17th World Conference on Earthquake Engineering, 17WCEE

Sendai, Japan - September 13th to 18th 2020

IMPROVED DISASTER RESILIENCE THROUGH INTEGRATED DATA COLLECTION AND MANAGEMENT

S.-L. Lin[(1)], F. Scheele[(2)], S. Singh[(3)], C. Chan[(4)], P. Holland[(5)]

[(1)] *Senior Risk Engineer, GNS Science New Zealand, s.lin@gns.cri.nz*
[(2)] *Risk Specialist, GNS Science New Zealand, f.scheele@gns.cri.nz*
[(3)] *Senior Geospatial System Architect, SPC Fiji, sachindras@spc.int*
[(4)] *Project Assistant, SPC Fiji, carrolc@spc.int*
[(5)] *Environmental Economist, NIWA New Zealand, Paula.Holland@niwa.co.nz*

Abstract

Comprehensive and coherent asset data (information on items that may be affected by natural hazards (e.g. earthquake and tsunami) such as people, buildings or infrastructure) and data on those assets after events (damage survey) have long been recognized as an information need to support improved disaster risk management (DRM). Potentially useful asset, exposure and damage data and related information from past events are available, albeit usually dispersed around different government departments and agencies, regional and other organizations.

To enable better data management, to assist in the efforts of proper safe keeping of this information and to further assist with the disaster modelling tools, an integrated system and methods to enable the accurate collection, management and dissemination of this information is proposed and implemented.

In this paper, a brief review of existing data management for disaster risk management is presented, followed by an overview of the proposed integrated data collection and management system toward improved disaster resilience. The paper discusses: (i) the importance of data management for risk assessment, (ii) information on various types of data for risk assessment, (iii) identification of open data principles and available data management platform, and (iv) information on data collection histories and methodologies. Next, applications of the proposed system in Pacific region are described. Finally, the proposed data management system and its application are discussed.

Keywords: Disaster resilience; Data management; Data collection; Disaster risk management

1. Introduction

The countries along the Ring of Fire (the Pacific Ocean) are prone to many natural hazards (e.g. earthquake, volcanic eruption and tsunami) which can cause significant impact to communities due to economic losses, damage to the built environment and crops, and threat to populations. Risk assessment is a fundamental tool underpinning evidence-based decision making on risk management and could be used to improve disaster resilience in the communities.

Examples of how risk assessments can improve the disaster resilience of a community include:

- Land use planning decisions can quantify cost and benefits relating to risk management.
- Immediately after an event, a near-real time impact assessment will provide emergency response officers information to advise the public and prioritise resources.
- An asset owner seeking to purchase insurance can understand what losses could be incurred, and with what probability.
- An insurance company can use the same information to set premiums and to purchase reinsurance.

These issues can all be addressed by a risk assessment that identifies the likely losses, in terms of severity and frequency of occurrence. When attempting to estimate risk we need to know three things: the magnitudes of the hazards affecting the assets, the vulnerabilities of the assets, and the quantities of the assets. Of those the understanding of natural hazards, such as the modelling of seismic hazards over the region is now relatively advanced, with the beginnings of models for volcanic hazards and tsunamis.

Much the same is true for vulnerability which relates the probability of a given asset type reaching or exceeding a given damage state as a result of a seismic hazard. Soundly based damage states/ratios exist for some types of buildings, equipment, and infrastructure affected by earthquakes and tsunami, and rudimentary damage ratios for buildings affected by eruption products phenomena also exist. Nevertheless, much work remains to be done.

The case of assets (or a dataset on them to represent exposure) remains difficult. When small numbers of buildings are involved, say, less than a thousand or so, it may be possible to acquire the information necessary for modelling with a reasonable level of confidence. For regional or national-scale projects, viewing/ surveying each building is impossible. Nevertheless, the results of a survey of a sample of buildings could be useful and practical in developing regional and national-scale projects.

Traditionally, paper-based surveys by teams/ individuals have been carried out to gather the asset information required for modelling. Such a method takes considerable time and effort. Recent changes in the technology used for spatial data acquisition, computing and information management provide a framework that has the potential to allow to acquire, analyses and manage data in innovative ways. This includes the development of tools such as RiACT [1], GEM DO [2], and HHI KoBoToolBox (www.kobotoolbox.org).

With the significant increasement of these data available for risk assessment, many attempts have been proposed to provide a central space where data can be detailed documented, safely archived and shared, and managed to ensure quality, appropriate use and ongoing maintenance. PacRIS (Pacific risk information system) is a project which developed an exposure model for the Pacific Region. It was used for estimating monetary losses and casualties in the Pacific Catastrophe Risk Assessment and Financing Initiative (PCRAFI) project [3]. The initiative aims to provide the Pacific Island Countries (PICs) with disaster risk modelling and assessment tools for enhanced disaster risk management, and to engage in a dialogue with the PICs on integrated financial solutions to increase their financial resilience to natural disasters and to climate change.

Though the above achievements, improving data collection and management in post-event reconnaissance surveys and in long-term impact assessments, is an ongoing challenge to improving the quality of datasets for use in catastrophe risk assessment [4, 5].

This paper offers practical approaches into the development of a data management system for use in risk assessment toward the improvement of disaster resilience. The authors first discuss the importance of data management for risk assessment. An overview of various types of data for risk assessment is then provided, followed by a discussion on identification of open data principles and available data collection and management platform. Next, the development of an integrated data management system is illustrated and demonstrated via applications in Pacific region. Finally, the proposed data management system and its application are discussed.

2. Understanding Data Management for Risk Assessment

Risk assessment is the basis to improve disaster resilience, such as emergency planning, mitigation, response, and recovery. The assessed impact on civil infrastructure systems, national economies, and societal activities, provides the foundation for developing plans that adequately protect vulnerable communities. Generally, such impact assessment software is composed of three main components: (i) hazard, (ii) vulnerability, and (iii) asset.

As mentioned in previous section, through field investigations after major event, along with theoretical and experimental studies, our understanding of hazard and vulnerability components have significantly improved in the recent years. Similarly, rapid and profound changes in the technology used for data acquisition, computing and information management provides a framework that allows communities to collect asset data in new ways. While these significant advances in hazard, vulnerability and assets improve our understanding of risk assessment, they also produce unprecedented amount of data. The later has long been recognized as emerging ongoing challenge in disaster risk management.

Below, the various types of data needed for risk assessment are discussed, followed by histories and methodologies of data collection tools. Next recent development on data management platform and open data principles are briefed to illustrate recent attempts toward improving the efficiency, transparency and standardization of the increasing data.

2.1 Data types for risk assessment

Information of hazard intensity (e.g. the magnitudes of the earthquake shaking), vulnerability (e.g. performance of assets to earthquake shaking) and quantity and valuation of assets (e.g. people, buildings and infrastructure) is required when attempting to assess hazard impact to the affected community.

2.1.1 Hazard Data

As mentioned previously, the modelling of seismic hazards is now relatively advanced, with the beginnings of models for volcanic hazards, landslides and tsunamis. Typically, the seismic hazards are presented via various shaking intensity (e.g. PGA, the peak ground acceleration) versus various return years (e.g. 100, 500 or 2500 years).

This information is useful in providing an overall assessment of earthquake hazard, but must be combined with amplification, liquefaction and landslide susceptibility information before the full severity and distribution of the seismic hazard can be assessed. For example, sites on thick sediment tend to amplify shaking compared to adjacent rock sites resulting in more damage to buildings and infrastructure.

By contrast, very limited studies exist that provides a good understanding of the volcanic, landslide and tsunamic hazards. Alternative approaches might use hazard maps resulting from available safety maps (e.g. the ash-fall safety zone map [6], as shown in Fig. 1), and historical events.

Fig. 1 – Tanna Island ash fall safety map

2.1.2 Vulnerability Data

People, buildings, crop, infrastructure and other assets all comprise attributes that are susceptible to hazard impact. The collective susceptibility of attributes to damage or loss influences asset vulnerability. An assets susceptibility to damage often increases in response to increasing hazard intensity thereby enabling these relationships to be modelled. Vulnerability functions are typically a continuous or discrete series of statistical functions or values relating a hazard intensity parameter (e.g. earthquake shaking, ashfall thickness) to a level of direct or indirect damage to response of an asset. Three common methods are employed to develop vulnerability functions:

- 'Empirical' e.g. derived from a regression analysis of observed direct or indirect asset damages correlated with hazard or damage intensities.
- 'Analytical' e.g. the use of asset attribute and damage analysis techniques to develop a theoretical model of direct or indirect asset damage for various hazard or damage intensities.
- 'Synthetic' e.g. direct or indirect asset damage for various hazard or damage intensities based on the judgement of experts.

Continuous or discrete functions estimate levels of asset damage or loss in two ways (Fig. 2). Functions represent asset damage response to increasing hazard intensity as either a ratio or percentage of "cost to repair/cost to replace" the asset to pre-damaged condition (i.e. damage function) or the conditional probability that a damage state or level (i.e. fragility function) will be reached or exceeded for a given hazard intensity. Damage functions are commonly applied for estimating economic losses from hazard events or scenarios whereas, fragility functions are applied when information about the likely range of physical level or state of asset damage is required.

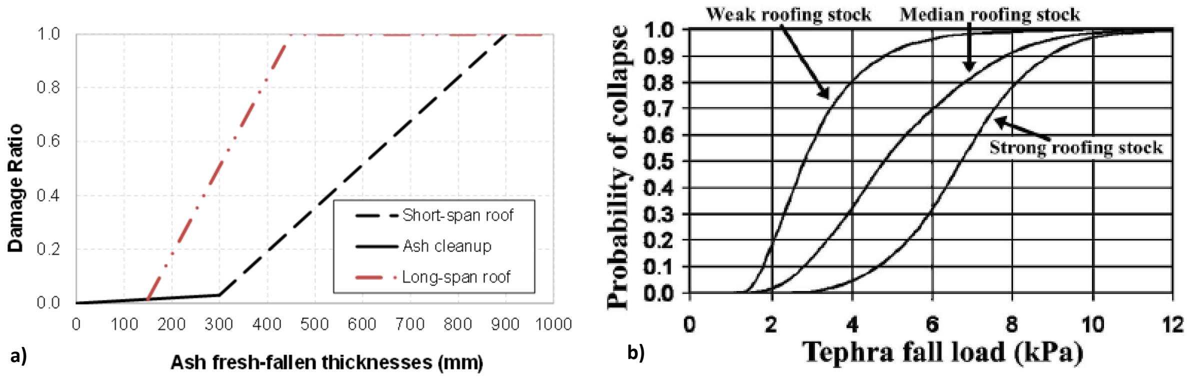

Fig. 2 – Examples of continuous curves representing a) damage functions and b) fragility functions [7]

2.1.3 Asset Data

The asset (or exposure) dataset defined here contains buildings, infrastructure, crops, as well as people because much of the post-disaster suffering, be it death, injury, or loss of shelter, is related to impacts on people. Hence building asset models need to be accompanied by occupancy models that describe the numbers of occupants at various times of the day and week.

The attributes in an exposure database will vary depending on its purpose as well as on the size of the portfolio area. More often, the quality and quantity of an exposure database is driven by the available data sources and resources (human and budget) to develop it. Ideally, when conducting risk modelling we would know the location and value of each of the exposure models and have enough structural information to underpin the assignment of vulnerability functions for each of several hazards.

Often, no single data source contains all the necessary attributes, so multiple available sources must be collated to map the same attributes from various sources and definitions. Furthermore, in-field data collection survey is another alternative to help fill the data gaps. More importantly, it provides a local understanding of asset portfolios within the context of risk reduction and planning, as well as supports better decision-making regarding disaster response. Details on the histories and recent development of data collection is discussed in the following section.

2.2 Data collection histories and methodologies

Traditionally, paper-based surveys by teams/individuals have been conducted to gather information on assets pre- and post-disaster. Such methods take considerable time and effort. Recent developments in technology for spatial data acquisition, computing and information management provide an opportunity for communities to more rapidly acquire, analyse and manage damage data in innovative ways.

In the rest of this section, the history of data collection is briefed, followed by discussion of the development and application of two commonly available data collection tools, namely RiACT and KoBo.

2.2.1 History of data collection

The importance and value of data collection and management have long been recognized and were highlighted in the Canterbury earthquake sequences [8] when the condition of many buildings needed to be assessed and reported as quickly as practical.

In recognition of the above needs of data capture and management, and by taking advantage of the rapid development of information and communication technology, several digital tools have been developed to directly record the built environment at risk from earthquakes. These include, for example, ERS developed by the Earthquake Engineering Research Institute [5], ROVER developed by the Federal Emergency Management Agency (www.fema.gov/rapid-observation-vulnerability-and-estimation-risk), the DO Androdi tool developed by Global Earthquake Model [2] and the KoBoToolBox developed by the Harvard Humanitarian Initiative.

Despite these advancements, challenges and limitations in data collection and management remain. For instance, support is no longer available for the ERS tool development since its first publication in 2004. As a result, its application might be limited due to out-of-date technology. The DO Android tool, by contrast, was developed using the current information technology of portable devices. Strengths of this tool include the use of widely available Android operation system, the display and use of a device-enabled global positioning system (GPS) and maps, and built-in relatively high-resolution cameras. However, asset information (i.e., attributes) collection using the tool has been restricted to Global Earthquake Model building attributes taxonomy and is not easily modified. In addition, data acquired using the tool is locally stored on each collection device and needs to be manually uploaded and merged following each field survey; an inconvenience when direct download is available, but a necessity when communication channels are disrupted.

Due to the above limitations, an improved data collection system has been developed by GNS Science [1]. This includes development of an integrated and extensible framework for the capture of attribute data that describes the characteristics (pre-disaster) and damage (post-disaster) of assets. The framework, referred to as Real-time Individual Asset Attribute Collection Tool (RiACT), Inventory Repository (IR) and asset repository web portal (WEB), enables data capture by direct field observations of asset attributes and/or damage data.

2.2.2 RiACT – a real-time individual asset attribute collection tool

Fig. 3 illustrates the RiACT field data collection framework and how its components are combined to achieve the main goal of this development. RiACT uses direct observations, its real-time telecommunication, its georeferencing (i.e., background satellite image maps along with a global positioning system, GPS) and photo-taking (i.e., on-board camera) abilities to collect the characteristics and/or damage condition of each asset feature. It should be noted that asset location determined via GPS has been a severe limitation in the past as often it is important to have more accurate location data. To improve this, RiACT uses a tablet GPS to locate observers in the asset vicinity, while the background map enables the actual location of the surveyed asset to be verified. Because of this improvement, the conventional geocode process of converting street addresses into geographic coordinates (e.g., latitude and longitude) is no longer required. The uncertainty from geocoding verification could therefore be removed.

Upon completion of each record, it is stored within the tool and transmitted via mobile internet connection (if available) to the inventory repository (IR). Meanwhile, the recorded information can be modified on / or off-field, and new assets (e.g., individual buildings) can be added to the IR via the WEB either as individual records or batch uploads from spreadsheet compatible files (i.e., comma-separated-values, CSV). The IR can also be updated with additional building information from design drawings or other PDF files. These can, on request from the field surveyors, be downloaded to RiACT to provide the field inspection team full access to previously available building data to which new or modified data can be added (e.g., damage data) and managed.

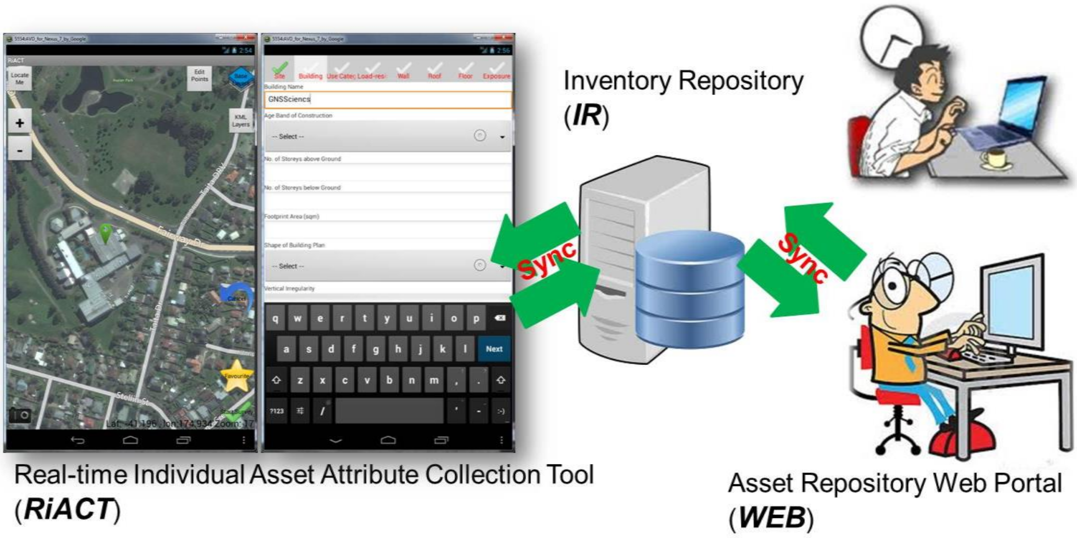

Fig. 3 – Schematic of the integrated exposure data development framework

2.2.3 KoBo Toolbox

KoBo Toolbox (www.kobotoolbox.org) is another tool which provides a flexible data collection option. The tool is free, and the software is open source. KoBo supports the full data collection cycle (Fig. 4), from inception, data collection (both online and offline), and analysis, which can be exported into any GIS (Geographic Information System) system should geo-location be recorded, or for spatial joins with existing layers.

Fig. 4 – KoBo Toolbox workflow covering the full data collection cycle - inception, collection and analysis

KoBo allows users to build survey forms from scratch, without the use of code, through an intuitive form builder using a range of question types including location, images, and ratings. Skip logic and validation is also included allowing more complex forms to be deployed. Once created, projects can be shared, using different levels of permissions and forms can be imported or exported.

Registration for accounts is simple, however users have the choice to have accounts hosted on either of two servers:

- Humanitarian Server: hosted by UN Office for the Coordination of Humanitarian Affairs (UN OCHA).
- Non-Humanitarian Server: hosted by KoBo Toolbox.

It is recommended that users utilize the humanitarian server due to the unlimited storage and submissions, as well as for the purpose of longevity. Users would not be able to share or collaborate on existing surveys if their respective servers differ. As with all open source software, all components of KoBo Toolbox can be installed on a local server (if needed), with the code located on GitHub (https://github.com). Installation requires advanced server administration and programming skills.

Users are able to control access to forms through sharing, however forms are by default inaccessible to both other users and system administrators. Once survey forms are deployed, users are able to collect data via Android (KoBo Collect) or browser-based form (Enketo) on other OS devices either online or offline. Submission of forms is done directly to the server once connected to the internet. Exposing Data to GeoNode (http://geonode.org, an open source data management platform, discussed in the following section).

Once data is properly prepped, it is possible to expose the datasets (if permitted) to GeoNode instances. Data collected using KoBo with attributes, are exported either as CSV or CSV (legacy). These files can be imported into QGIS (https://qgis.org, an open source GIS platform) or any other professional GIS suite as a delimited text file and the longitude and latitude used to create a shape file suitable for upload into the GeoNode instance.

Should the dataset not have any geometry, it is possible to load the layer as a table into QGIS or any other professional GIS suite, and perform a join to a related spatial layer using similar unique IDs.

2.3 Data management platform and open data policy

As suggested previously, no single data source contains all the necessary attributes required for risk assessment, so multiple available sources must be collated to map the same attributes from various sources and definitions. In this section, a brief overview of data management platform and open data policy which was developed and proposed to facilitate the data management for risk assessment are discussed.

2.3.1 GeoNode – an open source data management platform

GeoNode is an open source, web-based content management system for developing GIS and has been utilized in the Pacific region to deploy spatial data infrastructures (SDI). It is designed to be extended and modified and can be integrated into existing platforms. Broken into usability components, GeoNode is a web application used to:

- Upload, manage, and share geospatial data.
- Create and share interactive maps.
- Collaborate and interact with other users.

2.3.2 Open data for disaster risk and resilience

Through the implementation of a Pacific SDI, a set of interconnected components contribute to creating an environment where stakeholders across the region can interact with available datasets via PacGeo (http://www.pacgeo.org/) and PCRAFI (http://pcrafi.spc.int/) (both are established under GeoNode framework) as well as their own locally deployed databases. With support from the various projects in the region, local databases have been deployed with data management and administration training provided including respective stakeholder agencies for handover.

The existence of these multiple instances raises the need for effective communications, consistency, standards and sharing of information for national and regional resilience. Accordingly, the Open Data for Resilience Initiative (Open DRI) (see https://www.gfdrr.org/en/publication/open-data-resilience-initiative-field-guide for further information) aims to target these issues globally.

3. Overview of the Proposed Data Management System and its Application

To enable better data management, to assist in the efforts of proper safe keeping of this information and to further assist with the disaster modelling tools (hence the disaster resilience), the Pacific Risk Tool for Resilience project (PARTneR), along with the Pacific Resilience project (PREP) being led by SPC, has aligned efforts to provide pacific island countries with systems and methods to enable the accurate collection, management and dissemination of this information.

Fig. 5 illustrates a data management system developed under the PARTneR project and how its components interact. GeoNode was selected as the data host server which is linked to a real-time individual asset attributes collection tool (RiACT) or KoBo to collect required asset characteristics or damage information in the aftermath of a hazard event. Upon completion of each data collection activity, the data collected is then transmitted to the in-country GeoNode servers (e.g. Samoa GeoNode (www.samoageonode.ws), Vanuatu GeoNode (www.geonode.gov.vu)). The in-country GeoNode servers are currently hosted and maintained by disaster management office, namely DMO and NDMO in Samoa and Vanuatu, respectively. With the development of this centralized data management system, the asset, hazard and impact datasets could be easily shared among interested government ministries and departments (e.g. Ministry of Agriculture & Fisheries (MAF), Fire & Emergency Services Authority (FESA), Planning & Urban Management Agency (PUMA-MNRE), Vanuatu Meteorology & Geo-hazards Department (VMGD) etc.) for disaster planning, response as well as loss modelling.

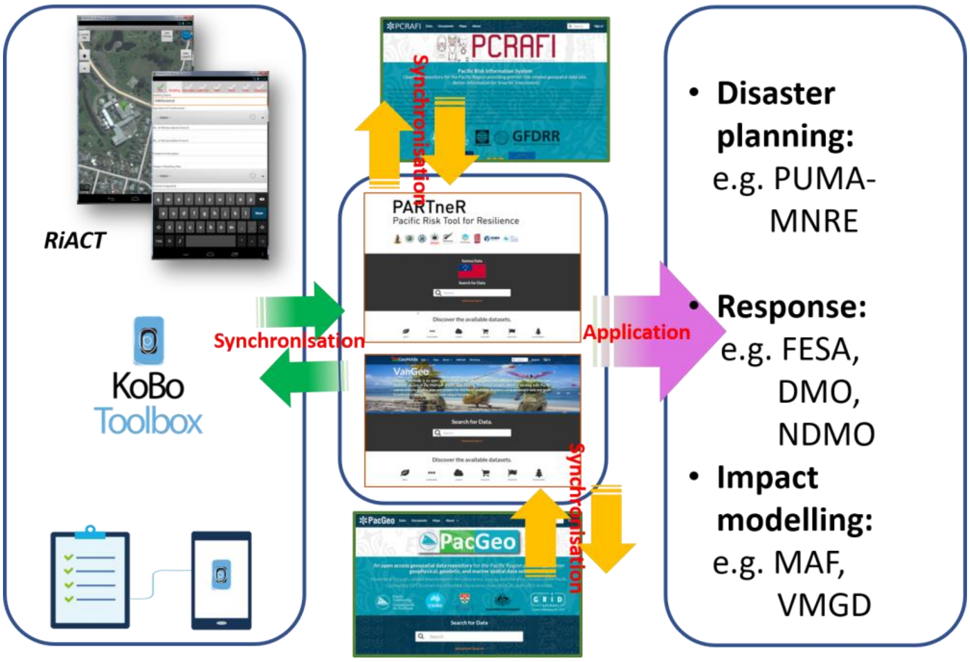

Fig. 5 – Schematic of the proposed data management system

Since the establishment of the GeoNode servers in Samoa and Vanuatu, many data management training (including GIS skills and data capture) have been conducted to allow the potential users to get familiar with this system. The on-going task is to encourage data sharing for sustainability. For example, the use of risk tools to model and plan for disaster events relies on a whole-of-government approach as different agencies have varying responsibilities related to the same disaster event.

In practice, data sharing can be a challenge. Opportunities to enhance data sharing for the sustained use of risk tools discussed from various training workshops include: (i) securing strong political buy-in, promoting data sharing via GeoNode, (ii) developing standardized collection templates and a seamless data collection workflow, and (iii) engaging with data stakeholders in-country.

On the other hand, with the rapid development in the information technology field, it is expected that further development using a range of new technologies and tools may instil more efficiency, transparency and standardisation of the developed data management framework.

4. Concluding Remarks

In this paper, the importance of data management toward improved disaster resilience is first discussed, followed by a detailed description of the proposed integrated data management system for risk assessment. The components required for such a framework include a spatial data management platform (i.e., GeoNode) and data capture tools (e.g., RiACT and/or KoBo).

Collection of asset and impact data (hence the understanding of the exposure) can provide useful information for disaster risk management planning, including readiness and response initiatives and for long-term risk reduction decision-making. The history and recent development in the field data collection of pre-event detailed building types and post-event impact has also been introduced.

The implementation and feedback from the pilot examples in Samoa and Vanuatu not only demonstrate the successful development of the framework for data management, but also the achievement of the objectives of this development. However, further improvement along with more use-cases are being conducted to further demonstrate and refine the feasibility of this framework according to targeted user needs and adoption of new IT technology.

Acknowledgements

The funding provided by the PARTneR project which is funded by the New Zealand Ministry of Foreign Affairs and Trade – Partnership Fund. Further, the authors wish to thank the anonymous reviewers for providing valuable comments and suggestions that helped to improve the quality of the paper.

References

[1] Lin S-L, King AB, Horspool N, Sadashiva V, Paulik R, Williams S (2019) Development and Application of the Real-Time Individual Asset Attribute Collection Tool, *Frontiers in Built Environment: Earthquake Engineering*, **5**:15. doi:10.3389/fbuil.2019.00015

[2] Foulser-Piggott R, Vicini A, Verrucci E, Bevington J., Shelley W (2013) IDCT Mobile Tools – *Field Test Reports, Produced in the context of the Inventory Data Capture Tools Risk Global Component, Version 1.0*, GEM Foundation, Pavia, Italy.

[3] Air Worldwide (2011) Pacific Catastrophe Risk Assessment and Financing Initiative (PCRAFI) Component 2: Exposure Data Collection and Management. Air Worldwide, Boston.

[4] Porter, K.A. 2002. "Learning from Earthquakes: a Survey of Surveys". *EERI Invitational Workshop: An Action Plan to Develop Earthquake Damage and Loss Data Protocols*, Pasadena, California, September 19-20, 2002.

[5] Greene, M.; Grossi, P.; Tubbesing, S.K.; Basoz, N.; Love, R.J. 2004. "Learning from earthquakes: new directions and initiatives". *Proceedings of the 13th World Conference on Earthquake Engineering*, Vancouver, B.C., Canada, 1-6 August, 2004.

[6] VMGD, 2016. Tanna communities safety map (ashfall). *1 map sheet scale approx. 1:90,000 at A3*. Vanuatu Meteorology and Geohazard Department (VMGD), Port Vila.

[7] Spence, R.J.S.; Kelman, I.; Baxter, P.J.; Zucarro, G.; Petrazzuoli, S. 2005. Residential building and occupant vulnerability to tephra fall. *Natural Hazards and Earth System Sciences*, **5**: 477-494.

[8] Galloway, B.D.; Hare, H.J.; 2012. "A review of post-earthquake building control policies with respect to the recovery of the Christchurch CBD". *Proceedings of the NZSEE Annual Technical Conference & AGM*, Christchurch, New Zealand, 13-15 April, 2012.

