Open Science Cyber-Infrastructure Framework for Next Generation Disaster Analytics

Nurfadhlina Mohd Sharef, Deden Witarsyah Jacob, Masaharu Hayashi

Abstract: The open science movement is gaining popularity due to the stability of data storage and network technologies besides the availability of open data portal in many countries. However, a case study that focuses on the requirement and design of the cyber-infrastructure for open science is limited. This paper reports the assessment of existing infrastructure for disaster information management as an open science activity based on the Sendai Framework. A framework that combines the open data quality and the next generation repository system requirements based on a case study on the flood and forest fire management in Malaysia and Indonesia is proposed. This paper fills the gap between the focus on open data framework and the next generation repository system based on the requirements from a recent international collaboration on climate research studies.

Keywords: Open science, Next generation repository system, Open data, Disaster analytics.

I. INTRODUCTION

Open data [1] is an emerging phenomenon happening around the world due to the disruptive technology caused by the advancement of data storage, network and data analysis technologies. The openness and connectivity empowered open data to gain traction among governments, public and businesses due to its promises to bring socio-cultural and technology change [2]. However, the data quality is still a risk and threat to the transform the data into value [3].

Data quality issues have been studied mainly on datasets internal to organizations, where the metadata and domain knowledge of the data semantics are known. Despite the hype from the potential value for open data, there is also some evidence that the time-to-value from these datasets remains prohibitively long [3], primarily due to lack of knowledge on the quality characteristics of the data and resulting effort of making the data ready for consumption of the public. Besides data discoverability issue, the underlying quality of the metadata is said to be deficient due to the lack of accessibility and schema descriptions, has representation issue. Furthermore, the published data contains duplication, inconsistency and missing data [4]. Existing works on data quality are divided into data and process analysis, data quality requirement analysis and data quality analysis [3].

Revised Manuscript Received on May 22, 2019.

Nurfadhlina Mohd Sharef, Department of Computer Science, Faculty of Computer Science and Information Technology, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, MALAYSIA.

Deden Witarsyah Jacob, Department of Information Systems, School of Industrial Engineering, Telkom University, Bandung, 40257, INDONESIA.

Masaharu Hayashi, Research Centrer for Open Science and Data Platform, National Institute of Informatics Chiyoda Ward, Tokyo, JAPAN

Among the approach for data quality assessment and requirements identification is the assessment of information quality conducted through individual experiences and understanding of certain criteria. The examples of criteria are interpretability and concise-representation. This paper is prepared by following this user-centred setting with specific focus on data exploration and data profiling strategy. The aim of this paper is to discuss the present open data quality challenges based on two case studies from Malaysia and Indonesia specifically for natural disaster data management and outline an agenda for future research to address this bottleneck so that valuable investment from open data can be harvested. The hypothesis is that the information and media management of natural disaster could be improved by an open data framework that could facilitate the existing business function, reliable and easy to use.

This paper is distinguished from existing literature based on its emphasis on the need of change for both countries towards preparing the next generation open natural disaster data management. Besides, we fill the gap from the lack of knowledge on the scale and impact of data quality problems across datasets from various international open data portals. Accordingly, we see a need for a comparative nations study to explore the open data problem and suggest efforts that can be made to address it.

The paper is divided into four parts. The first section gives a brief introduction to the proposed work while the second section highlights the related work. The third section presents the proposed framework continued by the conclusion in section four.

The mostly used definition of open data binds to the eleven requirements that open data should conform to which emphasizes on how to enable the free use, reuse, and redistribution of data. Moreover, open data should not discriminate any person and must not restrict the use of the data to a specific field or venture [4].

II. OPEN DATA INITIATIVES

Open data has very close connection to the government as open government data is said to be able to unlock more innovation to drive the knowledge base economy towards global competitiveness of the country [5]. Despite various efforts including openness index [6] to improve the quality of open data in e-government, insufficient heed to the management of datasets, resources and associated metadata publishing on the online portal still lingers today's organization [7].

This is revealed through a study that utilizes an Analytical Hierarchical Processing framework to compare over 250 open data portals by organizations across 43 different countries. However, this study did not include both Malaysia and Indonesia open data portal. Indeed, Malaysia and Indonesia are both in high support and preparation towards transforming the e-government towards a full-fledged open data initiative. Therefore, the following subsections present about the open data initiatives in these countries.

data.gov.my

The data.gov.my is a portal delivered by the Malaysian Administrative Modernisation and Management Planning Unit (MAMPU) that offers one-service-centre for citizen to access and download online the government's open datasets. Through the portal, open data is also envisaged to increase the nation's digital economy productivity through new industries or innovations with the involvement of the people and the business community besides improving the transparency of government services delivery through sharing of data.

The open data portal presents the data provided by the organizations according to national statistics, education, international trade, agriculture, transport, census, environment, health, mapping, land ownership, legislation, crime, government speed, budget, company registration, public contract and national election clusters. Users can also view and download the available data as published by dataset publishers (eg., ministries and agencies). The portal allows users to request for a dataset by providing their information which are name, email address, dataset title, dataset description, dataset publisher for the source of dataset and the requested data format (i.e., CSV, Excel, JSON, HTML, RDF and XML). A term of usage is also provided for the reference of users of the dataset.

Portal Satu Data Indonesia

The Nawa Cita program is the foundation of the Indonesian government's mission under President Joko Widodo to realize a clean, effective, democratic, and reliable government [8]–[10]. One of the steps taken is to continue efforts to build more clean, effective, democratic, and reliable government through the OD commitment. OD commitment is a concrete form of the Indonesian government's seriousness in conducting public sector reform in Indonesia. OD Indonesia is expected to be born ideas, initiatives, and practices of government openness to all levels of society in collaborating and accelerating the achievement of national targets and priorities. It is significantly would strengthen the quality of public policy innovations to fit the public needs and implementation of the government. Furhermore, open data movement is began with the government support for publishing public data by institutions in the form of One Data Initiatives .

Data.go.id is the official portal of Indonesian open data as a form of operationalization of one data initiative. This portal contains data of ministries, government agencies, local government, and all other related agencies that produce data related to Indonesia. Utilization of government data is not limited to its use for policy-making, but also as a form of public data fulfillment for the public. Through a one-data portal, the government supports and strives to reform the Indonesian government data. The data is available in an open and easy to use format, with the aim of improving government transparency and accountability, as well as to increase community participation in development.

Data Quality Evaluation

Data quality expectation basically includes usability, accuracy, completeness, consistency, timeliness, accessibility and openness [4]. Many governments now aim towards increasing the quantity of published data (and slowly many are moving towards making sure the data is machine readable to encourage reusability) but not necessarily matching the needs of the open data consumers.

This is partly because of their current focus to comply with the metrics of the Open Barometer assessment[1]. Produced by the World Wide Web Foundation, the Open Data Barometer covers 115 jurisdictions in the fourth edition. The barometer ranks governments based on the readiness for open data initiatives (including government, citizens, and enterpreneurs), implementation of open data programmes (including accountability, social and innovation) and the impact (including political, social and economical) that open data is having on business, politics and civil society. Known for data quality in the online portal, the focus of the barometer is not scoped to a specific purpose. Thus, the consumers of the portal have to survive with available data (although on-demand dataset request could be expressed, this takes time and not guaranteed for access) have to invest a significant effort to generate results only to find it either inadequate, low quality, or even worse; rely on erroneous results. This indicates that an exploratory tools and approaches (i.e., comprehensive query engine within the online data portal) that allow user to be aware of the data incomprehensiveness and easy to understand with the support of schema properties could be considered.

There are seven metrics of data quality [3], [11] which covers the cell and dataset level. The traceability characteristics focus on the track of creation and updates while the current focuses on the percentage of rows of a dataset that have current values and the delay in publication. The delay after expiration indicates the ratio between the delay in publication and the period of time referred by the dataset. The completeness characteristics indicates the percentage of a non-empty values in a dataset while compliance is addressed through the percentage of standardized columns in a dataset. The understandability characteristics indicate the percentage of a dataset that has associated descriptive metadata and the representation of file in a machine-readable format. The accuracy of the data focuses the correctness of the values according to the domain and the type of information in the dataset. Accuracy also indicates the ratio between the error in aggregation and the scale of data representation.

Open Science

Open science infrastructure is becoming crucial as scientific work becomes more computational and data-intensive, besides research processes and results become more difficult to interpret and reproduce. Open science emphasizes digitization, transparency, reproducibility, and accountability and refers to a worldwide repository of documents, data sets, tools and services, analysis scenarios, and meta knowledge about these, for researchers from different disciplines to publish, share, exchange, and easily reedit and federate resources together to help in new contexts and for new purposes [12]. This is closely aligned to the Industrial Revolution 4.0 movement that is gaining popularity in many countries. This is also an extension of the generous data collected through the internet-of-things technologies that are applied by science related agencies throughout the world. Currently the Open Science readiness is relatively low in terms of the acceptance among scientist and availability of technologies that could facilitate the sharing; despite the common usage of social media platforms to share less formal content.

The disruptive Open Science phenomena which emphasizes on the transparent and accessible knowledge that is shared and developed through collaborative networks [13]. The pillar of this phenomenon is the inseparable social media dependency [14] among researchers nowadays where they could explicitly receive credit for and give credit to datasets with machine-readable metadata, provenance, and reproducible workflows [15]. The publicly accessible content digitalization [16] also ensure data quality through the data stewardship concept which requires careful planning and thought from the beginning of a research project to cater for data management issues related to long-term data reusability and interoperability.

Typical mode of knowledge dissemination has been continuously practiced through academic and popular writing, as well as exhibition and oral presentation. However, being able to track the related studies still require huge effort and time consuming mainly because the digital and direct connection between the content are still scarce. Online communities such as the Jupiter, ResearchGate and Academia provide a platform for scientist to make their research more visible but since these enforcement to usage of these portals are missing and the owner of the portal could raise doubt among the users, the adoption is still limited.

The Open Science could be connected to the Next Generation Repository System [17] which is based on the Creative Commons and targets quality open digital content. The key to these studies are reproducibility and data sharing to accelerate innovation [18]. The next generation repository system manages and provides a wide diversity of scholarly contents including research data. It is designed with resource-centric architecture and supports machine-friendly features. The next generation repository does not only store scholarly content, but also captures user activity on the repository and actively cooperates with other services and systems. Moreover, it realizes a global repository network by mutually connecting distributed other repositories. Over the global repository network, there are overlay services with various purposes. The overlay service provides various services while processing scholarly contents on the global repository network.

Disaster Information and Media Management

Proper disaster management initiatives such as utilizing the Information and Communication Technologies could improve the vulnerability resilience of the community towards disaster. Various focus has been given to the disaster information management, such as (i) public perception based study to find the causes [19], (ii) study on the effects and mitigation of floods such as through runoff estimation and forecasting (Tahir, Aminuddin, Ramli, & Jaafar, 2017), (iii) the impact, vulnerability and risk assessment such as through severity index and hazard map [20], [26], [27], (iv) policy [25][28], as well as (v) education [29].

The Prevention Web portal[2] by the United Nations Office for Disaster Risk Reduction (UNISDR) provides the Global Assessment Report of Disaster Risk Reduction and monitors the recent trends in disaster risk. The risk of a country for each hazard is defined by the multiplication of hazard, exposure and vulnerability of each; and these are included in a risk modeling. The Sendai Framework for Disaster Risk Reduction 2015-2030 aims to achieve the substantial reduction of disaster risk and losses in lives, livelihoods and health and in the economic, physical, social, cultural and environmental assets of persons, businesses, communities and countries over the next 15 years [30].

The Sendai Framework Monitor also functions as a management tool to help countries develop disaster risk reduction strategies, make risk-informed policy decisions and allocate resources to prevent new disaster risks. The portal also collects reports on studies of the global disaster loss besides recommending policy on disaster loss data, which includes establishment of basic data infrastructure for disaster loss reporting and sharing. Links to other sources provided by UNISDR are also provided such as disaster inventory system, evacuation tool and loss database. Other materials by various countries such as the Global Disaster Identifier Number and catastrophes loss database provided by other sources are also listed.

The Sendai Framework report for data readiness review indicate that Malaysia[3] will start collecting data attributed to disasters from 2016 onwards, as according to 11th Malaysia Plan 2016-2020. The data readiness report for Indonesia[4] indicate that further detail survey will have to be conducted to obtain data about the number of people whose livelihood were disrupted or destroyed, attributed to disasters. The purpose of this review is also to measure the global targets and identify the current gaps.

However, currently in many countries disaster data is in heterogeneous mode and gathering them requires huge effort including financial, capacity and technology infrastructure. Tracing the complete representation of a disaster event is also challenging due to the unknown source and inaccessible content although a formal source

in each country has been created. The information presented in the news by reporters are usually after the endorsement of the authoritative government agency [31]. Although both open data and open science initiative are gaining popularity, no ideal implementation is existing that connects the content displayed in the authoritative government agency portal to with the published news. Furthermore, researches that use related information may provide acknowledgement in their scientific publication, but no explicit connection is available that could ease the interested parties and avoid reinventing the wheels.

III. OPEN DATA FOR DISASTER INFORMATION MANAGEMENT

Disaster risk diagnosis through historical sources is fundamental [32]. Various data sources are referred for the disaster risk management including statistics, description on the situation of the victim, law and disaster management plans, damage database, meteorological agency documents, and published documents. There are also approaches that use heterogeneous data from multiple sources by harvesting the social media for temporal and spatial information and fusing with remote sensing observations both for retrospective analysis and for real time monitoring and decision making [33].

The sources are diverse and varies from quantitative information as to the number of effected houses or infrastructure and geographic location, the number of human loss (deaths, injuries, missing or displaced people) [34]. Qualitative information should also be included such as the type of damage (ranging from indirect losses like traffic interruption and disruption of functions to direct economic losses). External information such as the weather/triggering conditions could also be utilized. This section describes about the digital disaster management efforts in Malaysia and Indonesia. Existing open data portals are analysed and suggestions for the infrastructure improvements are discussed.

Disaster Management Portal in Malaysia

Flood is the most devastating natural disaster experience in Malaysia. The stakeholders responsible for disaster management in Malaysia cover the disaster awareness and education, forecasting, mitigation and prevention, alert and emergency declaration, situation diagnosis, and recovery. Malaysia's 11th version of the Five Year Plan (2016-2020) focuses on strengthening disaster risk management across five phases (prevention, mitigation, preparedness, response and recovery). Malaysia continues to develop its disaster management structure and policies to meet emerging and chronic disaster risks, as well as enhance its evolving role as an Humanitarian Assistance and Disaster Relief (HADR) leader in the region.

The National Agency for Disaster Management (NADMA), an agency under the National Security Council is the agency which is responsible for coordinating the national disaster management and responsibility for establishing and ensuring all policies and management mechanisms of the national disaster is followed and implemented at all levels (i.e., district, state and central) of disaster management. NADMA maintains the http://portalbencana.ndcc.gov.my/Portal/ for the announcement of disasters in Malaysia such as flood, haze, landslide, storm and landslide. It also provides weather warning and road closing information, statistics, disaster information archive and photo gallery. However, the portal has several limitations such as access to the information is only restricted to the content being displayed. The information is not current and has not been updated for the past couple of years. The shared information is also not reusable friendly which could be used for integration with other data for analytics purpose.

Another governmental portal on hydrological data is available through http://publicinfobanjir.water.gov.my/ (as in Fig. 1) which is hosted by the Malaysian Department of Irrigation and Drainage (JPS) which provide the map and information from the telemetry station such as daily rainfall last hour rainfall and water level. The map also visualizes the hourly rainfall and water alert level at each station. The portal also provides real-time access to video stream from flood camera at selection stations. Similarly to the MKN's portal, the data provided is mainly only for view. Although the provided values are meaningful for many scientific studies purposes, a download function is not made available. The information specific to the flood occurrence is also not existing (low accuracy) besides traceability, completeness and expiration problem.

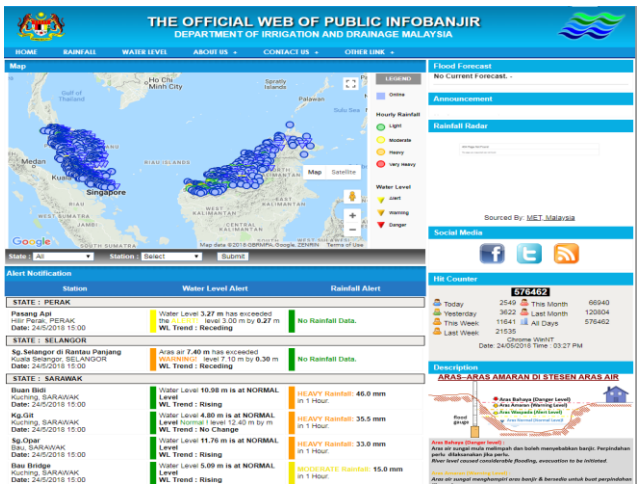

Fig. 1 InfoBanjir Portal

In the data.gov.my portal, the JPS and the Ministry of Natural Resources and Environment has shared several information pertaining to the flood events such as the inventory of flood warning system (eg., water level system, rainfall, flood warning board) in each state, a record on the list of flood hazard map for each state, a record of the impacted area in each state and number of people affected, and the statistics of flood events for each state by focusing on the number of events, average daily rainfall, maximum flood duration, number of victims, maximum depth of flood and loss estimation. This information are in very limited number and mostly are secondary data which provides constraint for external entities such as research organization to perform relevant analysis. The flood risk analysis and flood forecasting systems for Malaysia5 help

emergency response teams coordinate their actions more effectively. The systems are built by combining a range of data including live weather radar data, rainfall measurements, topographical information, details about soil types, land use and drainage. However, the system is not accessible to the public.

Table. 1 Assessment level definition of data quality criteria

Country	Source	Type	Current Situation	Changes situation	News	#affected	#supplies	Logistic info	Emergency Info
Malaysia	http://data.gov.my	Data repository	N	N	N	P	P	P	N
	http://publicinfobanjir.water.gov.my	Alert	Y	Y	Y	N	N	N	Y
	http://saifon.my/	Alert	Y	Y	Y	N	N	N	Y
	http://portalbencana.ndcc.gov.my/portal	Alert	Y	Y	Y	N	N	N	Y
	http://ebanjir.kelantan.gov.my/	Information	N	N	N	P	P	P	Y
Indonesia	https://data.go.id	Data repository	N	N	P	N	N	N	N
	http://inarisk.bnpb.go.id	Risk portal	N	N	P	P	P	P	P
	www.bmkg.go.id/	Alert	Y	Y	P	N	N	N	Y
	www.menlh.go.id	Information	Y	Y	Y	N	N	N	N
	www.bnpb.go.id	Information	Y	Y	Y	P	N	P	Y

Disaster Management Portal in Indonesia

Badan Nasional Penanggulangan Bencana (BNPB) or National Board for Disaster Management launched inaRISK[6], an internet-based portal application to identify disaster risks in Indonesia. In addition, population data living in areas exposed to disaster threats and calculation of potential losses by province, district and city can be accessed through inaRISK (see Fig. 2).

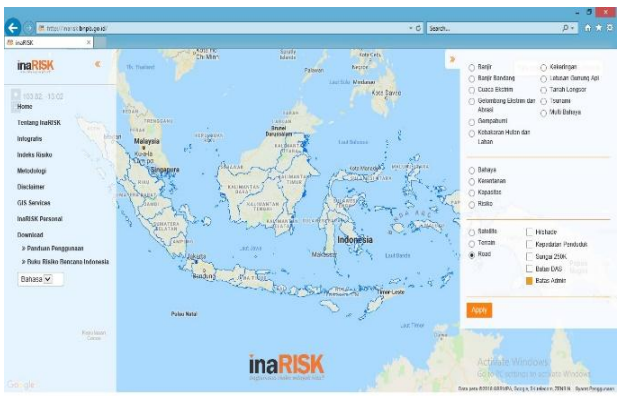

Fig. 2 inaRisk Portal

The inaRISK portal is a portal for disaster risk assessment and monitoring disaster risk index in Indonesia. The portal is built on a Geographic Information System (GIS) server that allows users to use it and very user friendly where users are not required to understand the GIS when exploring the required information. Furthermore, BNPB categorizes ten disaster threats in Indonesia that identified its spread, such as earthquakes, tsunamis, volcano eruptions, floods, flash floods, landslides, droughts, forest and land fires, extreme weather and extreme waves.

The effort to build the inaRISK portal have taken about 7 years, starting with data collection and agreeing on the methodology used. In 2009, each agency still used different methodologies. Agreement on methodology among stakeholders including academics is very important in developing disaster risk assessment. In 2011 the methodology of disaster risk assessment was approved and legalized as Peraturan Kepala BNPB No. 2 of 2012, which is currently in the process of becoming the Indonesian National Standard. Furthermore, the Indonesian Multi-threat Disaster Risk Index is based on the data of 2013. By 2014, the methodology has been developed with insertion from disaster experts from Japan, Australia, New Zealand, Germany so that with the 2015 data generated disaster risk assessment this is used in inaRISK.

Based on disaster risk data, a mainstream disaster planning plan can be formulated with national, provincial, district and city development plans. Even disaster risk data is the first step in the early warning system that BNPB is developing into a multi-hazards early warning system. Furthermore, disaster risk data must be known by all parties, both government, local government, private sector and also the community. All parties can make efforts in anticipation in accordance with the duties and functions of each. BNPB hopes that disaster risk identification can be used by the government, both central and regional to adjust their respective regional development planning. Included in mapping disaster risk types at the district and city level in anticipating and mitigating.

Indonesia Disaster Risk Index Report (IRBI) writes that Indonesia's disaster risk rating cannot be avoided.

Indonesia's geological and geographic location is in the meeting of three giant Eurasian, Indo-Australian, Pacific Plate, and is in the Ring of Fire area (IRBI Report, 2013). The index compiled by BNPB calculates disaster risk in all parts of Indonesia. Many aspects are used, such as the impact of geological, environmental, geographical and demographic positions, economic, and disaster profiles that have occurred, and supplemented with the estimated value of losses incurred.

The Indonesia Disaster Risk Index is calculated based on the formula; Risk = Hazard *(Vulnerability/Capacity). Hazards are calculated based on the average of hazard ranges of frequency and magnitude data from natural hazards such as floods, landslides, earthquakes, tsunamis, and others. Vulnerability is observed based on socio-cultural, economic, physical and environmental parameters. For capacity, data is performed using capacity assessment methods based on parameters of regulatory capacity, institutions, warning systems, skills training education, mitigation and preparedness systems.

Analysis on Information Readiness

In this study the open data portal is defined as the official government open data portal and the accessible disaster data (with comprehensive disaster record attributes) which could be accessed through the web. According to the Sendai Framework, disaster data inventory is among the crucial element to allow good preparation of a country. Although the number of disaster may not be lower, but its impact can be lessened through better preparation of response team. Due to this, we have performed an analysis of the content of open data portal with regards to disaster information availability.

We posit that disaster information accessibility would encourage the nations to perform intelligence and suitable preparation before a disaster strike. This would hopefully contribute to achieving the global targets by year 2030 indicated in the Sendai framework such as reducing (i) disaster mortality, (ii) number of affected people, and (iii) disaster economic loss in relation to global GDP. The Data Readiness Review of Sendai Framework measures the national database for collecting disaster losses including records of comprehensive information on affected people by age, gender, disability type, income, losses (injured/ill, deaths and missing person) and damage (dwellings, livestock, forest) according to geospatial and hazard type. In the following we provide our analysis to match the review content to identify the gaps and proposed a framework which is detailed in section 4.

Table 2 shows the comparison of the portals according to the information required by Sendai framework where 'Y' indicates that the function is available in the portal, 'N' indicates the function is not available and 'P' means part of the function is accomplished. Although source no.5 has most of the information requested as in the Sendai framework, the portal is outdated since the last information dated in year 2014. The alert-based portals have provided intensive information regarding the current situation but the function that allows retrieval of past record is not available for the public.

Beside the explicit comparison as shown in Table 2, we further investigated the usage of DesInventar Sendai[7] which is an open source system of acquisition, collection, retrieval, query and analysis of information about disasters of small, medium and greater impact, based on pre-existing official data, academic records, newspaper sources and institutional report. It is a conceptual and methodological tool for the generation of National Disaster Inventories and the construction of databases of damage, losses and in general the effects of disasters. Data is collected following a set of standards by focusing on disaggregating and is time-stamped and geo-referenced to a minimal geographic area unit of resolution. Malaysia has not been using DesInventar and in the Sendai Data Readiness Report, it is stated that Malaysia will start collecting disaster loss information in year 2018. Although the data on Indonesia disaster from year 1815 until current has been recorded DesInventar, only the occurrence of a disaster by type, date and location are mainly recorded. Data on damage and losses are mainly sparse. The reasons to this are shared by both Indonesia and Malaysia, in terms of current governance of collection of disaster information. The reasons to this scenario are as follows:

i) DesInventar provides an inventory platform but data governance for systematic data collection is not within its scope.

ii) Although support in terms of training, software and consulting service is available, the penetration of DesInventar is still not world-wide because of inability of the countries to establish DesInventar as the tool as to measure the effectiveness of the risk mitigation plans. Some countries carried on the historical research and used it for initial analysis, but did not continue to support the project in an ongoing basis because of high turnover rate in the staff of agencies involved, budget constraints to keep human and other types of resources assigned to the project and over-the-average occurrence of disasters also prevent resources to be kept assigned and with availability.

iii) Collecting data throughout disaster period is challenging because during relief operation, action is performed through endorsement of a committee and according to expert's gauge of the situation. Due to the pressing time for decision to be made, metadata of the event and reason for decision is given less priority and has not been captured. This is mainly because there is no existing system that facilitates the decision-making communication and coordination which results to untraced best practices which is critical for analytic purpose.

iv) Currently some of the required data regarding disaster damage are separated to various operational agencies (eg., the water level records under the Irrigation and Water department, observation on water reaching alert levels is monitored by Public Works department, information on logistics asset is hosted by the Civil Defense Force) and although the emergency relief committee and disaster coordination office could have access to these data, it is time consuming. No direct integration and manipulation of these data

are allowed as each were designed in separate platform specification.

v) Some of the required data are not under the responsibility of specific or existing agency such as the total building area of certain residential building.

vi) Estimating and reporting the damage of certain assets are not performed according to a standardised mechanism and data governance for this purpose is not in place.

In the next part of this section we highlight the analysis of the open data portals with regards to data quality metrics and Next Generation Repository System characteristics. We posit that among the reason that the current open data portal could not accomplish the goals expected in the Sendai Framework is because of the data quality metrics and Next Generation Repository System incompliance. The lack of data ready info-structure has resulted to not merely poor achievement in the Sendai Data Readiness Review but also hamper the research progress in improving disaster management support during pre-disaster, during relief and post-disaster phase. In section 2 we have discussed that whilst more alert-based applications are available, proactive measures based on data is still very scarce, mainly because of the weakness of the info-structure. Table 2 illustrates the comparison of the portal for disaster information management based on data quality metrics [3], [11] where 'L' indicates low achievement, 'M' indicates medium and 'H' indicates high achievement. The definition for each criteria level is listed in Table 3. The results mainly show that currently the provision of digital information has low support from the stakeholders. The definition for each criteria level is listed in Table 4. The data quality criteria is chosen since the focus of this study is on extending the quality of the available data.

Table. 2 Assessment of Open Data Portal based on Data Quality Metrics

Criteria	Malaysia	Indonesia
Currentness	L	L
Understandability	M	L
Accuracy	M	M
Expiration	L	L
Traceability	M	M
Completeness	L	L
Compliance	L	L

Table. 3 Assessment level definition of data quality criteria

Criteria	Low	Medium	High
Current	Mainly old data (before 2014)	Most are date in year 2014-2016	Most are data in 2016-recent
Understandability	No metadata is provided	Some files have accompanying metadata	Most of the files has metadata
Accuracy	Many of the data is inaccurate	Some of the data is inaccurate	Most of the data is accurate
Expiration	The gap between the event and data availability is big	The gap between the event and data availability is moderate	The gap between the event and data availability is small
Traceability	Creation and update tracking is poor	Creation and update tracking is moderate	Creation and update tracking is good
Completeness	Major data loss and incomprehensive	Some data loss and comprehension is moderate	Minimal data loss and comprehension is good
Compliance	Compliance to standard is poor	Compliance to standard is medium	Compliance to standard is good

Table. 4 Assessment level definition of Next Generation Repository System criteria

Criteria	Low	Medium	High
Resource-centric	Limited resource is available	Moderate resource is available	Many resource is available
Networked	Connection to other repositories are limited	Moderate connection to other repositories	Good connection to other repositories
Machine-friendly	Many data are not machine-friendly format	Some of the data are in machine friendly-format	Many of the data are in machine-friendly format
Linking Services	Data are not linked to external resources and traceability are poor	Data are linked to external resources and traceability are moderate	Data are linked to external resources and traceability are good
User activity	Limited user activity is allowed	Some forms of user activity is allowed	Many forms of user activity is allowed
Active	Data is not updated	Some of the data has continuous update	Many of the data has continuous update

Table. 5 Assessment of Open Data Portal based on Next Generation Repository System Criteria

Criteria	Malaysia	Indonesia
Resource-centric	M	L
Networked	M	M
Machine-friendly	L	L
Linking Services	X	X
User activity	X	X
Active	X	X

Table. 6 Assessment of Open Data Portal Content based on Disaster Information Management Services

Service	Malaysia	Indonesia
Structured Information	Location, Water level, Rainfall	Location, Date, Hour Victim, Disadvantages, Descrivtion.
Unstructured Information	News, Picture	News, Picture
Realtime Telemetry Data	Y	Y
Relief Centre Information	Y	Y
Statistics	Y	Y
Visualization	Y	N
Manipulation		
Alert	Y	N
Data Query	N	N
Simulation	N	N
Emergency response	N	N
Awareness education	N	N
Forecast	Y	Y

Table 5 shows the comparison of the open data portal for disaster information management purpose based on the guidelines of technological perspectives provided in the Next Generation Repository System [17], [35] where 'L' indicates low readiness, 'M' indicates medium and 'X' indicates that the criteria is not provided. The linking service highlights the ability to link the provided info with other related resources while the user activity records the user behavior in harvesting the data as well as commenting to the shared scientific data and research outputs. Meanwhile, the active criteria focus on the support of versioning and updating the resources based on the available event evolution.

The comparison of the available services in Malaysia and Indonesia regarding flood and forest fire disaster information management is shown in Table. 6 where 'Y' indicates that the service is available and 'N' indicates that the service is unavailable. The observation results indicate that the portal have provided data mainly for reference through the online based viewing, but no extended support is provided for specific purpose such as raw data viewing and browsing, data query and analytics.

Open Science Cyber-Infrastructure Framework for Next Generation Disaster Analytics

Sustainability of open scientific communities requires the availability of open challenges, problems, and resources so that participants can get motivated and stimulated [12]. Three factors of cyberinfrastructures are the fundamental of open science technology, which includes the management, publication, and discovery infrastructure for research data (Komiyama and Yamaji 2017).

Fig. 3 illustrates the gathered general requirements of open science cyber-infrastructure extracted from the literature referred in this paper which emphasizes the quality metrics (as adopted from [3], [11]) as the base component of the framework to support the management and the technology components of the portal. These components are selected because they could contribute to the sustainability of the framework especially in terms of public usage. The sustainability, social and maintenance component addresses the service promotion, stabilization and version improvements of the portal to energize the technology adoption towards large scale acceptance. The interoperability, diversity and accessibility are the principle of the open data infrastructure to complement the data quality principles.

Existing portals on disaster management mainly focus on statistical-based report and news-based content such as the location impacted, the duration of event, and the number of victims involved. It is proposed that more detailed information is provided such as the elaboration of the event both in the structured data (eg., water level, volume, current speed) and unstructured data (e.g. news, images, video). Shows the scope of disaster information management that should be captured in the open science platform based on the data that are related such as follows:

- the geolocation of the occurred event
- victims impacted (ie., gender, age, health status, length of stay)
- logistics information (eg., vehicles used for evacuation and relief, survival kits/relief supplies distribution)
- damage record (i.e., human, pets, building)
- health records (i.e. disease, treatment)

- evacuation activities (which centre is used, the maximum capacity of the centre
- the population in each centre
- specific demographics in each centre
- related programs conducted during the evacuation)
- relief support parties (e.g., non-governmental or volunteer groups).

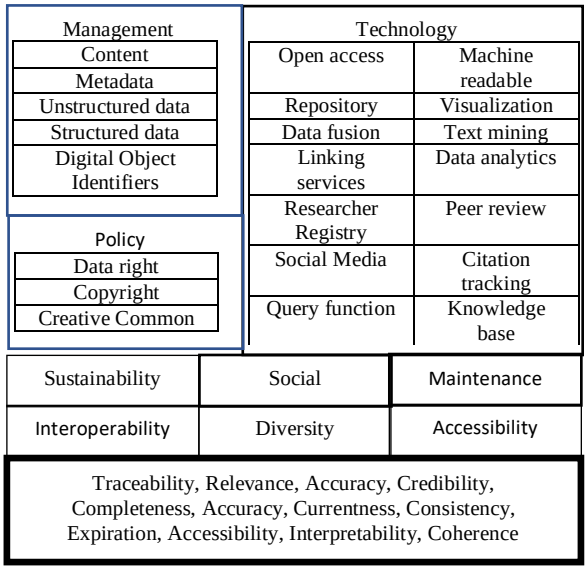

Fig. 3 General Requirements of Open Science Cyberinfrastructure

Besides, the current platforms mainly provide data in single modality only. It is also common for the data providers to separate the representation of a disaster event according to each data and modality type. Since current digital community mainly utilizes mobile applications and social media networks to coordinate data and information with disaster responders, agencies should also use this channel to leverage both the situational awareness and collecting the generated data such as textual posts, mentions and pictures as part of the disaster information source.

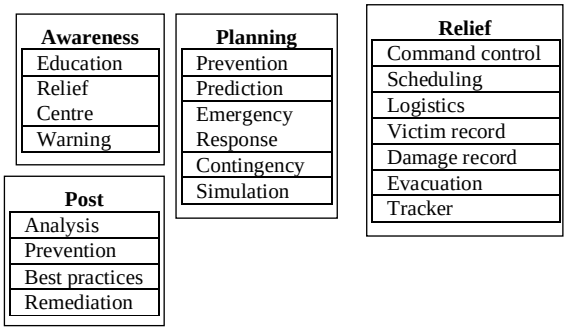

Fig. 4. Disaster Information Management Scope in Open Science Cyberinfrastructure

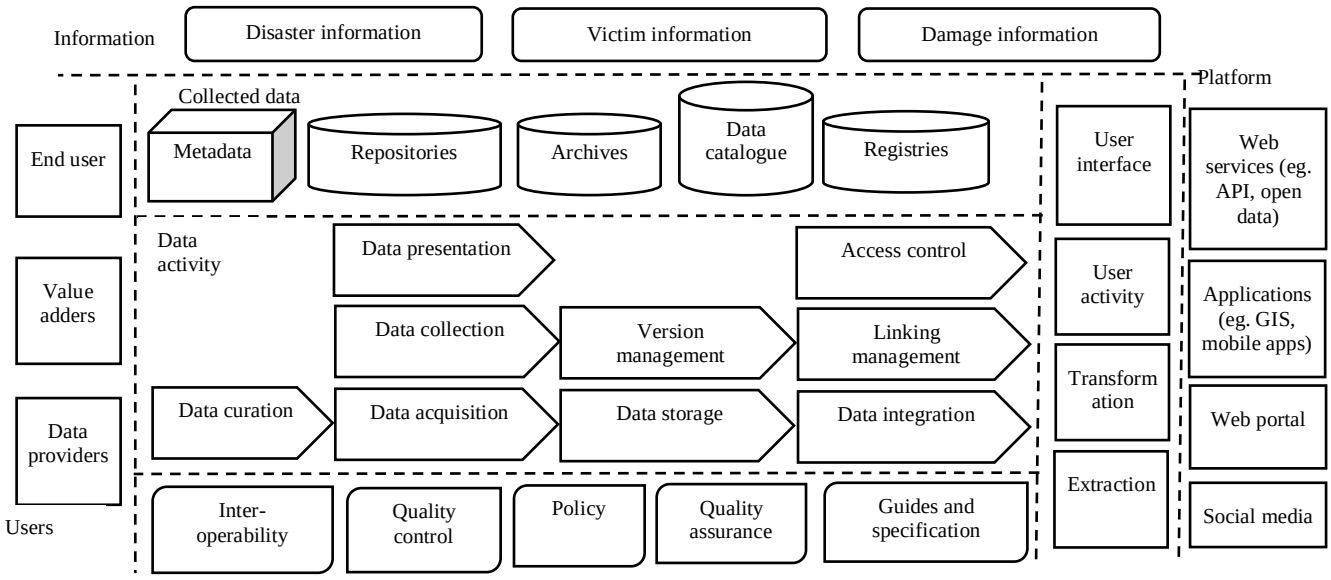

Fig. 5 Open Science Cyber-Infrastructure Framework for Disaster Information Management

Fig. 4 shows the scope of disaster information that should be included in the open science infrastructure. Fig. 5 shows the proposed framework where the availability of the data would allow research and development of many services and tools for the use of the community, government agencies and other relevant stakeholders. The framework is inspired by [37], [38] and modified based on our analysis described in section 3.1.3.

The framework addresses both user and data-centric management of disaster information by emphasizing the data activity capturing which is supported by the interoperability, quality control, quality assurance, policy, and guides and specification.

This is surrounded by the data providers, value adders (eg. users of related apps and reporters), and end user (eg. NGOs, enforcement bodies and researchers) at the right side and the specification of the system at the right side. The collected data in various modalities (capturing text, audio, images and videos of disaster event representation, victims and damage information which are stored through metadata, repositories, archives, data catalogue and registries) enrich the structured data deposited in the inventory such as DesInventar.

The implementation of data activity monitoring could ensure reliability and integrity of the availability of the data. The user interface, user activity, transformation and extraction are components in the framework which indicate the functions that should be created so that the collected information could be exported for web services usage, ingest by applications, web portal and social media. By applying dynamic analytics tools, specific management purpose such as visualization and reporting could be performed. The tool could be benefited also to process the continuously collected data to be transformed into more critical disaster risk reduction actions such as scheduling the response team to address more critical areas and optimizing expenditure to plan for structural based prevention projects. Alert services to the community can also be extended to allow more dynamic communication.

Relevant agencies with technology experts and policymakers with expertise in data sharing, disaster response and analysis, and recovery efforts should establish data sharing policies for disaster response so that stakeholders know how to share data and the data collected are in useful formats. The accuracy of the deposited data should be routinely verified to ensure its usefulness in the event of an emergency. The coordination of available resources would avoid duplicative efforts, and this could be further supported by establishing and maintaining situational awareness to facilitate effective responses. Fraud, waste, and abuse of recovery assistance programs should be identified and prevented if the monitoring of the outcomes of disaster response and recovery programs is conducted.

IV. IMPACTS AND CHALLENGES

Naturally, comparing between Malaysia and Indonesia, more efforts and resources are performed by Indonesia on understanding their risk of disaster due to their geographical location. The results from the current assessment conducted in this research indicate many potential studies including financial, technological, and capacity building. The data readiness is the heart of the disaster risk understanding and loss reduction, as it is reflected as Priority 1 in the Sendai Framework. The development of cyber-infrastructure should be in parallel with the development of policy making, financial budgeting and capacity building.

Considering diverse needs of each geographical locality in the country, regional efforts such as the ASEAN movement to allow mutual-based study for best-practices learning and resource sharing could optimise the process towards improving disaster management. A roadmap and foresight studies by think-tanks for identifying the disaster management decision making improvement should be placed a high priority, such as through special attention by officials in power.

An information exchange framework should be designed by highlighting the storage, machine-understandable data format, network facilities, computer speed specification to enable computation of the unprecedented amount of available data to enable for interoperable and common information communication. Evaluation of current information inventory practise against the Sendai Framework requirement and the facilities provided by UNISDR will be executed especially in terms of financial, technology transfer and capacity building. Related change management issues handling through a structured action plan should be constructed once the infrastructure are in place, so the technology adoption can be smooth.

Besides transformation at enforcement level, a change of culture within the community could also contribute towards producing a reliable information inventory. The integrity of the internet user in contributing into generating data related to the disaster reporting could avoid miscommunication, and worse, catastrophic event from occurring. Some of the actions for redefining the mechanism for data governance may impact the community in terms of their personal financial. For example, revealing the correct private residential building area could lead to increase in certain kind of tax, but accurate knowledge on this could save life during rescue operation. Systematic profiling of citizens in a residential area could smooth evacuation process although some people may not support explicit knowledge on this due to problems such as citizenship legalities.

The promise from an idealistic disaster information inventory could allow enforcer such as welfare office to prepare in advance the life supports according to the potential damage that a person, family and the residential perimeter could be impacted. Proactive measures such as controlling recreational activities nearby streams and river with high risk. Devastating consequences from disaster events through erosion control and river overflowing that could cause flood would help in saving lives.

V. CONCLUSION AND FUTURE WORKS

A shift to data-driven research gradually allows conducting scientific research studies completely in cyber worlds, after the required data sets are obtained or through real time data streams access. Disaster information management is one of the most critical focus in the climate and emergency response of a country due to its disastrous impact to the human, nature and economy. The Disaster information management today, though advanced, are still evolving and their evolution toward more complex and sophisticated systems is expected to continue for the foreseeable future.

The success of the depends on the availability of a state of the art, robust, and flexible cyberinfrastructure, and on the timely access to quality data, products, and tools to process, manage, analyse, integrate, publish, and visualize those data.

Existing scientific works have mainly been in silo and utilize specific sources which are obtained based on special communication for both primary and secondary data request. In the era of big data the sharing of information this should be catered through an open data portal that has the capability to maintain data quality yet offering services that meet the consume needs.

This paper addresses the gap between the open data quality research and the next generation repository system research by utilizing a case study that compares the open data portal in Malaysia and Indonesia. The proposed framework combines the open data portal criteria as and the next generation repository system components which are built based on the requirements of disaster analytics. The implementation of the framework depends on the support of the governmental agency including managerial and administration, financial and technology. This paper is hopefully will provide a reference on the current state of both Malaysia and Indonesia as developing nations which need technological advancement so that systematic disaster management could be facilitated. The open data essentials would encourage many innovations that could help in reducing disaster loss.

The study is currently the main input for an ASEAN research group for risk understanding. The future work from this study is to promote the proposed infrastructure for implementation and focus on the pilot development of specific disaster management analytics based on the requirement of the stakeholders in each country. The quality of the proposed work could be evaluated through a study on the impact of improved data readiness and monitoring the creation of applications that could exploit the collected information.

ACKNOWLEDGEMENT

This study is conducted cross-country thanks to the great cooperation and support across the institutions. This study is a collaboration between the countries towards supporting the ASEAN disaster risk understanding research. The authors wish to express our appreciation to many parties including Prof. Dr. Kzutsuna Yamaji from NII Japan, Assoc. Prof. Dr. Suhaimi Napis from UPM, Malaysia, Mr. Adiratna Wira Adnan and Ir. Bibi Zarina from NADMA Malaysia, and Dr. Asnor Muizan bin Dato Ishak from JPS, Malaysia.

REFERENCES

1. E. Ruijer, S. Grimmelikhuijsen, and A. Meijer, “Open data for democracy: Developing a theoretical framework for open data use,” Gov. Inf. Q., vol. 34, pp. 45–52, 2017.
2. S. Dawes, L. Vidiasova, and O. Parkhimovich, “Planning and designing open government data programs: An ecosystem approach,” Gov. Inf. …, vol. 33, pp. 15–27, 2016.
3. S. Sadiq and M. Indulska, “Open data : Quality over quantity,” Int. J. Inf. Manage., vol. 37, no. 3, pp. 150–154, 2017.
4. J. Attard, F. Orlandi, S. Scerri, and S. Auer, “A systematic review of open government data initiatives,” Gov. Inf. Q., vol. 32, no. 4, pp. 399–418, 2015.
5. J. N. Lee, J. Ham, and B. Choi, “Effect of Government Data Openness on a Knowledge-based Economy,” Procedia Comput. Sci., vol. 91, no. Itqm, pp. 158–167, 2016.
6. S. Bogdanovi and L. Stoimenov, “Benchmarking open government : An open data perspective,” vol. 31, pp. 278–290, 2014.
7. S. Kubler, J. Robert, S. Neumaier, J. Umbrich, and Y. Le Traon, “Comparison of metadata quality in open data portals using the Analytic Hierarchy Process,” Gov. Inf. Q., vol. 35, no. 1, pp. 13–29, 2018.
8. “Portal Data Indonesia,” Unit Keria Presiden Bidang Pengawasan Pengendalian Pembangunan (UKP-PPP), 2014. [Online]. Available: http://data.go.id.
9. D. Gunawan and A. Amalia, “The Implementation of open data in Indonesia,” Proc. 2016 Int. Conf. Data Softw. Eng. ICoDSE 2016, 2017.
10. R. Indonesia, “Rencana Pembangunan Jangka Menengah Nasional 2015-2019,” BUKU I AGENDA PEMBANGUNAN NASIONAL, 2014. [Online]. Available: http://www.social-protection.org/gimi/gess/RessourcePDF.action?ressource.ressourceId=50077.
11. A. Vetrò, L. Canova, M. Torchiano, C. O. Minotas, R. Iemma, and F. Morando, “Open data quality measurement framework : Definition and application to Open Government Data,” Gov. Inf. Q., vol. 33, pp. 325–337, 2016.
12. G. Zou and L. Yilmaz, “Exploring the impact of social communication on the innovation potential of epistemic communities,” Proc. 50th Annu. Southeast Reg. Conf. - ACM-SE ’12, vol. 46, no. 2, p. 315, 2012.
13. Ruben Vicente-Saez and C. Martinez-Fuentes, “Open Science now: A systematic literature review for an integrated definition,” J. Bus. Res., no. in press, 2018.
14. T. Fries, “The Social Factor of Open Science,” in Opening Science, 2014, pp. 271–283.
15. A. Farnham et al., “Early career researchers want Open Science,” Genome Biol., vol. 18, no. 1, pp. 10–13, 2017.
16. S. U. Khan and A. Y. Zomaya, “Data Management: State-of-the-Practice at Open-Science Data Centers,” in Handbook on Data Centers, 2015, pp. 1–1334.
17. K. Yamaji, T. Aoyama, and M. Furukawa, “Development and Deployment of the Open Access Repository and Its Application to the Open Educational Resources,” in Genetic and Evolutionary Computing, vol. 388, 2016, pp. 395–403.
18. R. Weiss and B. Barish, “Social Media , Open Science , and Data Science Are Inextricably Linked,” Neuron, vol. 96, no. 6, pp. 1219–1222, 2017.
19. S. G. D/iya, M. B. Gasim, M. E. Toriman, and M. G. Abdullahi, “Floods in Malaysia: Historical Reviews, Causes, Effects and Mtigations Approach,” Int. J. Interdiscip. Res. Innov., vol. 2, no. 4, pp. 59–65, 2014.
20. A. M et al., “Flood Impact Assessment in Kota Bharu, Malaysia: A Statistical Analysis,” World Appl. Sci. Journa, vol. 32, no. 4, pp. 626–634, 2014.
21. W. Tahir, A. K. Aminuddin, S. Ramli, and J. Jaafar, “Quantitative precipitation forecast using numerical weather prediction and meteorological satellite for Kelantan and Klang river basins,” J. Teknol., vol. 79, no. 1, pp. 45–53, 2017.
22. J. J. Gourley, J. M. Erlingis, Y. Hong, and E. B. Wells, “Evaluation of tools used for monitoring and forecasting flash floods in the United States,” Weather Forecast., vol. 27, no. 1, pp. 158–173, 2012.
23. J. Hardy et al., “A method for probabilistic flash flood forecasting,” J. Hydrol., vol. 541, pp. 480–494, 2016.
24. K. Nagata, “Quantitative Precipitation Estimation and Quantitative Precipitation Forecasting by the Japan Meteorological Agency,” Japan Meteorol. Agency-Technical Rev., pp. 37–50, 2011.
25. Y. R. Chen, C. H. Yeh, and B. Yu, “Integrated application of the analytic hierarchy process and the geographic information system for flood risk assessment and flood plain management in Taiwan,” Nat. Hazards, vol. 59, no. 3, pp. 1261–1276, 2011.
26. A. J. Schroeder et al., “The development of a flash flood severity index,” J. Hydrol., vol. 541, pp. 523–532, 2016.
27. M. A. R. Estrada, E. Koutronas, M. Tahir, and N. Mansor, “Hydrological hazard assessment : THE 2014 – 15 Malaysia floods,” Int. J. Disaster Risk Reduct., vol. 24, no. June, pp. 264–270, 2017.

28. L. Alfieri et al., "A global network for operational flood risk reduction," Environ. Sci. Policy, vol. 84, no. December 2017, pp. 149–158, 2018.
29. P. Tanwattana, "Systematizing Community-Based Disaster Risk Management (CBDRM): Case of urban flood-prone community in Thailand upstream area," Int. J. Disaster Risk Reduct., vol. 28, no. February, pp. 798–812, 2018.
30. UNISDR, "Sendai Framework for Disaster Risk Reduction 2015 - 2030," Third World Conf. Disaster Risk Reduction, Sendai, Japan, 14-18 March 2015., no. March, pp. 1–25, 2015.
31. R. Giordano, A. Pagano, I. Pluchinotta, R. Olivo, S. M. Hernandez, and E. S. Lafuente, "Modelling the complexity of the network of interactions in flood emergency management : The Lorca flash flood case," Environ. Model. Softw., vol. 95, pp. 180–195, 2017.
32. A. Rilo, A. Tavares, P. Freire, P. P. Santos, and J. L. Zêzere, "The contribution of historical information to flood risk management in the Tagus estuary," Int. J. Disaster Risk Reduct., vol. 25, no. July, pp. 22–35, 2017.
33. L. Mancusi and A. Abbate, "Improving flood risk analysis for effectively supporting the implementation of flood risk management plans : The case study of ' Serio ' Valley," vol. 75, no. May, pp. 158–172, 2017.
34. V. Röthlisberger, A. P. Zischg, and M. Keiler, "Science of the Total Environment Identifying spatial clusters of flood exposure to support decision making in risk management," Sci. Total Environ., vol. 598, pp. 593–603, 2017.
35. T. Yamaji, K., Aoyama, T., Furukawa, M., Yamada, "WEKO and to JAIRO Cloud: Development and Deployment of the Open Access Repository and Its Application to the Open Educational Recourses," Adv. Intell. Syst. Comput., no. 388, pp. 395–403, 2015.
36. Y. Komiyama and K. Yamaji, "Nationwide Research Data Management Service of Japan in the Open Science Era," in Proceedings - 2017 6th IIAI International Congress on Advanced Applied Informatics, IIAI-AAI 2017, 2017, pp. 129–133.
37. A.Mansouriana, A. Rajabifard, Z. I.Williamson, and M.J.Valadan, "Using SDI and web-based system to facilitate disaster management," Comput. Geosci., vol. 32, no. 3, pp. 303–315, 2006.
38. B. Bjerge, N. Clark, P. Fisker, and E. Raju, "Technology and information sharing in disaster relief," PLoS One, vol. 11, no. 9, pp. 1–20, 2016.