Open Data Reuse, Recycling and Sharing as Potential Solution to Data and Information Resource Inadequacies

Kyla Matias, Tetsuo Kidokoro, Scira Menoni, Ouejdane Mejri, Negar Aminoltaheri

Open Data Reuse, Recycling and Sharing as Potential Solution to Data and Information Resource Inadequacies

Kyla Matias
Department of Urban Engineering, The University of Tokyo

Tetsuo Kidokoro
Department of Urban Engineering, The University of Tokyo

Scira Menoni
Department of Architecture and Urban Studies (DASTU), Politecnico di Milano

Ouejdane Mejri
Department of Electronics, Information and Bioengineering, Politecnico di Milano

Negar Aminoltaheri
Faculty of Civil, Environmental and Territorial Engineering, Politecnico di Milano

ABSTRACT. This paper explores the reuse, recycling and sharing of open data (OD) as a potential solution to bridge gaps in existing baseline data and information in areas with scarce data resources. It focuses on open data generated during disasters, and analyses how these voluminous 'free disaster data,' such as social media posts, images, damage assessment reports, etc., could be reused and recycled to serve purposes other than emergency response and relief. To illustrate this, the paper makes use of a previous research that analysed how the open data of Super Typhoon Haiyan, the hydrometeorological disaster which affected several nations, and gained both local and global attention, could be reused and recycled as inputs for development planning, especially in post-disaster recovery and rehabilitation, and pre-disaster mitigation and prevention planning.

1   INTRODUCTION

Information and communications technologies (ICT) are the impetus of this information age (Masuda, 1981) wherein data and information are the new raw materials (Maeng & Nedovic-Budic, 2007). Given this crucial role in the development and growth of societies, it may be said that data and information are the most valuable commodities of planning (Lord, 2012).

As an inherently knowledge-intensive exercise (Goodspeed, 2011), planning requires a steady stock of readily available and reliable data and information that could provide causal knowledge (Wildavsky, 1973) that is needed to support strategy development and inform decision-making.

Fortunately, current ICT advancements are able to fulfill this demand through new and innovative data collection technologies and techniques, and wider data distribution channels (Sustainable Development Solutions Network (SDSN), 2015), which has made data generation and exchanges easier and faster. However, this is a privilege that is not enjoyed by all due to disparate development conditions and unequal distribution of resources (Graham, 2002). These conditons pose limitations and create tensions, especially in developing regions (Maxwell & Reuveny, 2000), where varying issues and interests abound and compete for the allocation of scarce resources. These areas struggle to sustain and boost growth and development while also trying to address other pressing issues, such as climate change, environmental hazards or disasters (UNISDR & ESCAP, 2012). As a result, the obvious and perceptible problems get prioritised, while other "trifling" issues, such as the need to improve one's baseline data and information or geospatial data infrastructures, end up filed away and forgotten until the moment of necessity presents itself again.

Although data and information only form a part of the requirements for promoting and boosting one's growth and development, it plays an important part in accurately identifying vital concerns that must be put on top of the agenda. It is also required for monitoring and evaluating the performance and effectiveness of implemented programs and strategies so that adjustments could be made accordingly in order to avoid further resource wastage, which is crucial for disadvantaged communities. As such, greater attention must be accorded to these data and information gaps if we were to plan effectively and efficiently towards promoting sustainable development, and safeguarding achievements and gains.

In response to the issue, the paper offers an insight on how present ICT trends and developments could help deal with the problem, and explores potential alternatives that could ease data and information resource inadequacies, such as harnessing the potential of open data (OD). The next section discusses factors contributing to the problem, including conceivable hindrances to its resolution. It is followed by a case study that presents a practical application of OD reuse, recyling, and sharing, and concludes by discussing the perceived benefits and challenges that could be the subject of future investigation.

2 DATA AND INFORMATION RESOURCE INADEQUACY ISSUES

Data and information gaps come in various forms such as being unavailable, inaccessible or having questionable integrity and quality (World Bank, 2014). While it is obvious that the absence of data and information can pose serious consequences, the utilisation of poor quality data and information may also lead to undesirable outcomes. Both were witnessed in past disaster events, such as the 2010 Haiti Earthquake where unmapped areas and roads had to be traced to aid emergency response, and the 2013 Typhoon Haiyan in the Philippines, wherein storm surge hazard maps underestimated the maximum inundation levels, which led to a false sense of security and complacency. The aforementioned examples showed that the lack of information slowed down emergency response and relief efforts, while the use of imprecise data aggravated the disaster's ill effects. Both of these unfortunate situations have contributed to the larger number of casualties, and could have been averted if better information had been available and accessible.

This problem is not only evident in disaster management, its crucial effects also extend to other components of development: social services, economy, etc., and could could translate to poorer services, imbalanced growth and development trends, and although indirectly, lesser investment opportunities.

In everyday life it may seem that data and information issues do not pose grave immediate threats nor evoke a "real" sense of urgency, but the examples illustrate that leaving it unaddressed could impact future events. It also stresses the importance of having and sharing reliable data and information in order to fully understand various issues, and avoid making erroneous choices.

This problem is usually more critical in underdeveloped and developing regions where baseline datasets and information are scarce and often outdated. Understandably, there is a cost in addressing these issues, and it would require upfront investments being made. This cost becomes a major inhibiting factor for those that are typically plagued by this problem. For in most cases, the creation and access to data and information comes at a price tag that these lower and middle-income economies could hardly afford.

These are great challenges that could discourage and hinder willingness to genuinely address the issue, but the need to protect development gains and societies' well-being, as well as ensure the sustainability of progress, should be accorded with much greater significance, and this can only be attained when there's a full understanding of issues and situations so that the appropriate decisions and actions are made with sound judgment.

3 REDUCE, REUSE, RECYCLE: THE POTENTIAL OF OPEN DATA AS SOLUTION TO DATA AND INFORMATION INADEQUACIES

There is no skirting around the fact that resources must be allotted to resolve data and information inaedquacies. Fortunately, recent ICT developments could be utilised to devise more affordable and innovative solutions to this problem.

With the proliferation of the internet and the World Wide Web, data and information creation, access, and sharing have become convenient for all, especially the public (Stauffacher, Hattotuwa, & Weekes, 2012). These incessant technological advancements have also led to the breaking down of physical boundaries. Previously, data and information sharing have been confined within common spaces, but now it is possible to access and share data and information even when you are in a different geographical region. This has been observed in the recent disaster events where in people from all over the globe actively participated in disaster information exchanges, as well as "remote disaster response efforts." The ease of access and sharing is a great feat made possible by modern technology, and it has allowed greater civic engagement and unrestricted information exchanges that has further led to an exponential growth of freely accessible information (Stauffacher, Hattotuwa, & Weekes, 2012) such as ODs.

In order to demonstrate how ODs can be reused and recycled for purposes other than its original intent, the case of Tacloban City in the Philippines was selected. The reasons for this are (1) its involvement in a recent event that attracted immense attention and efforts to the area, (2) the availability of widespread materials generated by various volunteers and organisations around the globe, and (3) access to official maps and information supplied by the local authorities.

3.1 TYPHOON HAIYAN QUICK FACTS AND FIGURES

In the early weeks of November 2013, Typhoon Haiyan, a Saffir-Simpson Category 5 typhoon, struck the Asia-Pacific region. It affected several countries including the Federated States of Micronesia (FSM), Philippines, Vietnam and China. Nonetheless, the massive destruction was concentrated in the Philippines where 9 out of 17 administrative regions of the country were disturbed. It was later reported that Haiyan claimed approximately 6,300 lives and caused immense damage amounting to an estimated $2B US dollars (National Disaster Risk Reduction and Management Council (NDRRMC), 2014).

Tacloban City in the Province of Leyte, Philippines suffered the most due to the unexpectedly high inundation from the storm surge, which was further aggravated by the robust winds. Despite carrying out pre-emptive evacuation, the number of casualties still soared as some of the evacuation centres were also submerged (ABS-CBN News, 2013). As shown in Table 1, in Tacloban City alone, there were 2,678 casualties, which accounts for the 42.5% of the total number of fatalities in the country, and nearly 60,000 homes were destroyed and damaged causing the displacement of families, and leaving many parents and children bereaved.

Table 1. Data on Affected Households and Population in Tacloban and the Philippines

Haiyan Affected Statistics	Total Count (Tacloban City)	Total Count (Leyte Province)	% Share of Tacloban vs. Leyte Province	Total Count (Philippines)	% Share of Tacloban vs. PH
Damaged Houses	58,823	381,810	15%	1,140,332	5.15%
Household (HH)	58,823	441,588	13%	3,424,593	1.7%
Persons	552,836	2,364,023	23.4%	16,078,181	3.4%
Fatalities	2,678	5,402	49.6%	6,300	42.5%
Injured Persons	-	15,672	-	28,688	-
Missing Persons	701	931	75.3%	1,062	66%

Source: National Disaster Risk Reduction and Management Council (NDRRMC), 2014

3.2   DATA DELUGE: MINING HAIYAN'S VOLUMINOUS OPEN DATA

Recent technological advancements have allowed both the authorities and the public to forecast imminent disasters, and prepare for it. The case of Typhoon Haiyan was no exception. Despite the soaring number of casualties brought about by the disaster's overwhelming destruction, it was a well-anticipated event whose enormous magnitude piqued various interests, and captured both local and global attention and concern.

Countless contributors from all over the globe, and all walks of life took advantage of the world wide web to collect and share data and information regarding the super typhoon, which eventally led to the generation of innumerable ODs. These ODs came in various formats and from different platforms such as text-based or image-based social media posts, interactive disaster simulation models, satellite images as well as dynamic maps.

These various data and information became useful during the course of the disaster as it aided preparedness, and response and relief efforts. However, after the catastrophe, many of these ODs and information will not be utilised, but will still remain in the wild. Given that many of these contain significant information that were once outdated or inaccessible, especially to those with scarce data and information resources, it may be worth considering to recycle and reuse these, and harness their potential.

With the primary objective of identifying how ODs generated during disaster events could be applied for other purposes other than the original intent, Typhoon Haiyan's case was used and its ODs were later collected manually using search engines, RSS feeds, social media, etc.

The resulting collection had diverse characteristics, which offered a wealth of information, but required tedious work to organise, and analyse prior to application and use. These were simplified through tabulation as shown in Table 2, which contains a partial list of the collected ODs. These were later grouped into different typologies using taxonomy and ontology, such as Figure 1, which provided a visual representation of the collected data's typologies and relationships. This process of grouping and reorganising of the data helped uncover similarities in the nature of the collected data,

and its potential usage. Most of the collected ODs were spatially-enabled, and had similar characteristics to the data requirements of local development planning. As such, reusing ODs as alternative inputs for local development planning of the typhoon afflicted area could be one of the practical applications of disaster-related ODs, and a suitable model for OD reuse and recycling.

3.3   HAIYAN'S OPEN DATA REUSE AND RECYCLING

In the Philippines, land use planning is commonly used by local authorities to guide area development. Coincidentally, there is also a prevalence of gaps in the existing planning datasets and information stock. Bearing this in mind, Haiyan's ODs were tested as alternative inputs for planning. This would also determine whether the retrieved open data may be reused and recycled to serve another purpose, and if it could help bridge existing data and information gaps.

Given the huge variety of the ODs, only those data that were spatially-enabled, and specfic to the study area were chosen to be used. These data are listed in Table 2, and were applied as inputs in a land use planning exercise that is used to evaluate land suitability, commonly known as sieve mapping. It is done by placing maps over another to produce new images that would reveal potential opportunities and threats for the area that is proposed to be developed.

In order to carry this out, the collected ODs need to be combined with other relevant data from the study area, such as administrative and hazard maps. These were obtained from the local and national authorities. (See Figures 2 and 3.)

By combining these official maps and the open data collected, new maps were generated. (See Figures 4 and 5.) Unfortunately, the collected ODs and the information supplied by the authorities came in varying formats and scales, which restricted its manipulation. As a consequence, scale has been sacrificed during the sieve mapping. Thus, the resulting overlays are more of approximations for illustrative purposes. Nevertheless, the resulting overlays fairly give us an idea how an area might have changed due to the recent disaster. It is also useful to highlight issues and problems that might have been previously overlooked during plan creation or review. Any finding from the new overlays could be later verified and further studied prior to making the necessary amendments in existing plans.
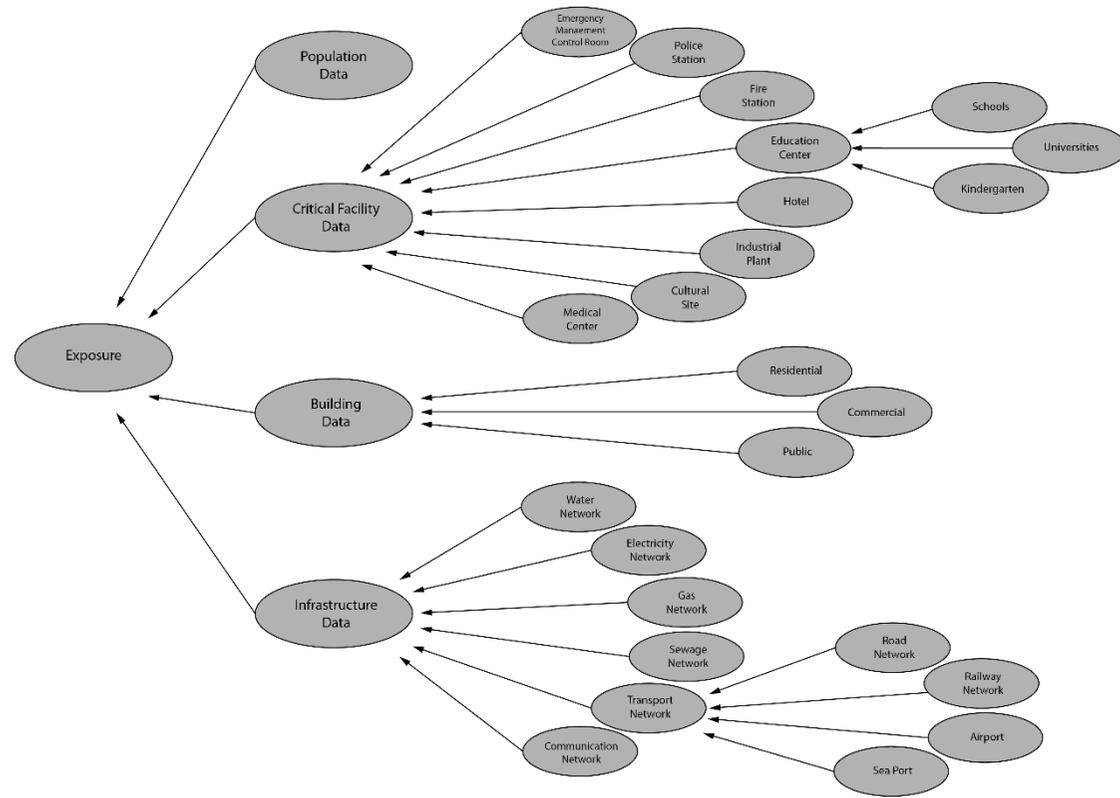
In the case of the generated maps using Haiyan's OD, it highlighted the weakness of the previous storm surge hazard map that have only illustrated modest levels of inundation as depicted in Figure 3. Comparing this with Figure 4, the overlay of the administrative map and the storm surge simulation map from Harvard's World Map site clearly shows that inundation could, and actually extended beyond the identified areas in the official storm surge hazard map. The OD map depicted a scenario that was closer to the actual flooding that happened during the typhoon.

Additionally, Figure 5, the overlay of the storm surge model, crisis information and Tacloban's proposed land use, highlights other potentially hazardous areas, such as areas designated as future residential zones, that could be further evaluated during future amendments.

Table 2. Haiyan Data Sorting Table

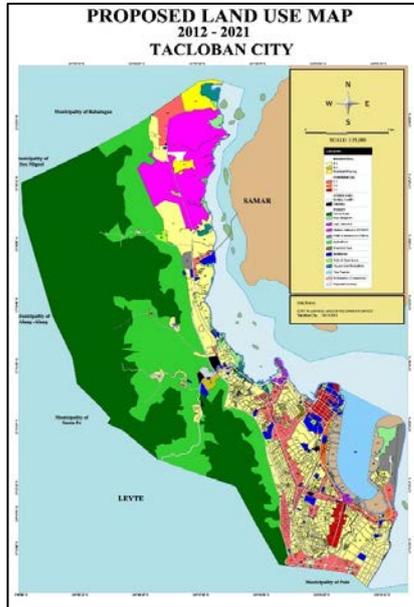| Document Name	Document Type					Area of Interest (Location of Incident or Event)	Source / Origin (Name of generator, producer, owner)	OD Prod.		EM Data Ontology																	Remarks / Other Details	Link to Document Location (if any)
										Context										Crisis								
										Exposure				Resource						Hazard			Disaster Consequence					
	Text	Image	Static Map	Georeferenced Map	Others (State file type)			Volunteered Data	Traditional	Population	Building	Critical Facility	Infrastructure	Human	Monetary	Vehicle	Vital Line	Shelter	Equipment	Monitoring	Alerts	Incidents	Direct Loss	Direct Damage	Indirect Loss	Indirect Damage		
Phils. Typhoon Haiyan/ Yolanda Map (Aggregated)							Harvard / World Map	x	x			x	x					x				x		x			Dynamic / Interactive Map	http://worldmap.harvard.edu/maps/typhoon_haiyan
Crisis Map						PH	GISCorps / Micromappers	x														x	x	x			Urgent need Tweets/ Info about needed resources not available resources.	http://giscorps.maps.arcgis.com
Tacloban Damage Assessment Report					Vector Files	PH	Copernicus		x															x			Scale 1:12,500	http://emergency.copernicus.eu/mapping/list-of-components/EMSR058
Various Land Use Maps						Tacloban	Tacloban LGU		x		x		x														Scale 1:25,000	
Flood Susceptibility Map						Tacloban	MGB		x				x							x							Official map for validation use / Scale 1:50,000	
Tacloban Storm Surge Hazard Map						Tacloban	Tacloban LGU		x											x							Official map for validation use / Scale 1:17,000	
Landslide Hazard Map						Tacloban	Tacloban LGU		x											x							Official map for validation use / Scale 1:17,000	
Comprehensive Land Use Plan					Table/ Graph	Tacloban	Tacloban LGU		x	x	x	x	x							x								
Tacloban Infrastructure						Tacloban/ PH	Tacloban LGU		x				x															

Source: Author's own work

Fig. 1. The Taxonomy of Exposure Data

Table 3. Charting the Overlapping of Emergency Management and Spatial Planning Data
Requirements based on the Disaster / Emergency Management (D/E)M Data
Taxonomy and Ontology

Disaster / Emergency Management (D/EM) Data Ontology			D/EM Cycle Phase Use				Plan Use	Remarks
			M/P	P	R	R		
Context Data								
1	Exposure	Population	x	x	x		x	May be useful for future planning
		Building	x	x	x		x	
		Critical Facility (CF)	x	x	x		x	CF damage and consequence required in recovery phase
		Infrastructure	x	x	x		x	May be useful for future planning
2	Resource	Human	x	x	x		x	May be useful in recovery phase and future planning
		Monetary	x	x	x		x	Necessary for resource allocation and prioritisation decisions
		Vehicles		x	x		x	May be useful in future planning (esp. considering response)
		Vital Lines		x	x		x	May be useful in recovery phase and future planning
		Shelter		x	x		x	Proper location identification during SP
		Equipment	x	x	x		x	Specific equipment considered in SP
Crisis Data								
3	Hazard	Monitoring	x	x	x	x	x	Regular monitoring including historical data is useful for SP
		Alerts		x	x			May be useful in recovery phase with secondary effects
		Incidents			x		x	Past incidents can be used as basis for future mitigation through SP
4	Consequences	Direct Loss	x	x	x	x	x	May be used as references in M/P, P, R to cut down future losses
		Direct Damage	x	x	x	x	x	
		Indirect Loss	x	x		x	x	
		Indirect Damage	x	x		x	x	

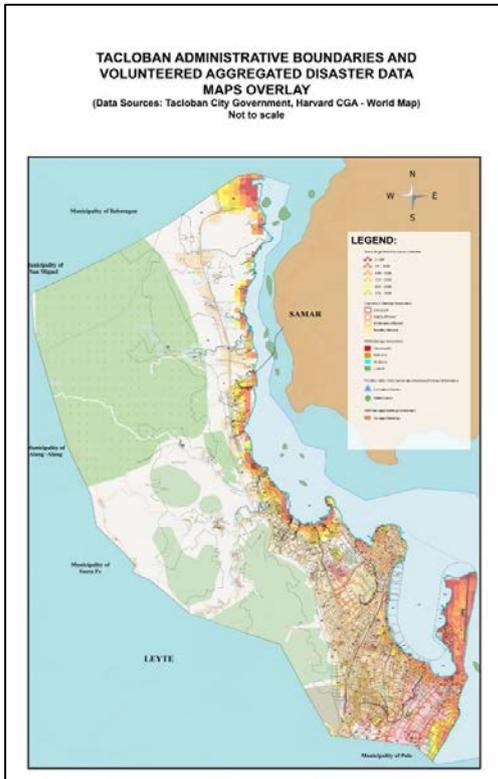**Fig. 2.** Tacloban City's Proposed Land Use Map

Source: Tacloban LGU, 2012

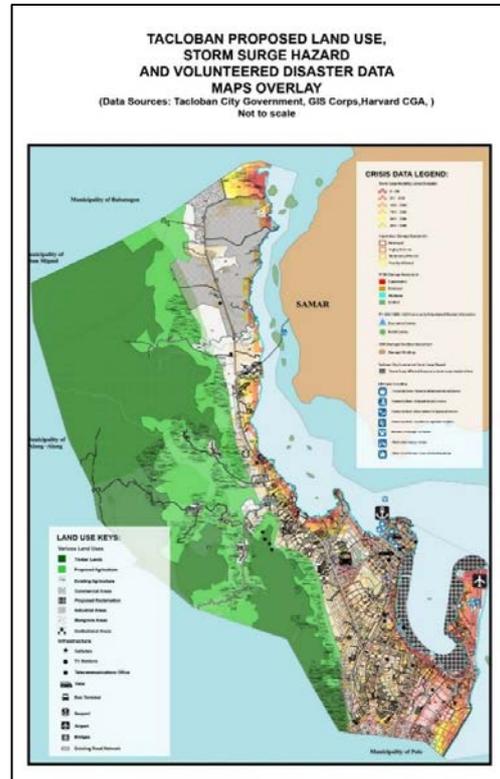**Fig. 3.** Official Storm Surge Hazard Map before Haiyan's Occurrence

Source: Tacloban LGU, 2012

Fig. 4. Overlay of Administrative Boundaries and Storm Surge Model Map

Source: Author's own work

Fig. 5. Overlay of Land Use Map and Storm Surge Model Map

Source: Author's own work

4   CONCLUSION

Scarcity of resources has been the classic excuse for incompetence, slow growth or stunted development. Though it may be true that the availability of resources such as data, information and intellectual capacities influence one's own ability to be comprehensive (Lindblom, 1959) when tackling problems, there are novel ways to deal with these inadequacies.

In response to these insufficiencies in data and information, the example cited in this paper investigated how open data can be reused and recycled to fill in these gaps. It analysed the case of Typhoon Haiyan in Tacloban, Philippines to show how freely available and publicly accessible data that were generated during disaster events could be retooled for local development planning, especially in pre-disaster mitigation and prevention, and post-disaster recovery and rehabilitation scenarios. It could be useful in aiding planners to craft more risk-sensitive development plans that would integrate appropriate disaster management strategies for their respective areas.

Various ODs of Typhoon Haiyan were collected, but priority was given to those with spatial references. These were utilised as inputs in a mock sieve mapping, an exercise that was commonly utilised for land use planning within the study area. The resulting overlay maps have highlighted some inconsistencies in the proposed land use plan and existing hazard maps, which might have been overlooked during plan creation and review, or could have been altered by the recent event, and may be investigated further in future planning exercises and research. Even though it was a test case, it demonstrated the possibility of reusing and recycling open data as an alternative input, specifically in situations where data and information resources are limited.

However, reusing and recycling ODs is not without challenges. Despite its potential, various concerns may arise, such as the need for data validation, source credibility check, the addition of requirements for technical capacity building for data manipulation, interpretation and utilisation, and for ensuring interoperability among different datasets. These concerns underline the problem's complexity, and underscores the fact that while scarcity of resources poses limitations, its abundance does not necessarily translate to more effective plans and strategies.

Such complex problems will always be difficult to solve using traditional methods and policy-making as more often than not, existing resources are not as unbounded as it assumes it to be. This is not just true for money and time, but also for knowledge and intellectual capacities (Lindblom, 1959). Innovation and education can help overcome these. Teaching communities about the significance of data and information in decision-making, and promoting a culture of data appreciation, literacy and sharing will enable communities with minimal to no technological resources to still benefit from modern ICT advancements, and adapt using local means to create and retrieve pertinent information required to wield sound judgment, and take appropriate courses of action until such time that better data and information infrastructures could be afforded.

REFERENCES

1. ABS-CBN News. (2013, November 11). *Children, elderly drown as flood waters swallow Tacloban evac center.* Retrieved November 11, 2013, from ABS-CBN News.com: www.abs-cbnnews.com/focus/11/09/13/children-elderly-drown-waters-swallow-tacloban-evac-center

2. Goodspeed, R. (2011, February 14). *Knowledge Management for Planning Organizations.* Retrieved from Planetizen: http://www.planetizen.com/node/48125

3. Graham, S. (2002). Bridging Urban Digital Divides? Urban Polarisation and Information and Communications Technologies (ICTs). *Urban Studies, 39*(1), 33-56.

4. Kowalczyk, S., & Shankar, K. (2011). Data Sharing in the Sciences. *Annual Review of Information Science and Technology, 45*(1), 247-294.

5. Lindblom, C. (1959). The Science of "Muddling Through". *Public Administration Review, 19*(2), 79-88.

6. Lord, A. (2012). The Planning Game: An information economics approach to understanding urban and environmental management. Oxon: Routledge.

7. Maeng, D.-M., & Nedovic-Budic, Z. (2007). Urban Form and Planning in the Information Age: Lessons from Literature. *Spatium*, 1-12.

8. Masuda, Y. (1981). *The Information Society as Post-Industrial Society.* Bethesda, Maryland: World Future Society.

9. Matias, K., & Aminoltaheri, N. (2014). *Linking Spatial Management and Emergency Management through Open Data.* Politecnico di Milano. Milan: (Unpublished Master's Thesis).

10. Maxwell, J., & Reuveny, R. (2000). Resource Scarcity and Conflict in Developing Countries. *Journal of Peace Research, 37*(3), 301-322.

11. National Disaster Risk Reduction and Management Council (NDRRMC). (2014). *Final Report re Effects of Typhoon "Yolanda" (Haiyan).* Quezon City: NDRRMC. Retrieved from National Disaster Risk Reduction and Management Council (NDRRMC).

12. Serote, E. (2008). *Rationalized Local Planning System in the Philippines.* Bureau of Local Government Development - Department of Interior and Local Government (DILG).

13. Stauffacher, D., Hattotuwa, S., & Weekes, B. (2012, March 23). *The potential and challenges of open data for crisis information management and aid efficiency: A preliminary assessment.* Retrieved from ICT4Peace Foundation: http://ict4peace.org/?p=2334

14. Sustainable Development Solutions Network (SDSN). (2015, April 17). *Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development.* Retrieved from Sustainable Development Solutions Network (SDSN): http://unsdsn.org/wp-content/uploads/2015/04/Data-for-Development-Full-Report.pdf

15. UNISDR & ESCAP. (2012). The Asia-Pacific Disaster Report 2012: Reducing Vulnerability and Exposure to Disasters. Bangkok: United Nations.

16. Urbancic, T., Stepankova, O., & Lavrac, N. (2006). Enhancing Human Choice by Information Technologies. In J. Berleur, M. I. Numinen, & J. Impagliazzo (Ed.), *IFIP International Federation for Information Processing, Volume 223, Social Informatics: An Information Society for All? In remembrance of Rob Kling. 223*, pp. 255-264. Boston: Springer.

17. Wallis, J. C., Rolando, E., & Borgman, C. L. (2013, July 23). If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology. *PLoS ONE, 8*(7).

18. Wildavsky, A. (1973). If Planning is Everything, Maybe it's Nothing. *Policy Sciences, 4*, 127-153.

19. World Bank. (2014). *Open Data for Resilience Initiative: Field Guide.* Washington DC: The World Bank.