] argue that current research has primarily considered the functional aspects of distributed sensor systems, focusing on techniques to sense, capture, communicate, and compute over sensor networks, whereas in more complex and diverse sensor applications, non-functional application needs (such as timeliness, reliability, accuracy and privacy) become important. Thus, it is necessary to perform quality-aware sensor data management to make AERS an effective reality in the future.

AERS are heavily reliant on the quality of XML-based messages sent to various smart devices for automated processing of disaster management tasks. Efforts have been made throughout the last decade or so to standardize emergency-related data formats and use them to effect in emergency situations. For example, XML-based EDXL (Emergency Data Exchange Language) (OASIS, 2006) [5] messaging standards, including CAP (Common Alerting Protocol) (OASIS, 2010) [6], enable information exchanges between emergency information systems and public safety organizations, automatic reports by sensor systems to analysis centers, and aggregation and correlation of warnings from multiple sources. In the early stages, users have reported problems when implementing CAP messaging over multiple systems that include commercial satellite and terrestrial network technologies, such as C/L/X-Band, GSM, and CDMA in modes of voice and text (Waidyanatha et al, 2007) [7]. However, recent developments have enabled CAP to be used as the standard of emergency alerts. In recent years, CAP has been deployed in the U.S., Canada, Australia and parts of the Asia/Pacific region, including Taiwan and Japan. In the case of AERS, it is necessary that all smart devices used in the system are compatible with data standards and produce error free, quality messages in emergency situations.

2.3 Disaster data quality assurance and control cont...

Recent developments in building information models, smart/intelligent homes and the concept of smart cities will become important aspects in deploying AERS. A building information model illustrates the geometry, spatial relationships, geographic information, quantities and properties of building elements such as facility equipment that can be used for lifecycle management of buildings (Bazjanac, 2006) [8]. On the other hand, smart and intelligent homes and environments now offer us devices, applications and services for our comfort, convenience, and social connectivity, while also providing services such as monitoring elderly people and for healthcare (Chan, 2008) [9]. The concept of smart cities uses the availability of the ICT infrastructure, human capital and the plethora of generated information in the process of urban development and management (Caragliu et al, 2011) [10]. All the above-mentioned concepts are based on smart devices operating in networks to collect, store and process data online and offline. So far, few attempts have been made to extend these concepts to support emergency management. Therefore, the future challenge is to utilize the data and information from the emerging gigantic network of smart devices known as the internet of things (IoT) for the benefit of emergency response. However, the concern is the quality of data and information from the gigantic network of smart devices which is still in its infancy in the context of emergency management.

2.4 Disaster data standards and format

Disaster data response requires large amounts of data. However, data is produced by different organisations and stored in different formats, which leads to difficulties for agencies in terms of data sharing and interoperability. Inconsistent standards and disaster data formats are key challenges to collecting and using disaster data efficiently. The sharing of data resources in networked cooperation has become standard practice in some fields, particularly in the more economically developed countries. In many cases, researchers and their institutions experience considerable difficulty in data sharing, which acts as a barrier to developing and using shared data in new ways [11].

Disaster databases around the world play important roles in disaster reduction, such as the global historical disaster database DesInventar [12] (http://www.desinventar.org/), global level emergency disaster database EM-DAT (http://www.emdat.be/), space disaster events and loss database in the United States [13] (http://sheldus.org), the reinsurance company database, NatCatService and Sigma (http://www.swissre.com/sigma/).

Within both international and national projects, a range of earthquake impact databases have been created and are constantly being updated to calibrate seismic intensity attenuation models and vulnerability functions of elements at risk, in order to increase forecast estimates of losses in the emergency mode. The need for systematic data for disaster mitigation and prevention has been an increasing concern of both development and response agencies [14].

Different databases have different standards for disaster data management and storage. The organization that serves the corresponding database usually follows its own specific processes and protocols to collect and manage disaster data. For example, when criteria are set for entry on the EM-DAT disaster database, the data should meet at least one of the following three conditions: 1) 10 or more people deaths 2) 100 or more people affected/injured/homeless 3) declaring a state of emergency and/or an appeal for international assistance [15]. The EMA (Emergency Management Australia) disaster database requires that the disaster must cause more than three deaths, more than 20 injuries or at least AUD10 million in total losses. Although these databases have different disaster data standards, the data standards have a clear and concise definition which is important for the use of the disaster data. In addition, disaster data standards need to meet demand for disaster reduction activities. Many key components such as regional scale, time scale, accuracy of information, timeliness of information, and the comprehensiveness of a disaster should be considered in the standards. Different disaster databases comply with their respective standards, therefore cause difficulty in data sharing and interoperability. The Disaster Loss Data (DATA) project, under the umbrella of the Integrated Research on Disaster Risk (IRDR) programme, proposed a standard data collection system (Figure 2) which has been adopted by many countries since 2017 (Fakhruddin et al., 2019 [35]).

[1] **NatCatService (http://www.munichre.com)**

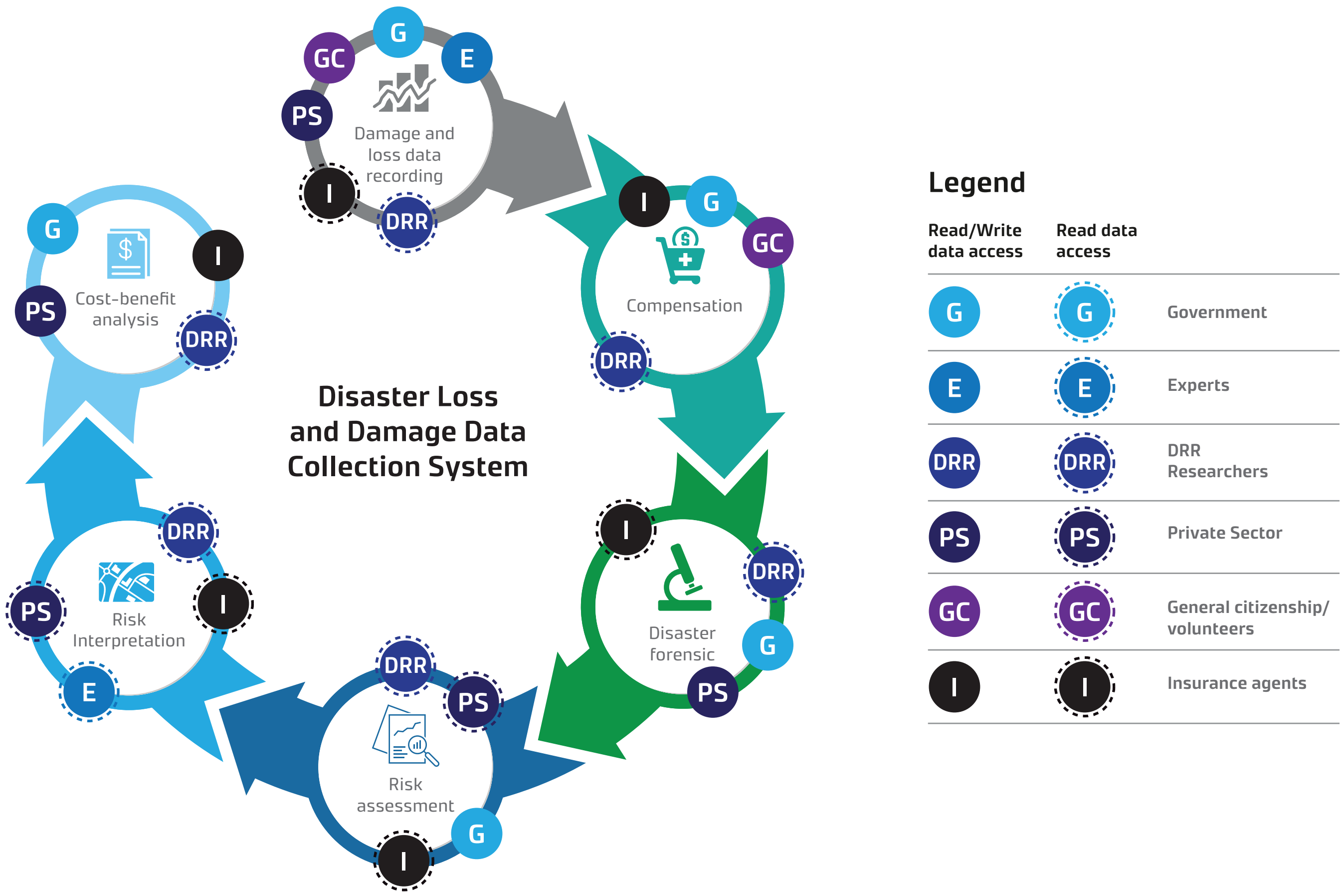


Figure 2 - Disaster loss and damage data collection system (Fakhruddin, et al., 2019 [35])

To be efficiently shared during the disaster mitigation, the data should follow agreed standards and be stored in standardized formats. The categories of disaster data can be divided by three ways as follows:

1. From a performance point of view, disaster data includes graphical data (e.g. topographic maps, plan, layout drawings, point notes, structure charts and images), text data (e.g. descriptive text, various statistical reports, attribute data related to geographical entities, sound).

2. From a data carrier point of view, disaster data includes traditional paper graphics, tables and documents. In addition, a variety of graphics, charts, and documents which are stored in a computer, and recordings and videos which are stored in cassettes or CD-ROMs.

3. From a data sources point of view, disaster data includes basic geographic data, ground observation data, ground survey data, model simulation data, historical data, census data, disaster report and integration. According to different data categories and data collecting processes, there are many regional or global standards due to the diversity and multi-disciplinary form of disaster data.

According to the different stages of disaster data acquisition, management, analysis and application, there are a range of standards and guidelines as shown in Table 1. For disaster classification, there are various types of disaster events such as flood, earthquake, drought, and hurricanes. Defining and classifying a disaster event is the most important aspect for disaster mitigation around the world. The standard such as Peril disaster classification proposed by IRDR (Integrated Research on Disaster Risk) provides a specified classification of disasters. Due to the variation in methods, the UN Sendai Framework defined a Data Collection Protocol to standardize disaster data collection. Impact database formats used for earthquake loss models calibration are based on the requirements of individual systems, such as PAGER, EXTREMUM, GDACS and others. For disaster data access and interoperability, the international organizations for standardization such ISO and OGC (Open Geospatial Consortium) have defined many standards on geographical data sharing and interoperability. This can be used for disaster data access and data exchange.

Table1. Examples of disaster related standards and guidelines

Standards category	Name	Organization	Scope	URL
Disasters classification	Peril Classification [16]	IRDR	global	http://www.irdrinternational.org/2014/03/28/irdr-peril-classification-and-hazard-glossary/
Data Collection Protocol	Data Collection Protocol presented in the 5th GP-DRR, 2017	UN Sendai Framework (UNISDR)	global	https://www.unisdr.org/we/coordinate/sendai-framework
Framework	National Disaster Recovery Framework	FEMA	regional	https://www.fema.gov/media-library/assets/documents/117794
OGC data services	WMS, WFS, WCS, Sensorweb, Opensearch-geo	OGC	global	http://www.opengeospatial.org/
Flood	FEMA Policy Standards for Flood Risk Analysis and Mapping	FEMA	regional	https://www.fema.gov/media-library/assets/documents/35313
WHO standards on disaster	Classification and minimum standards for foreign medical teams in sudden onset disasters	WHO	global	http://www.who.int/hac/global_health_cluster/fmt_guidelines_september2013.pdf
Disaster loss data sharing	Guidance for Recording and Sharing Disaster Damage and Loss Data	JRC	EU	http://publications.jrc.ec.europa.eu/repository/bitstream/JRC95505/lbna27192enn.pdf
ISO /TC211	Geographic information/geomatics	ISO	global	http://www.isotc211.org/
GEO	GEOSS framework	GEO	global	https://www.earthobservations.org
FAIR	Guidelines to improve the findability, accessibility, interoperability, and reuse of digital assets		global	https://www.go-fair.org/
CoreTrustSeal	Universal catalogue of core requirements for trustworthy data repositories	WDS	global	https://www.coretrustseal.org/

3. Big disaster data collection and transmission

Both rapid urbanization and continuous changes in climate have resulted in increased disaster risk. Whenever disasters occur, various incidents happen simultaneously (i.e. traffic congestion or a road block due to a landslide) which leads to a rescue and response mechanism. An essential part of the process is the collection and transmission of an accurate situation report collected by the right person or officials so that the time of response can start without wasting a single second. The role of data collection and transmission options is therefore essential in mitigating and managing the effects of disaster.

3.1 Available communications infrastructure

Disaster communications are largely dependent on existing Land Mobile Radios (LMR) and Mobile Cellular network infrastructure. Most of the access points associated with these are vulnerable to damage in disasters (i.e. cyclonic storms, earthquake). Tower collapses, antenna misalignments, power outages, cable damages are just a few of the examples of how the network can fail during disasters. Disasters like Hurricane Katrina destroyed essential infrastructure, rendering most wireless base stations, land line cables, central offices and data centers unusable. The non-availability of network severely hampers tactical operations. At the strategic level, the non-availability of relevant data from the disaster sites would be a great disadvantage.

Telecommunications holds the key to exchange of information and invariably is the most vulnerable link in the chain. Big data can be of assistance in this particular field as it can bring in disaster resilience in network planning. This is a fundamental component of disaster planning in order for information exchange to continue through a disaster situation, enabling operations to continue. Device-to-device communications or Proximity Services are required to create an ad-hoc mesh network from available working devices and terminals. This allows information exchange to continue, even during a network failure. A mobile-to-mobile mesh network can be created using the 3GPP ProSe standards (3GPP LTE Release 12) and mesh networking standards used by B.A.T.M.A.N or by Serval Mesh, which uses existing device-to-device communications to set up off-grid communication networks. This enables mobile devices to create a network using device-to-device communication such as ProSe, WiFi, Bluetooth etc. It is therefore feasible for the mesh network to function even during grid failure, allowing data and information to be transferred through optimal nodes.

The KoBo tool box is a free and open source tool for mobile data collection used by many developing countries. The toolbox was a joint initiative between UNOCHA, Harvard Humanitarian Institute and the International Rescue Committee. Emergency responders are able to develop forms quickly to collect needs information post event. Data can then be analysed and viewed in summary reports, graphs, tables and on maps, as data can be georeferenced and is able to be exported . KoBo toolbox is effective in gathering needs data in an emergency response situation however, it is not as useful for capturing past impact information. Rapid Pro, another tool, collects data via short message service (SMS) and other communication channels (e.g. voice and social media channels, such as Facebook Messenger, Telegram, WhatsApp) to enable real time data collection and mass-communication with target end-users, including beneficiaries and frontline workers. Rapid Pro was developed by UNICEF, allowing users to gather accurate real time information on vital areas such as health, nutrition, education, water and sanitation, and child protection—even in remote and hard-to-reach places—and allowing that data to reach those most in need . Rapid Pro also has the potential to link information with the KoBo Toolbox under the upgraded system Rapid Pro+. Rapid Pro however cannot view data geospatially.

3.2 Available data rate

Data requires collection and there is significant potential in developing big disaster data collection and transmission for remote monitoring. The objective of big disaster data collection and transmission is to develop an effective mechanism for data transmission between a monitoring device, which acquires data using DRR parameters, and a remote receiver, which allows storage of, and access to, the data - for instance, by a rescuer geographically distant from victims.

Smartphones with wireless communication are an excellent and readily available technology for data collection and transmission. One in every two people has one and the commonly used Android, iOS and Bluetooth applications can be easily used to facilitate big disaster data collection and transmission.

Along with the abovementioned, there are other operating systems such as Kindle. Android and other iOS devices are becoming increasingly useful and are a symbol of technological progress whereby the compilation of big data and transmission can be done in no time. Android/other iOS is already being used in providing solutions to its users, including weather conditions, hazard information, maps, routes, places, and early warning systems, therefore helping to analyze real and correct information. There are some developments focused on acquisition and visualization of risk and hazard parameters on smartphones. At a commercial level, there are Android/other iOS applications for various purposes and the big disaster data collection and transmission is clearly an area with plenty of promise.

Bluetooth uses short wavelength radio transmission for exchanging data wirelessly and without internet connectivity. A low cost, low power, technology for exchange of data information over short distances, it can be useful in big disaster data collection and transmission of information sharing to tablets and smartphones.

Well-built transmission technology ensures data is kept safe and can be used effectively, efficiently and quickly for taking the DRR measures required in big disaster data collection and transmission. Therefore, to network devices that have Bluetooth technology, Android and a transmission technology will provide an effective platform to collect big disaster data and transmit it as required.

4. Big disaster data processing

Big disaster data is processed and analyzed to develop perception, comprehension and projection of the emergency event as explained in section 3. Generally, data analytics are classified into four categories depending on their goal and purpose: descriptive analytics, diagnostic analytics, predictive analytics, and prescriptive analytics. In the emergency response phase, perception of the emergency related incident can be developed through descriptive analytics. To develop comprehension, both diagnostic analytics and prescriptive analytics can be used, whereas diagnostic analytics recognize the cause(s) for the incident and prescriptive analytics determine which actions need to be taken to manage the situation. Projection into the future can be developed through predictive analytics.

A primary disaster is the triggering event and a secondary disaster is the consequence of disaster. An example of this is an earthquake which causes a building to collapse. Authorities dealing with disaster risk reduction (DRR) must make decisions that range from one-off strategic decisions, to monthly and weekly tactical decisions and high-volume, high-speed front-line operational decisions. The requirement of time constrained decision support is an important parameter in order to mitigate the effects generated by the primary disaster. When the emergency alarm starts, the time-based mitigating measures can only be effective if the authorities are able to read data provided by the database and time-based mitigating measures implementation policy works.

4.1 Competence & hypothesis for disaster risk reduction data processing

Geographic Information Systems (GIS), mobile devices, cloud computing, social media, sensors, and cameras are now found everywhere and produce massive amounts of data. To extract the maximum value from dynamic and perishable disaster risk reduction data, authorities need to process data much faster and take timely action to save thousands of lives and livelihoods. Whether communicating to first responders and officials, offering proactive support, detecting and preventing risk, or managing the Internet of Things, real time decision making is essential. Responding in real time requires systems to make operational decisions automatically.

Currently, for example, there are three global systems for modeling damage and losses due to earthquakes in near real time. GDACS has additional options to both to facilitate and stimulate response. Extreme, in addition to assessing possible damage and loss, provides estimates of the first response resources required for response. Acute, Systematic and Robust Decision Management Systems combine proper rules and predictive analytics to render tailored recommendations. Event processing brings large-scale correlation and pattern detection to risk events, along with big disaster risk reduction data streams which are transferred in microseconds.

Event-based decision management systems enrich event-based disaster risk reduction data with traditional and big data sources to determine when, and why, a real time response might be required. Through leveraging decision engines based on DRR rules and analytics, decision management systems can determine what the best and most effective response is. For authorities to respond to early predictions or warnings in real time, they need to acquire or develop decision management systems to capture, filter and analyze data, and make decisions in real time. Such systems need to be able to rapidly determine that a response is required and intelligently determine both what, and when, the relevant and appropriate response should be. Authorities need to ensure that response mechanisms are delivered in real time, so more event-centric DRR decision management systems are required.

In a disaster event, the combination of DRR real time decisions and risk event processing delivers the core capabilities for building a real time, hazard or risk-based Decision Management System - correlating events, managing decision logic, embedding predictive analytics, optimizing results, and monitoring and improving decision making. Key features of the solution include certain competencies and hypothesis as shown in Table 2.

Table 2: Competence & Hypothesis for Disaster Risk Reduction Data Processing

Competence	Hypothesis
Event correlation, Disaster Risk Reduction rules, and predictive analytics in combination	Some real time response solutions focus on event correlation, on DRR measures, or on predictive analytics. With DRR Event Processing and Real time Decisions working together, the solution balances these capabilities, maximizing the flexibility and power of the decision making systems that can be built with it.
Scalability and flexible distribution	Extreme scalability and deployment ability for disaster risk reduction measures is required. The architecture, event detection and correlation are widely distributed and close to the event source. Therefore, monitoring is needed for structuring disaster risk reduction strategies. This improves responsiveness and contextual awareness while lowering latency. At any point, the Real time Decisions can be invoked or used to learn from the patterns detected due to their flexibility of deployment and suitability for a mutually shared approach.
Broad support for divergent environments	Externally managed DRR data, content, response and mitigation rules, and predictive analytic models must be supported. A wide range of complex and interpretable data can be handled.
Adaptability and robustness in the face of change	The overall solution is robust in the face of ongoing change. Event Processing allows new patterns and queries to be deployed to a live instance, while Real time Decisions allows similar changes to DRR rules and analytic models. Real time Decisions for Disaster Risk Reduction provides support for analytic models that are based on risk finding and hazard mapping, while automation of the full analytic lifecycle allows professionals to use hundreds of regularly updated DRR based analytic models in mitigating disaster affects.
Extensibility	DRR event processing provides new functions and capabilities that can be made available for pattern-matching analysis for disasters. The support of Real Time Decisions for external DRR rules and analytic models allows decision makers to achieve extensible prospects in DRR planning.

The first step in developing a DRR real-time response solution is to configure the event processing engine. This involves identifying the message streams in the environment, i.e., disaster locations. These message streams are fed into the optimization engine and are correlated with different disaster situations. Where a DRR event processing network is in place, various disaster related practices are defined to process these events. After the event processing engine is configured, it can be connected to the decision engine. This involves defining the options to be selected between the hazard and vulnerability measures that will allow the decision engine to choose the best option. Any DRR rules that constrain the choices are specified and analytic models are built to predict behavior and segment risk.

In case of any hazard increment, or risk enhancement, the DRR decision engine is invoked to determine the eligible choices and identify the best choice for taking mitigation steps, given the performance measures and analytics. These recommendations can be fed back to the disaster event processing engine or passed out as a response. The DRR decision engine then closes the loop by recording decision performance information. The decision engine automatically divides activity into test and control groups, and the competent authority of the decision decides how the activity will be split between the test and the control groups. This logic can be adjusted over time as necessary. Outcomes from DRR decision making are fed back into adaptive analytic models and used to monitor the overall performance of the system for taking DRR measures.

Social media has become a fast and efficient mechanism for disaster data collection. For example, if an earthquake were to occur, the reporting mechanism would be immediately activated as soon as any information was transmitted from any user who was not part of the disaster risk reduction system.

The features of social media used for disaster data collection and transmission are as follows:

- The most up to date data: The general public who are at the location of a risk or hazard can often provide a completely new set of information for data researchers. With every post, conversation and site or app visit, a user leaves behind pieces of information about themselves. The data transmitted is comprehensive - from simple demographic information through to robust details such as coping capacity, damage assessment, impact, risk information and early warning systems.
- Instant data: Every organization has historical data upon which data transmission is based, e.g., census data for establishing male to female ratios, or how many children live in an area. However, in the case of a disaster, instant data is required to understand the actual situation. Therefore, the social media data generated by people located in the disaster-affected area plays a vital role in situation analysis and data transmission, which is in turn critical to an effective disaster response and operation mechanism.
- Fine-grained data: Social media best represents people due to its ability to capture people's beliefs, attitudes and actions. No other source of data provides the same kind of granular detail of a person's life in terms of disaster reduction measures. Authorities can use social media to obtain personal information in an emergency which can save lives.

4.2 Big data processing for disaster management

Response and recovery phases of disaster situations require effective processing of historically collected information (government data, open data, linked data etc.), as well as big data from various channels such as sensors, satellites, crowd sourced information (social media feeds, photos, video) and cell phone GPS signals. Therefore, the challenge is real time integration of the archived information and big data streams through seamless interaction and collaboration with different platforms. Furthermore, it is necessary to have data integration and data fusion capabilities to integrate multiple distributed and heterogeneous data sources to produce more consistent, accurate, and useful information to support rescue operations (Haghighat et. al., 2016) [17]. The necessity of handling larger data volumes with different data formats such as structured, unstructured, and semi-structured data with high velocity constraints and limitation of human interpretation can overwhelm decision-makers. Therefore, advanced data querying and analytics such as machine learning techniques and in-memory computing will be required.

An important use of crowd-sourced data is event detection, where significant incidents are recognized (Saran et. al., 2017) [18]. In emergency response systems real time event detection should be performed quickly without compromising for precession. The rating of Seismological Surveys providing real time earthquake parameters in different earthquake-prone regions should be estimated in advance. Near real time global loss assessment systems should be calibrated taking into account using a Surveys rating.

The advancements in location inference techniques using geo-tagged social media data would provide useful tools for precise identification of locations (Laylavi, F., 2017) [19]. The information regarding events and their locations should be used to develop crisis maps for orchestrating rescue operations and planing recovery strategies. This geo-referenced information, plotted on maps, should be updated continuously as new information is received and events unfold (Beatson et al., 2014) [20].

The effective management of disaster events also requires the collaboration and coordination of a range of government decision makers, emergency response stakeholders and community-based non-government organizations. The availability of real time location-aware information, as well as the capability to effectively integrate and utilize available information with different autonomous agencies, is key to effective decision making and resource deployment to respond to crises (Fosso Wamba et al., 2012) [21]. In most situations, response teams have to make decisions based on incomplete and inaccurate information. This may be due to limited availability of data network resources. In the case of Japan's Tohoku earthquake, it was reported that the number of outages of communication facilities such as access lines and cellular base stations increased during the first 24 to 48 hours after the strike (ref). Therefore, mechanisms should be deployed to achieve high availability of the networks and systems. When developing emergency management systems, essential aspects to consider are: re-configurability of the network, network virtualization, and cloud-based systems that are not affected by the damaged infrastructure.

Predictive data analytics can play a wider role in disaster readiness and reduction phases. Natural hazards are extreme and unexpected phenomena resulting from natural processes of the earth and atmosphere. Therefore, the prediction of natural hazards is a major aspect in improving readiness. Though a massive amount of data is available across almost all disciplines related to disasters - such as geosciences, weather and medical insurance - the knowledge extraction from such massive data cannot always be performed by using standard statistical techniques. It is necessary to use new approaches such as recognizing relevant patterns of natural hazards through automated machine learning techniques to make predictions.

5. Big disaster data quality control

Big data is a new concept both to expert and the general public. Therefore general quality concepts related to Big data are still in their infancy. Hence the quality criteria (dimensions), the data quality management principles and methodologies are yet to be developed (Cai and Zhu, 2015) [22]. Big data has three main characteristics - volume, velocity and variety, where specific challenges are faced when assuring the quality of data. In this case, traditional data quality dimensions and frameworks become obsolete and new challenges emerge.

5.1 Challenges of disaster data quality control

The large volume and high velocity of crowd-sourced data requires an instant quality assurance process. Data from social media such as Facebook, Twitter, Instagram and various other communication channels such as Viber, WhatsApp multimedia messages and text messages are huge (The Harvard Humanitarian Initiative (2011) [23]). To use the collected data for decision making, information should pass quality checks pertaining to dimensions such as accuracy, currency and completeness. Therefore, the disaster response systems should have the capability to validate the data and information prior to its use for decision making. Sources such as GIS data and extra information pertaining to the device and the timestamp of the message can be used to perform validations. In addition, verification services can be designed to send collected data back to members of the public and request their feedback.

Due to the rapid changes happening in the environment during disaster situations, some big data remains valid only for a very short time (e.g., readings of thermal sensors and flood levels). If such data is not collected and used in real time, then people (including first responders) may receive outdated and invalid information as processing and analysis could produce misleading conclusions. Hence special capabilities should be built to handle the quality of time sensitive data. Further, due to multiple sources and the rapid changes, the necessity for entity resolution and record linkage will grow where matching of entities relies on good reference information for similarity scoring and linkage. Therefore, it is necessary to develop common reference domains and provide an environment for capturing and sharing disaster related terms, data element definitions and logical semantics.

The diversity of data sources brings verity of data types and complex data structures, therefore increasing the difficulty of data integration. Examples of data types are: Unstructured data (e.g. documents, video, audio) semi-structured data (e.g. software packages/modules, spreadsheets), financial reports and structured data from various databases. Semi-structured and instructed data play major roles in disaster situations and hence innovative quality criteria is required to assess them. Another aspect in using crowd-sourced data is the translation and interpretation of meaning in multilingual environments. The quality of data depends on how well the translations and interpretations match with the real-world disaster context to help save lives and assist relief operations effectively and efficiently. The presence of a wide variety of data and their sources requires joint efforts to assess the rating of sources for certain types of hazards.

Tapia et al. (2013) [24] argues that in a disaster context, crowd-sourced data may never meet the standards of quality required in situations such as search and rescue operations, while in others, such as resource and supply management, they may be useful as long as appropriately verified and classified. Goodchild and Glennon (2010) [25] reveal that many aspects of data quality problems surrounding disaster response require further research.

Identifying both the purpose of, and who will use, big disaster data is the most important component of big disaster data quality control and assurance. The methods of examining should be precise, directed by the purpose of the monitoring programme and the quality of disaster data that users need. The level of disaster data quality assurance will depend on its end users at the time of a crisis, and measures taken to check the reliability of disaster data for its intended purpose. For some monitoring groups, the main objective is identifying earthquake risk for the local community or school, where the focus is on earthquake awareness, rather than producing high quality disaster data on other types of disaster. However, groups that collect data to inform decision makers, or as part of an integrated monitoring programme with authorities, research organizations, regional bodies and state agencies must take measures to ensure the data is credible and reliable. When appropriate quality assurance and quality control measures are implemented, we can be confident that authorities' decisions are based on sound and reliable data.

Parameters for consideration in big disaster data quality control and assurance includes disaster data from contamination, illegal data filters, outlier detection, checking test assumptions with normal probability plots, diagnostic measures and GRUBBs testing. All these parameters should be taken into consideration for quality control and assurance for big disaster data. Data used for DRR projects for visualization, analysis, compilation, and sharing should meet a defined standard for quality. Data quality requirements vary from project to project and are determined by how accurate or complete a dataset needs to be, which in turn is based on how the data will be used. Further to this, data quality is influenced by technical, product, and location requirements.

5.1 Challenges of disaster data quality control cont...

Quality control and quality assurance is required in order to work through the issues identified, after which a quality control related report can be generated. All these issues can be implemented as tools for assessing big disaster data quality control and assurance.

GIS can be a useful platform in big disaster data management. GIS provides a platform for the management of geographic data and disparate documents (e.g. plans and photographs) necessary to meet emergency management requirements. GIS provides the capability to access information based on the geographic location to which it pertains, allowing users access to various types of information from the map display. GIS will be an efficient tool in data quality control and assurance and will help to build a real time picture for disaster risk mitigation exercises. Furthermore, the graphs and pictograms help analyze big disaster data quality. Dynamic data (e.g. camera feeds, weather, traffic, hospital status, automated vehicle location (AVL), incidents and sensors) provides situational awareness for decision support. Some points used in GIS provide better quality disaster data, therefore assurance of data can be gained up to a certain level.

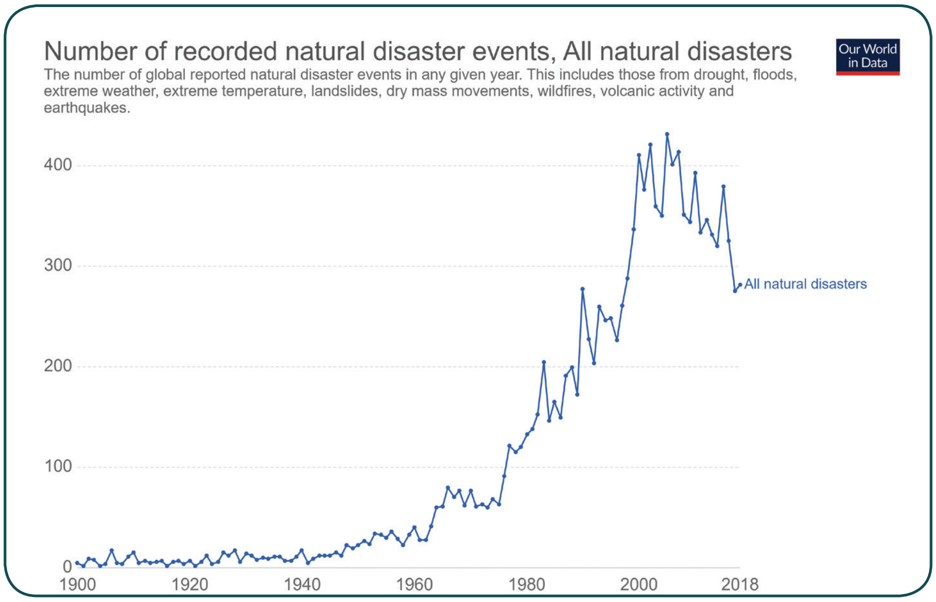


Figure 3a: Total damage cost from global disasters

Source: EMDAT: OFDA/CRED International Disaster Database, Universite Catholique De Louvain-Brussels- Belgium

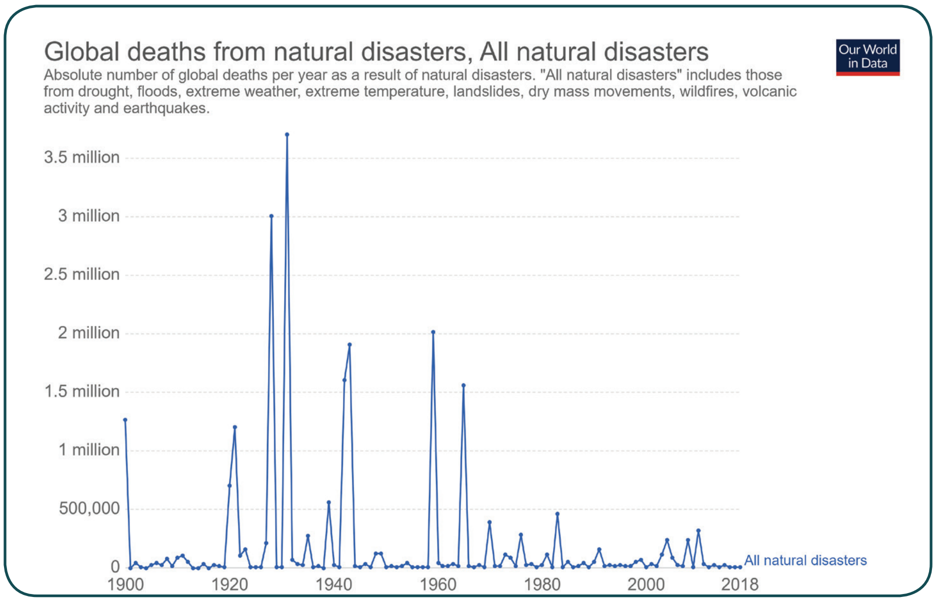


Figure 3b: Number of deaths from natural disaster

Source: EMDAT: OFDA/CRED International Disaster Database, Universite Catholique De Louvain-Brussels- Belgium

Figure 3a and 3b, wherein the first figure shows the damage assessment and next figure shows mortality data, can be used to analyse disaster data quality. These parameters can be integrated to obtain better quality and assurance of disaster data.

The user and system design requirements from the aspect of big disaster data quality control and assurance needs an interface to control big disaster data documentation. The interface document describes the internet capabilities of two separate platforms in information dissemination and big disaster data quality control and assurance. The interface document in disaster risk reduction and big disaster data also describes the interaction between two different sources and how to gather the correct information in one place. This interface is also similar to other communication interfaces, wherein the role of a user and the system, software component - or two software components – and a hardware device are interrelated. This class of document is typically used where complex interfaces exist between components for disaster risk reduction and are being developed by different people in different locations. It is jointly prepared by the interfacing groups.

The big disaster data requirements describe what the interface is to achieve, together with any hindrance to its design as follows:

- Identification of the disaster interfacing systems/ sub-systems
- The reason for the disaster interface's existence, including that the user requirement is satisfied
- A description of what the interface does for disaster risk reduction
- Specification of the information to be exchanged for mitigating disaster risk
- Timing and sequencing constraints pre-disaster as well as post-disaster
- Capacity and performance requirements for disaster risk reduction
- Requirements for communications protocol standards compliance to mitigate effects of disaster
- Identification of any safety requirements discovered in a Disaster Interface Hazard Analysis.

Data cleansing is a useful method for quality assurance in disaster data. Data profiling techniques can be used to identify the problems in historically collected disaster-related data sets such as weather data, loss data etc. Data profiling is the process of examining the data available in an existing data source and discovering statistics and information about that data. Profiling can unveil basic data quality problems (the number of null values, duplicate records, outliers etc.), data types and the most frequent patterns of data values. These discoveries can be translated into constraints or rules that are then enforced in the data cleansing phase. Data profiling methods can also be used to uncover heterogeneities (syntactic, structural, and semantic) in multiple data sources and provide guidance on how to integrate data sets (Naumann 2013) [39].

Disaster data quality is still in its infancy and not many attempts have been made to develop a data quality management framework for disaster management systems. However, in the context of business information systems, and in many other disciplines such as medical and healthcare, data quality is a well matured area with prominent frameworks and methodologies (Batini, 2009) [38]. Therefore, more research should focus on disaster data quality requirements, specifically focusing on the Active Emergency Response Systems (AERS) scenarios that will be the future of disaster management. As explained in section 4.2, some of the data used in disaster management are big data and new and innovative approaches are required to develop data quality management strategies. But traditional quality management approaches are useful for other data such as previously collected data (demographic data, geographical data, data from various government departments etc.) and real time data (sensor data, GPS data, alerts, instructions etc.)

5.2 Disaster data interface between different entities

The design of a big disaster database describes how the interface will be implemented for disaster risk reduction. For example, in the case of a control room interface, the following technical details are provided. A description of the control room communications protocol may include:

- Message format for hazards, risk, vulnerability and description including user error messages, user information messages and inter-process messages
- Message component names (like earthquake, tsunami etc.)
- Message initiation (heat wave or cold wave started)
- The processing of message interruptions. Fragmentation and reassembly of messages
- Error detection, control and recovery procedures
- Disaster data synchronization, including connection establishment, maintenance, termination and timing and sequencing for the disaster risk reduction mechanism
- Disaster risk reduction data transfer rate
- Data transmission services including priority and grade
- Disaster data security including encryption, user authentication and auditing
- Error codes in disaster data

Figure 4

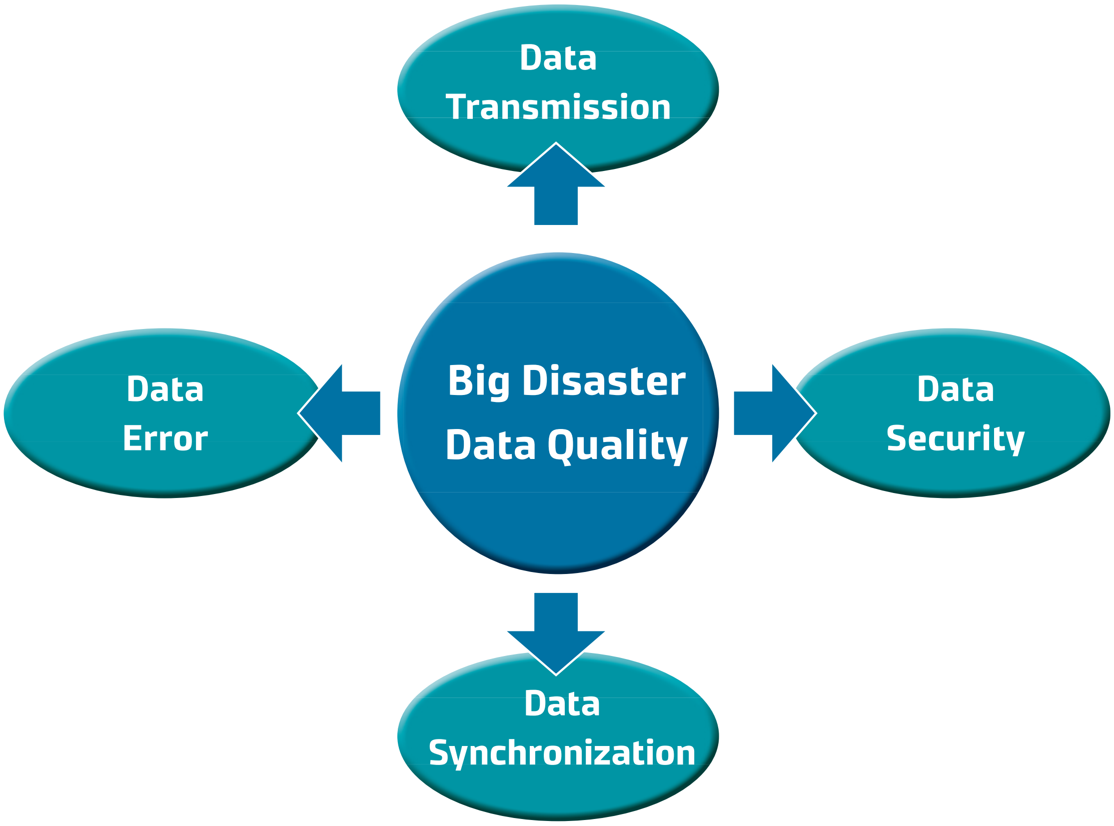


Figure 4: Interrelation of disaster data quality

Figure 4 above shows the impact of transmission, errors, security and synchronization in maintaining big disaster data quality which can be processed in terms of other parameters. Data quality is a perception or an assessment of whether the data is able to serve its purpose in a given context. Various aspects of big disaster data quality include other parameters of consideration: Accuracy, completeness, update status, relevance, consistency (across data sources), reliability, appropriate presentation and accessibility. In the following section, disaster data visualization and its applications and limitations are discussed.

6. Big disaster data visualization

Big Disaster Data visualization technologies can be both powerful and easy to use, thus allowing decision makers to quickly and easily understand, articulate and share insights across the organization to others. The main objective is easy sharing of information and to channelize the response mechanism as rapidly as possible.

More often than not, dealing with disasters has historically focused on emergency response. However, from the beginning of the 21st century it has become increasingly recognized that disasters are not natural (even if the associated hazard is). Only by reducing and managing conditions of hazard, exposure and vulnerability can losses be prevented and the impacts of disasters alleviated. As it is not possible to minimize the severity of natural hazards, the main task for reducing risk lies in mitigating vulnerability and exposure. Reducing these two components of risk requires identifying and reducing the underlying drivers of risk, which are particularly related to poor economic and urban development choices and practice, degradation of the environment, poverty and inequality and climate change. These drivers create and exacerbate conditions of hazard, exposure and vulnerability. Addressing these underlying risk drivers will reduce disaster risk, lessen the impacts of climate change and consequently, maintain the sustainability of development. DRR is a part of sustainable development, so it must involve every part of society, as well as government and non-governmental organizations and the professional and private sectors. Big disaster data visualization creates a culture of prevention and resilience for implementing disaster risk reduction mechanism. Moreover, big disaster data visualization will also allow decision makers to build strategy for multiple, cascading and interacting hazards.

6.1 Data visualization for disaster management

Trying to discover relationships and understand risks and hazards can be difficult, as data can be comprised by thousands to millions of variables. Big disaster data visualization has become the de facto standard for DRR. Data visualization tools in DRR have been important in democratizing data and analytics and making data-driven insights available to disaster professionals throughout an organization. Data visualization tools are typically easier to operate than traditional statistical analysis, which is not software-based and based on partially incorrect data. This has led to an increase in hazard mapping, vulnerability analysis and risk reduction implementing data visualization in isolation, without IT support. Data visualization software also plays an important role in big data disaster visualization and advanced analytics projects. Big disaster risk reduction measures resulted in massive accumulations of data in relation to various hazards and related risks and disasters during the early years of big data advancement. However, by utilizing visualization tools, it was a way to quickly and easily get an overview of the data.

The greater the quantity and type of disaster data collected, the more we need to experiment with how to better present the data in formats which can be understood. The requirement in mitigating disaster risk is to start with a plain sheet of paper and then experiment with custom visualization. Even minute details can enhance basic charts to reveal further disaster risk reduction measures.

The ideal process in creating big disaster data visualization is to amass all the disaster related data in a specific tool, pick it from unconventional sources and frame it so that the entire task can be completed within a couple of clicks. Simplified solutions are required to solve big disaster data collection and transmission methodologies. In order to achieve clarity, big disaster data needs to be compared, synthesized and concluded, then transmitted to users as a single piece of information. Big disaster data-driven doesn't mean that it is absolutely correct, because data and the tools that collect it are man made. DRR data is not completely factual, but can be viewed as evidence that filters reality in a very subjective way. The intention is not to get a black and white result, but rather visualized data in a satisfactory form that delivers meaning to people reading that information.

Figure 5

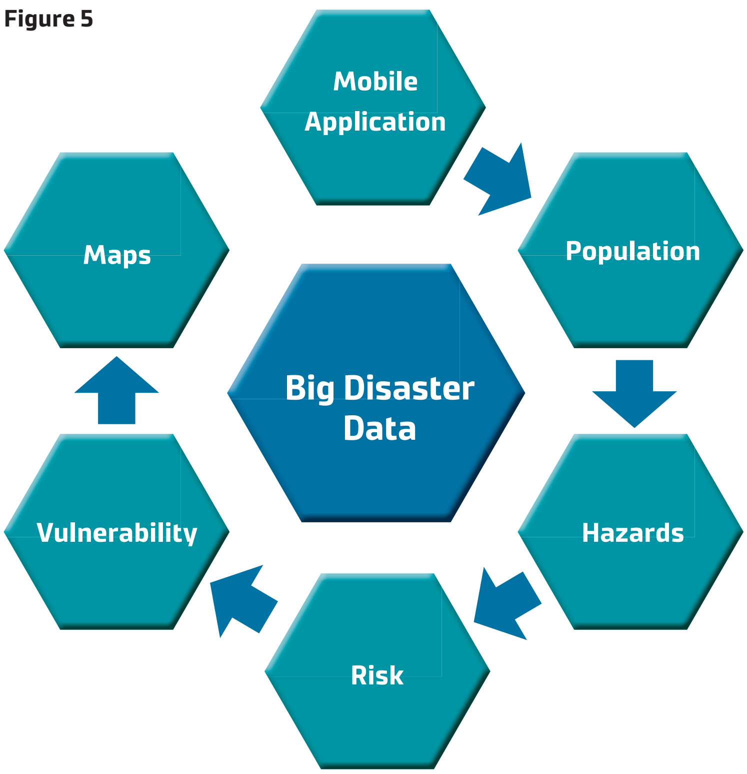


Figure 5: List that is required to fetch big disaster data

As shown in Figure 5, the data visualization approach is to analyze, compile and combine parameters such as the use of Android/other iOS, population data, hazard analysis, exposure mapping, vulnerability studies and maps to form a visualized big disaster data. This is the same approach used in preparing traffic congestion data if the number of vehicles on roads, the travelling population and nearby residential population is known, as shown in Figure 4. The important parameters always start with a hazard analysis and exposure mapping, along with a vulnerability study.

Big disaster data visualization is central to advanced analytics for similar reasons. When a data scientist is writing advanced predictive analytics or machine learning algorithms on hazards, vulnerability, risks and disasters as shown in Figure 7, it becomes important to visualize the outputs to monitor results and ensure that models are performing as intended. Visualizations are generally easier to interpret than numerical outputs. Most of today's data visualization tools come with connectors to popular data sources, including the most common relational databases, such as Hadoop and a variety of cloud storage platforms. The visualization software pulls in data from these sources in the case of normal data management and applies a graphic type to the data or in broader disaster data.

Figure 6

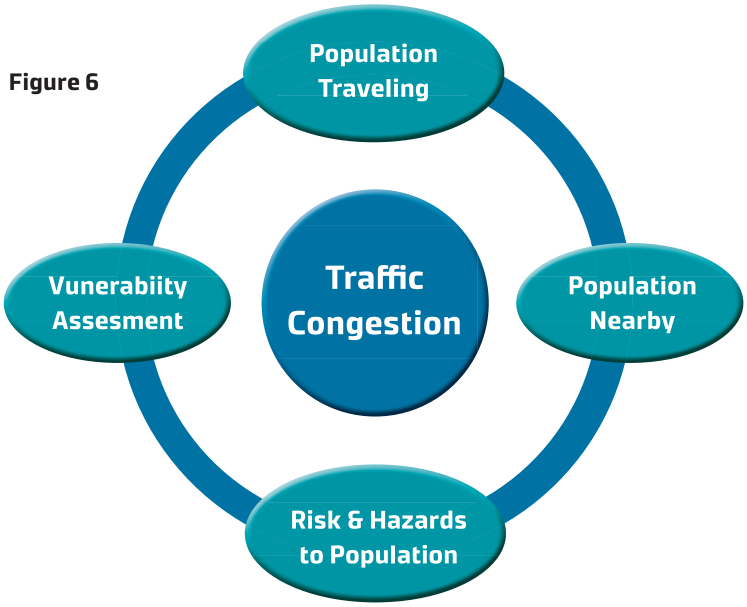


Figure 6: Data visualization for traffic congestion

Figure 7

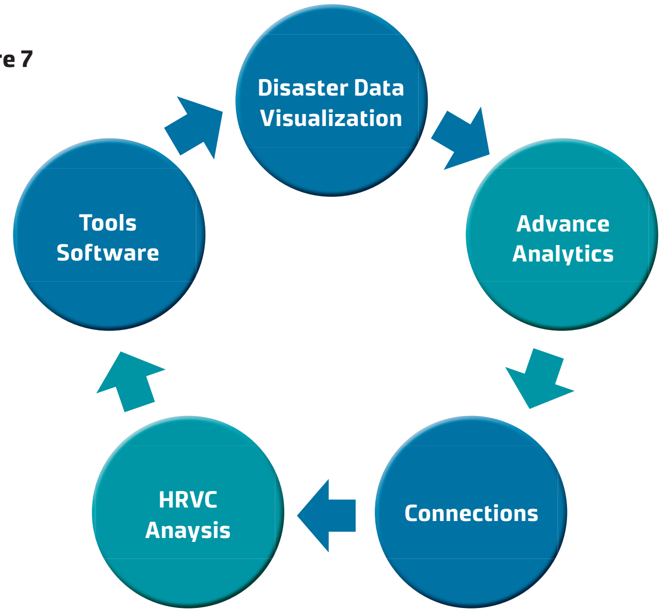


Figure 7: Big disaster data visualization cycle (HEVC*- Hazard, Exposure, Vulnerability and Capacity)

Big disaster data visualization will provide the best of both worlds; specific, precise, uniform analytics that can be relied upon, along with ease-of-use and speed if changes are needed. Disaster data visualization will explore parameters including:

- Search for response, recovery and rehabilitation points and the ability to access data quickly due to guided navigation
- Analyze disaster data, anywhere with instant mobile
- Highlight one visual to automatically see disaster risk-related information in the others
- Add new data and deliver, on-the-go, intelligent updates
- Use pre-built data connections to load and integrate data from a wide variety of sources
- Combine information to uncover new insights
- Capture insights and add comments to create visual stories
- Collaborate with research teams by sharing calibration parameters of loss models

Big disaster data visualization is an important step in mitigating disaster risk because of the way the human brain processes information. Using charts or graphs to visualize large amounts of complex data is easier than sorting through spreadsheets or reports. Big disaster data visualization will be a fast, feasible, easy way to deliver concepts in a universal manner - and we can experiment with different disaster-related scenarios, either natural or man-made by making slight adjustments.

Sense-making (also called data analysis) and communication is the graphical display of abstract information for two purposes in data visualization and understanding. Important stories live in the collected data and data visualization is a powerful means to discover and understand these stories and present them to others. Whether data incorporates hazards findings, vulnerability, risks or gap analysis, capacity building, or anything else, and even though it does not pertain to the physical world, it can still be displayed visually. However, to generate such outcomes we must find a new way to give form to that which has none. This translation of the abstract into physical attributes of vision (length, position, size, shape, and color, to name a few) can only succeed if we understand a bit about visual perception and cognition. In other words, to visualize disaster data effectively, we must follow design principles that are derived from an understanding of human perception in disaster risk reduction.

Generally, data visualization features key relationships between quantitative values. It can also display certain relationships that are not quantitative in nature and can derive some unique features. For instance, the connections between people near a natural disaster site or affected by suspected terrorists on social networking sites such as Facebook or WhatsApp can be displayed using a node and link visualization. The mobile technology provides a gateway to get instant disaster data by using the application which is inside that mobile device. Therefore big disaster data visualization can create an efficient and effective platform in helping to preserve the lives and property of people by using readily available technology.

The bigger challenge of big disaster data collection is that we rely on grass-roots workers for data sharing, who are never sensitized to, or trained in, that particular hazard or risk data. Intellectuals want to analyze disaster risk data and simulate models, but have no control over policies and mechanisms that exists to collect them. United Nations Office for Disaster Risk Reduction (UNISDR), KoboToolbox (A suite of tools for field data collection for use in challenging environments), EM-DAT (Platform for maintain international disaster database) and several country-specific mechanisms currently exist for post-disaster needs assessments by estimating several parameters. All these platforms are used for short and long term loss estimation but have a limited scope in terms of future prediction and recovery.

6.2 Information systems design for disaster management

All four phases of disaster management; readiness, response, recovery and reduction (MCDEM, 2007) [26] involve heavy use of data and information that belongs to all three categories of data; structured data, semi-structured data and unstructured data (Batini and Scannapieco 2006) [27]. The data used during disaster situations can be classified as big data. One of the world's leading research and advisory companies, big data is defined when the volume velocity and verity of data is in high scale (Kailser et al, 2013) [28]. For instance, data supporting the response phase of a disaster generates significantly large volumes of data from sensor networks, satellites, social media, (Prasanna and Huggins, 2016) [29]photos, videos, GPS signals from cell phones and other multimedia devices, as well as results of near real time loss simulation. The generation of such data occurs in shorter period of time (Yang, Prasanna and King, 2009) [30]. Furthermore, the data focuses on multiple variables such as weather, medical related, supplies, relief, warnings, traffic, transport.

Creating situational awareness (SA) is the foremost task in an emergency response phase. Situational awareness is defined as "The perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future"(Ref). In order to develop SA, data should be visualised in an effective and efficient manner so that it supports three levels of knowledge: (1) perception (2) comprehension and (3) projection of the incident, so that the incident commanders can plan and carry out the rescue operations. In order to develop each level of knowledge, system interfaces are required for data visualization.

• Interfaces for Level 1 SA - Perception

The first step in achieving SA is to perceive the status, attributes and dynamics of relevant elements in the environment. Lack of basic perception on important information can easily lead to an inaccurate picture of the situation. These interfaces will support an end user to maintain a global picture relevant to a particular role at any given time during the incident (Yang et al, 2009) [30]. Thus, as shown in Figure 8, with these types of interfaces incident commanders will be able to have a high-level, summarized overview of the situation.

Figure 8

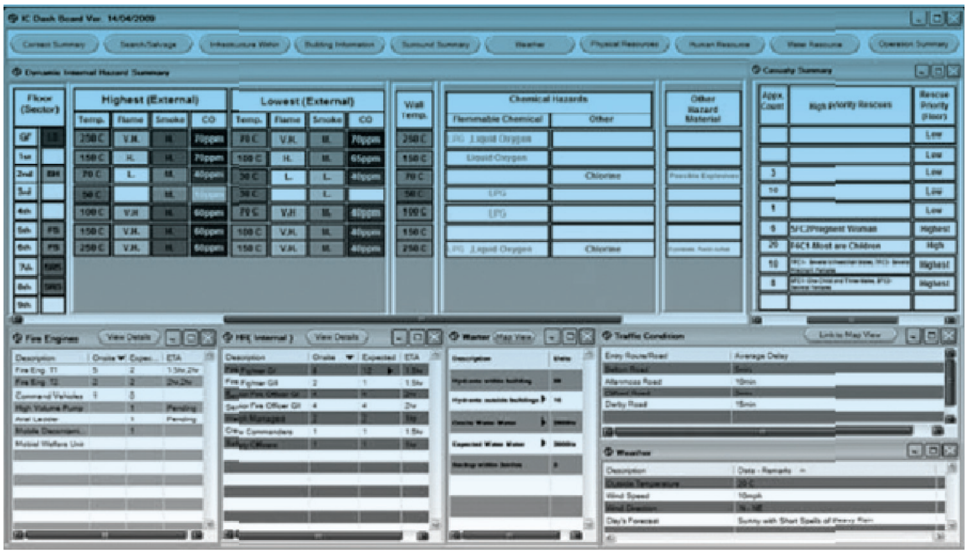


Figure 8: Interface supporting Level 1 SA of a IC: 'Dash Board'. [30]

• Interfaces for Level 2 SA - Comprehension

At this level, information obtained through observation is combined and interpreted. Rather than presenting a set of isolated information, mostly via numbers and text as in the perception level, as shown in the Figure 9, with this type of interface dynamic information is meaningfully integrated with static information using graphical presentations. It provides an appropriate level of comprehension of the situation at any given moment in time to further improve the SA.

Figure 9

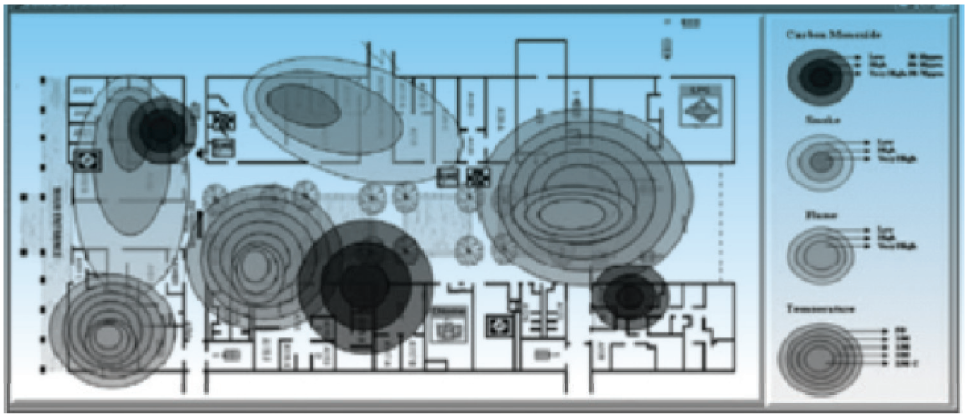


Figure 9: Interface supporting Level 2 SA of an IC. [30]

• Interfaces for Level 3 SA - Projection

By developing these interfaces, projection of future events is supported by providing the incident commanders with information on current and past trends on various situational parameters. Together with Level 1 - Perception, and Level 2 - Comprehension interfaces, Level 3 - Projection interfaces, as shown in Figure 10, can provide a higher level of SA for incident commanders in making difficult predictions with confidence at any given moment during the emergency.

Figure 10

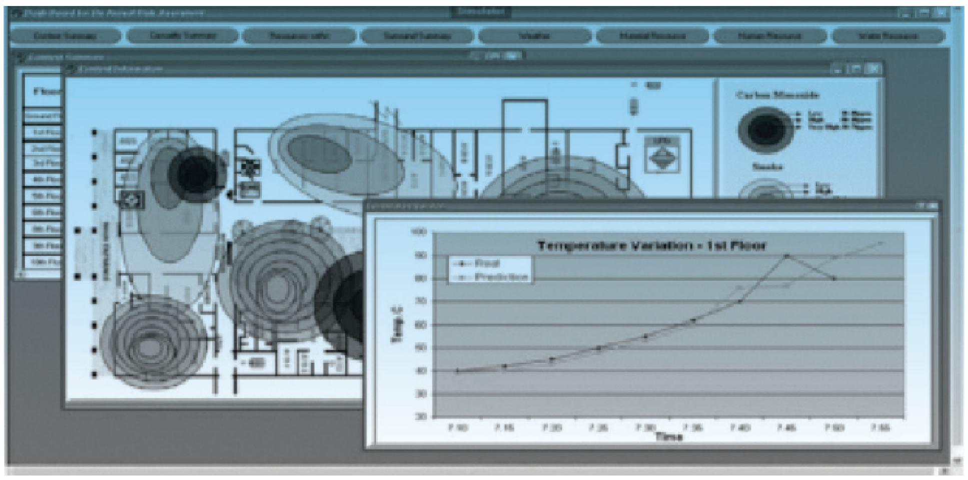


Figure10: Interface supporting Level 3 SA of a IC. [30]

The system design requirements for data visualization in each level explained above will differ, depending on the nature of the emergency or disaster. For example, in case of a fire-fighting scenario, nearly 350 different information interfaces have been proposed for the use of four fire-fighter roles (Prasanna et al. 2013) [31]. Thus, depending on the nature of the disaster, ground level data visualization requirements may change.

Visualization of disaster related data positioned on geographical data (area maps, building maps etc.) is an important aspect for all phases of emergency management. GIS technology provides the capability to map and analyze hazards of all types and visualize their potential impacts. Therefore, emergency management systems should be capable of producing interactive maps such as vulnerability, operations, logistics, tactical, air deployment, transportation and incident prediction maps which can be used in all areas of emergency management phases. An example of where visualizations in maps can be applied is when there is a need to identify information about trapped persons, medical resources, damaged buildings, closed roads, and the availability and whereabouts of specific needs such as food, water and shelter (Beatson, et al., 2014) [20].

Visualization of crowd-sourced information is another important aspect that should be considered in the disaster response phase. With the emergence of web 2.0 tools like Twitter, Facebook and YouTube, a massive amount of data is exchanged during disasters (Bruns and Burgess 2012) [32]. The majority of such crowd-sourced information is generated by the general public (Harvard Humanitarian Initiative 2011) [23] and is vital to developing real time live maps to produce and visualize an overall perspective of what usually is a complex and often rapidly changing environment (Beatson, et al., 2014) [20]. Organizations such as Volunteer & Technology Communities (V&TCs) need to visualize hot-spots of activity within a short timeframe in order to mobilize large numbers of internationally-dispersed volunteers, so they are able to collaboratively solve informational and logistical management issues.

7. Policy Recommendations

7.1. Future Scenarios

a. **Hazard warning and emergency response system:** It is recommended that those involved in developing and managing EWS, whether international organizations or national and local organizations, develop a coherent data/digital strategy, a digital roadmap of how to include big data into the different MHEWS components and into their internal processes. A major challenge is to help developing countries to better take advantage of the standard framework for an end to end multi-hazard early warning system.

- **Crowd-sourcing information:** Data fusion is necessary to develop a comprehensive view of threatened areas. The disaster information system should be able to classify the extensive amount of information collected from both social media and sensor networks to provide corresponding personnel with classification results.

- **Data quality assurance and control:** IoT is creating vastly more data at much lower cost. This does result in significant amounts of good quality data, but also an amount of poor quality data. Process can be developed to ignore the erroneous data and utilize the good quality data. As more and more sensors are deployed this process gets easier.

- **Disaster data standards and format:** Standardization of disaster loss data quantification can identify gaps in risk assessment, simultaneously improving disaster risk information which could provide common guidelines on methods of hazard, exposure and vulnerability assessments. It is vital to improve partnerships between intra-government agencies, academic, private sector, NGOs and insurance authorities at the global, national and local levels for data sharing and monitoring the Sendai Framework and its Global Targets.

7.2 Big disaster data collection and transmission

b. **Communications infrastructure:** Device-to-device communications or Proximity Services will be required to create an ad hoc mesh network out of available working devices and terminals so that information exchange can continue even during a network failure.

c. **Data transmission:** A well-built transmission technology is required in big disaster data and transmission to ensure data is kept safe and can be used effectively, efficiently and quickly for information dissemination during disaster periods. Thus, devices such as Bluetooth technology, Android/others and a transmission technology will provide everything to collect big disaster data and transmit it accordingly when it is required.

7.3 Big disaster data processing

d. **Data processing:** In the emergency response phase, perception of the emergency related incident can be developed through descriptive analytics. To develop comprehension, both diagnostic analytics and prescriptive analytics can be used where diagnostic analytics recognizes the cause of the incident and prescriptive analytics determine which actions need to be taken to manage the situation. Future projection can be developed through predictive analytics.

e. **Big data processing for disaster management:** It is necessary to use new approaches such as recognizing relevant patterns of natural disasters through automated machine learning techniques to make predictions.

7.4 Big disaster data quality control

f. **Data interface between different entities:** Data from social media such as Facebook, Twitter Instagram and various other communication channels such as Viber, WhatsApp multimedia messages and text messages are huge. In order to use them for decision making, the information should pass the quality checks pertaining to dimensions such as accuracy, currency, completeness etc. Therefore, the disaster response systems should have the capability to validate the data and information before it is used for decision making. Since big data has three specific characteristics - volume, velocity and variety - specific challenges are faced when assuring the quality of data. Advanced data querying and analytics such as machine learning techniques will be required to process the vast amount of available data and highlight what is significant.

7.5 Big disaster data visualization

g. **Data visualization:** In order to develop situational awareness data should be visualized in an effective and efficient manner so that it supports three levels of knowledge: (1) perception (2) comprehension and (3) projection of the incident so that incident commanders can plan and carry out the rescue operations. In order to develop each level of the knowledge system, interfaces are required for data visualization.

8. Conclusion

This white paper discusses the next generation disaster data infrastructure from four different aspects: Disaster data collection, disaster data processing, disaster data quality control and disaster data visualization. In disaster data collection, sensor data and crowd-sourced data, as well as results of near real time loss simulation, should be considered together to deliver a comprehensive view of threatened areas. In addition, the availability of real time location-aware information, as well as the capabilities to effectively integrate and utilize available information with different autonomous agencies, is key to effective decision making and resource deployment in crisis response. Predictive data analytics, including the results of near real time loss simulation, can also play a wider role in the disaster readiness and reduction phases. It is necessary to use new approaches, such as recognizing relevant patterns of natural hazards through automated machine learning techniques etc., to make predictions. Governments should also consider a cross reference platform for capturing and sharing disaster related terms, data element definitions and semantics. Further, when appropriate quality assurance and quality control measures are implemented for big disaster data, we can be confident that competent authorities decisions are based on sound and reliable data. Visualization of disaster related data positioned on geographical data is important in all phases of emergency management. GIS technology provides the capability to map and analyze hazards of all types, simulate loss estimations in near real time and visualize their potential impacts. Thus, emergency management systems should be capable of producing interactive maps such as vulnerability, operations, logistics, tactical, air deployment, transportation and incident prediction maps to be used in all areas of emergency management phases. Finally, governments should consider all the above-mentioned aspects and the needs of different categories of users and data dissemination in order to reduce the impact of natural hazards.

Reference

1. Liu, J. and E. Chu. Dependability of active emergency response systems. in Proc. DEPEND. 2015.

2. FEMA. Integrated Public Alert & Warning System Open Platform for Emergency Networks | FEMA.gov. 2017 Dec Available from: https://www.fema.gov/integrated-public-alert-warning-system-open-platform-emergency-networks.

3. Suri Vinay Sachidananda Abdelmajid Khelil, N., Quality of Information in Wireless Sensor Networks: A Survey. 2010.

4. Qin, Z., et al., Quality-Aware Sensor Data Management, in The Art of Wireless Sensor Networks: Volume 1: Fundamentals, H.M. Ammari, Editor. 2014, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 429-464.

5. OASIS. EDXL-DE: Emergency Data Exchange Language Distribution Element, V1.0. 2006; Available from: http://docs.oasis-open.org/emergency/edxl-de/v1.0/EDXL-DE_Spec_v1.0.pdf.

6. OASIS. CAP: Common Alerting Protocol Version 1.2. 2010; Available from: http://docs.oasis-open.org/emergency/cap/v1.2/CAP-v1.2-os.html.

7. Waidyanatha, N., G. Gow, and P. Anderson, Hazard Warnings in Sri Lanka: Challenges of Internetworking with Common Alerting Protocol. 2007.

8. Vladimir, B. Virtual Building Environments - Applying Information Modeling to Buildings. in European Conference on Product and Process Modeling in the Building and Construction Industry (ECPPM) 2004. Istanbul, Turkey.

9. Chan, M., et al., A review of smart homes-Present state and future challenges %J Comput. Methods Prog. Biomed. 2008. 91(1): p. 55-81.

10. Caragliu, A., C. Del Bo, and P. Nijkamp, Smart Cities in Europe. Journal of Urban Technology, 2011. 18(2): p. 65-82.

11. Uhlir, P.F., et al., Open Data for Global Science. Data Science Journal, 2007. 6: p. OD36-OD53.

12. Marulanda, M., O. Cardona, and A. Barbat, Revealing the socioeconomic impact of small disasters in Colombia using the DesInventar database. Vol. 34. 2010. 552-570.

13. Tschoegl, L., An Analytical Review of Selected Data Sets on Natural Disasters and Impacts. 2006.

14. R., G.-S.D.B., The Quality and accuracy of disaster data: A comparative analyses of 3 global datasets Working paper prepared for the Disaster Management facility, World Bank, Brussels CRED, 2002.

15. EM- DAT. EM- DAT criteria and definition [EB/OL]. 2006; Available from: https://www.emdat.be/.

16. Integrated Research on Disaster Risk. (2014). Peril Classification and Hazard Glossary (IRDR DATA Publication No. 1). Beijing: Integrated Research on Disaster Risk.

17. Haghighat, M., M. Abdel-Mottaleb, and W. Alhalabi, Discriminant Correlation Analysis: Real time Feature Level Fusion for Multimodal Biometric Recognition. IEEE Transactions on Information Forensics and Security, 2016. 11(9): p. 1984-1996.

18. Saran, J.F., et al. Data and information fusion in the context of emergency management: The DF100Fogo project. in 2017 12th Iberian Conference on Information Systems and Technologies (CISTI). 2017.

19. Laylavi, F., A framework for adopting Twitter data in emergency response. 2016.

20. Beatson, A., A. Buettner, and T. Schirato, Social Media, Crisis Mapping and the Christchurch Earthquakes of 2011. Vol. 14. 2014.

21. Fosso Wamba, S., A. Edwards, and R. Sharma, 'Big Data'as a Strategic Enabler of Superior Emergency Service Management: Lessons from the New South Wales State Emergency Service. 2012.

22. Cai, L. and Y.J.D.S.J. Zhu, The challenges of data quality and data quality assessment in the big data era. 2015. 14.

23. Initiative, H.H., Disaster relief 2.0: The future of information sharing in humanitarian emergencies, in Disaster Relief 2.0: The future of information sharing in humanitarian emergencies. 2010, HHI; United Nations Foundation; OCHA; The Vodafone Foundation.

24. Tapia, A.H., K.A. Moore, and N.J. Johnson. Beyond the trustworthy tweet: A deeper understanding of microblogged data use by disaster response and humanitarian relief organizations. in ISCRAM. 2013.

25. Goodchild, M.F. and J.A. Glennon, Crowdsourcing geographic information for disaster response: a research frontier. International Journal of Digital Earth, 2010. 3(3): p. 231-241.

26. MCDEM (Ministry of Civil Defence and Emergency Management). The 4Rs. 2017; Available from: http://www.civildefence.govt.nz/cdem-sector/cdem-framework/the-4rs/.

27. Scannapieco, M., Data Quality: Concepts, Methodologies and Techniques. Data-Centric Systems and Applications. 2006: Springer.

Reference cont...

28. Kaisler, S., et al. Big Data: Issues and Challenges Moving Forward. in 2013 46th Hawaii International Conference on System Sciences. 2013.

29. Prasanna, R. and T.J. Huggins, Factors affecting the acceptance of information systems supporting emergency operations centres. Computers in Human Behavior, 2016. 57: p. 168-181.

30. Yang, L., R. Prasanna, and M. King, On-Site Information Systems Design for Emergency First Responders. Vol. 10. 2009.

31. Prasanna, R., L. Yang, and M.J.R.M. King, Guidance for developing human-computer interfaces for supporting fire emergency response. 2013. 15(3): p. 155-179.

32. Bruns, A. and J.E. Burgess. Local and global responses to disaster:# eqnz and the Christchurch earthquake. in Disaster and emergency management conference, conference proceedings. 2012. AST Management Pty Ltd.

33. Comfort, L. K. (2005). Risk, security, and disaster management. Annu. Rev. Polit. Sci., 8, 335-356.

34. Ouyang, M. (2014). Review on modeling and simulation of interdependent critical infrastructure systems. Reliability engineering & System safety, 121, 43-60.

35. Fakhruddin, B., Virginia, M., & Fernando Gouveareis. (2019). Disaster Loss Data In Monitoring The Implementation Of The Sendai Framework. https://council.science/publications/disaster-loss-data-in-monitoring-the-implementation-of-the-sendai-framework

36. Fakhruddin, B., Murray, V., & Maini, R. (2017). Disaster loss data in monitoring the implementation of the Sendai Framework [Policy brief].

37. Frolova, N. I., Larionov, V. I., Bonnin, J., Sushchev, S. P., Ugarov, A. N., & Kozlov, M. A. (2017). Loss caused by earthquakes: rapid estimates. Natural Hazards, 88(1), 63-80.

38. Batini, C., Francalanci, C., Cappiello, C., and Maurino, A. 2009. "Methodologies for Data Quality Assessment and Improvement," ACM computing surveys (41:3), pp. 1 - 52.

39. Naumann, F. 2013: Data profiling revisited. SIGMOD Rec. 42(4), 40-49

40. Bapon Fakhruddin, Li Guoqing, Rebekah Robertson, 2018. Rapid Damage Mapping and Loss Data Collection for Natural Disasters - Case Study from Kaikoura Earthquake, New Zealand. China Scientific Data, 2018. (2018-06-14). DOI: 10.11922/csdata.2018.0033.en.

41. Endsley, M.R. 1995. Toward a Theory of Situation Awareness in Dynamic Systems. Human Factors Journal 37(1), 32-64

42. Bapon (SHM) Fakhruddin, Richard Reinen-Hamill, Rebekah Robertson, Extent and evaluation of vulnerability for disaster risk reduction of urban Nuku'alofa, Tonga, Progress in Disaster Science, Volume 2, 2019.

43. WMO, 2019. he Second Multi-Hazard Early Warning Conference (MHEWC-II) 13 - 14 May 2019, Geneva, Switzerland. https://mhews.wmo.int/en

44 SPREP, 2019. Progress on the Pacific Island Communication and Infrastructure (PICI) Panel. https://www.pacificmet.net/pmc/meetings/pmc-5/agenda