Utilizing Openstreetmap Tags for Road Toponyms in Post Tsunami Reconstruction Area: The Lesson from Aceh, Indonesia

Indra Roharatua HUTABARAT, Memby Untung PRATAMA, Indonesia

Keywords: Policy, Toponyms, Roads, Post Tsunamy Reconstruction, String Comparisons, Semantic meaning

summary

The Indonesian Government has been applying the policy to exploit Open Source Data in order to reduce cost, to improve and to enrich existing mapping products by using Openstreetmap (OSM). This project tries to demonstrate utilizing OSM tags for NLA road toponyms in urban and rural area in Aceh Province, Indonesia which had been impacted by tsunami in 2004. That disaster caused significant impact towards the geography of Aceh, especially man-made features like roads and buildings. During the post-disaster reconstruction, there had been major changes in topography and geography objects like roads and buildings. In addition, Indonesia government through The Ministry of Agrarian and Spatial Planning/National Land Agency (NLA) and Indonesian National Board for Disaster Management (BNPB) concern in Post Tsunamy Reconstruction and Hazard Mitigation. One of their project is InaSafe (GFDRR 2014), it is "a plugin for QGIS software aiming to produce realistic natural hazard impact scenarios for better planning, preparedness and response activities, using hazard and exposure geographic data". It significantly related with OSM because it can be utilized for analysis in Inasafe. Here, the community collected details and updated the geographic information, such as road toponyms which had been changed after tsunami and post-disaster reconstruction. (GFDRR 2014). By the fact, Aceh had been impacted by the tsunami makes it as a priority for the government in order to prevent such a great lost from ever happening again. Then, BPNB invites and supported community mapping in Aceh. So, OSM tags in Aceh Province can be assumed as updated data by given from local knowledge contributors.

This project is expected to improve the quality and coverage of toponyms of NLA Map, and can determine the benefits and limitation of OSM data. OSM tags were compared to NLA toponyms by using string comparisons, assessing road names semantic meaning, and measuring their road names existence. As results OSM tags can help to fix NLA toponyms mistakes that derived from toponyms survey, give another perspective about the road names in different local languages, and could enrich NLA roads which do not have attributes. Yet, the existence of road's OSM tags is the limitation on this projects as it is needed in string comparisons.

Utilizing Openstreetmap Tags for Road Toponyms in Post Tsunami Reconstruction Area: The Lesson from Aceh, Indonesia

Indra Roharatua HUTABARAT, Memby Untung PRATAMA, Indonesia

1. INTRODUCTION

Maps become part of our life recently. In our daily life, a map can help people for making their work easier. For instance, we can use the digital map for looking the address or even for driving in the traffic jam situation. Map has brought new perspective in understanding our environment through the use of geospatial information (Perdana, et al, 2012)

Maps have been already used as tools in planning and management and also for disaster management. The impact of disaster is a changing of location surface and the victim form the disaster. Regarding this paper topic, the authors try to utilize openstreetmap that developed In Post Tsunami Reconstruction Area for Indonesia's Authorize Map. Location is related to geographical position and name of place. When we have no good information about the location or we have lack of information regarding the number of topographic maps for the area where disaster happen, and in other side the only information that they know is name of place, so database of toponyms on maps is a key to solve it and connect or enter to related information they need.

Indonesia as one a big country in South East Asia realize how important the maps. With total land around 1.9million km^2 and total sea area around 3.5million km^2, they undertake to create maps for whole Indonesian territory. Yet, mapping in Indonesia, which has more than 17,000 islands and 100s languages (Statistics Indonesia 2014), is complex and expensive and this is a job that will never be completed in a quick time.

Ministry of Agrarian and Spatial Planning/National Land Agency (denoted as NLA) is responsible in creating Base Map large scale area, cadastral mapping and land title. The main task of NLA is to provide Base Maps which will be used for cadastral map in every municipality on whole country.

1.1 A brief data of Research Location and Tsunami

The chosen study area for this research is Aceh Province, Indonesia which is located in Northern end of Sumatera Island. The total Area is about 57.9thousands km^2 with approximately 5millions in population. The two municipalities are chosen to represent samples both in urban and rural areas. Banda Aceh will represent urban area while Aceh Besar will reflect rural area. Banda Aceh is the capital of Aceh province with total area is around 61.36 km^2 with 249,282 populations while Aceh Besar is about 29,741 km^2 with population 369,972 (Statistics Indonesia 2014).

Indonesia as a multicultural country has various kinds of local language (aside from Indonesia Language as the national/formal language) spreads all over the nation, including in Aceh. There are several local languages which is used by community in Aceh Province, which are Acehnese, Gayo,

Alas Aneuk Jamee, Singkil and Pakpak, Kluet, Tamiang, Sugalin and Lekon, and Haloban and Nias. Local people in Banda Aceh and Aceh Besar use Acehnese language. It is a language that is spoken by Acehnese who are in coastal areas, mostly inland and some islands in Aceh. Acehnese language is included in Chamic language family, a branch of the Malayo-Polynesian branch of the Austronesian language family. This local language, could affect the toponyms in that area.("Belajar Bahasa Aceh [Learning Aceh Language]" 2015).

They are several reasons why is Aceh province taken as study area on this project. First, Banda Aceh has been impacted by Tsunami in 2004. That disaster caused significant impact towards the geography of Aceh, especially man-made features like roads and buildings. During the post-disaster reconstruction, there had been major changes in topography and geography objects like roads and buildings. Secondly, Indonesia government through Indonesian National Board for Disaster Management (BNPB) concerns in hazard mitigation. One of their project is InaSafe (GFDRR 2014), it is "a plugin for QGIS software aiming to produce realistic natural hazard impact scenarios for better planning, preparedness and response activities, using hazard and exposure geographic data"(InaSAFE 2015). It significantly related with OSM because it can be utilized for analysis in Inasafe. Here, the community collected details and updated the geographic information, for example what the name of roads in area, how many houses will be affected by a specific hazards and what is the best routes for escape from that hazards (GFDRR 2014).

Figure 1 Location of Aceh Province in Indonesia (Google Maps, 2015)

1.2 The objectives of paper

This paper has an aim to improve the quality and coverage of toponyms of NLA Map, and can determine the benefits and limitation of OSM data in Aceh Province related to Rehabilitation Action for Post Tsunami Disaster. NLA as one of the ministries which has responsibility for developing Aceh Province after Tsunami, is in charge for mapping Base Map with large scale area, cadastral mapping and land registration. The main task of NLA is to provide Base Maps which will be used for cadastral map in every municipality. One of part in producing Base Map is toponyms survey. It is defined as visiting a geographical objects or features in real world to collect their geographical names. In mapping, it tends for giving textual information about object on the map like roads, villages, rivers, buildings, and mountains. Furthermore, with this identification of

geographic name, a map can be a tool for communicating people in the world and for supporting disaster management.

Currently, NLA is been seeking the advantage to exploit Open Source Data to reduce cost, to improve and to enrich existing mapping products. Openstreetmap (OSM) as the leading global Open Source Map Database would be an alternative to obtain information from many contributors. It is derived from various skill, knowledge and background. The amount of volunteers and contributors of OSM is very vast and quick since the ease of internet era. They map area where they are living, like roads, villages, buildings and rivers by using GPS or digitizing based on map given from volunteers. The contributors did not only give information about the positions, but also their attributes were undertaken in OSM. For example, a contributor track and record a road using GPS then he or she inputs the geographic name of that road. Thus, the information of this "crowdsourcing" map is an interesting to be used by NLA which is looking to enrich and to accelerate mapping NLA Base Map.

2. SOME THEORIES ABOUT TOPONYMS AND OPENSTREETMAPS

2.1 Basic theory of Mapping and Toponyms

A map, either raster or vector, contains information about the presence of real world model. Information about these models are represented in spatial and textual data. There are two ways to obtain spatial and textual data. First, by generating data through the other sources (e.g. remote sensing), the roads shape, for instance, can be obtained. In addition, taking textual information from imagery data also can be acquired such as object classification by analyzing images. Second, collect data by visiting geographical objects. For example, surveyors went to geographical objects (e.g. roads) then they measured using terrestrial measurements to acquire roads shape, length and width. Also, they can take specific textual information about the road, meaning their road names and road classification (specific textual data) by taking information from local knowledge.

Toponym means the place name or geographical name, while the study of toponyms, their origins, meanings, use, and typology is called toponymy. It is derived from Greek words *tópos* (τόπος) ("place") and *ónoma* (ὄνομα) ("name") (Wikipedia 2015). Basically the geographic name appears when the first humans to inhabit a territory and started to give the name of the object. Thus, there is a long history of human settlement in that name (Rais 2005).

2.2 The Standardization of Toponym in Indonesia

Indonesia realized about how important toponym, in 1987 the government of Indonesia reported to the United Nation (UN) conference on the standardization of geographical names about the amount of 13,677 islands transformed into 17,508 islands. However, what the UN want was about the process standardization of geographical names instead of the number of islands. This case impacted to toponym in Indonesia and other countries. Later on, UN included this case to UN Group of Experts on Geographical Names and Principles. Indonesia now responded with Presidential Decree No. 112 on 2006 about the National Standardization Team Name Topographic.

Toponym is consisted by two elements of names which are *generic name* and *specific name* (Rais 2005). The generic name illustrates the general configuration name an element of name in Indonesian language or local language, while specific name describes the place of name itself.

These are sample of generic names : Desa/Kampung (English : Village), Local language: "Lembur" in Sundanese and "Huta" in Bataknese ; Gunung (English : Mountain), Local language: "Dolok" in Batak ; Jalan (English : Road), Local Language : "Leumpang" in Sundanese, "Dalan" in Bataknese. The specific names are obtained from Place (e.g *Jakarta*, capital of Indonesia), People's Name (*Soekarno*, Former President of Indonesia), Animals (*Kucing*, English: Cat) or Common name that is given by communities. These two elements form a toponym each geographical objects that will be put on the gazetteer (a geographical index or dictionary), furthermore they would be seen on the maps as textual information of objects.

This project will not focus on the whole of type toponym objects but it will specify to road toponyms. These are the summaries of guidance for naming roads in Indonesia.

- The name of roads which is formed by People's Name, the configuration are written same with its full name instead of nickname and cannot be shortened, then followed by its title if the person had it. For example "Jalan Dr. Sutomo", "Jalan Ir. H. Juanda", "Jalan Jendral Gatot Subroto".
- The name of roads which is formed by Place's Name, Animal's Name, Plant's Name, Fruit's Name, the configuration are written as two words. For example "Jalan Blora", "Jalan Ciputat Raya", "Jalan Sumatera", "Jalan Mangga", "Jalan Ikan".
- The name of roads which is formed by generic name, then followed by its specific name, the configuration are written as one word. For instance, "Jalan Bukitsentul". Bukit is generic name for Hill in English, and Sentul is specific name from local people. Thus, it is written as one word.

2.3 Opestreetmap

Neis & Zielstra (2014) said that "OSM is a collaborative project to create a free editable map of the world. The goal of OSM is to build a free geographical database in the world and it can be accessed by every people by using the internet. There are two approaches that contributors can do in order to gather data in OSM. The first way is by recording data using GPS receiver, then the volunteers edit their data using free editors. Also, users can provide additional information about their data by adding attributes and other textual information (e.g. road names). Finally, the data that they collected store in OSM database (Neis and Zipf 2012). Another way is by digitizing aerial images data (e.g. remote sensed data or aerial photogrammetry) that given by volunteers to OSM database. So, the contributors can produce spatial information (e.g. footpath and building). Moreover, if the users have knowledge about the textual information such as name and feature class, they can add their attributes (Wikipedia 2014).

2.3.1 OSM in Indonesia

In Indonesia, there are applications in using OSM, such as Community Mapping for Exposure in Indonesia. The aim of this program is to minimize vulnerability to natural disasters. The students were involved to collect spatial and textual data both in urban and rural areas, by using GPS and digitizing images that given from volunteers. They collected previously unavailable data about buildings and their structure, and use appropriate models to calculate damage model in case of physical disaster. Finally they traced their results in the OSM platform. Thus, the thematic maps are created to show potential damage in case physical disasters. (Haklay, M et al., 2014). Another inovation is Flood preparedness through OSM in Jakarta. The Local Government represented by Jakarta's disaster management agency (BPBD DKI Jakarta) try to build geographical data to face the seasonal flood (Haklay, M et al., 2014). The students from University of Indonesia, Humanitarian Openstreetmap of Indonesia, AIFDR and UNOCHA helped the organization to map the potential area which is always impacted by the flood. The table below show the sub-category of project. The college students mapped the critical infrastructures by asking the heads of villages. Then, an InaSAFE open source software used for modelling their impact as part of a contingency planning process. Finally, based on that model, Local Government could create maps to report flood conditions and also every village has poster maps to plan logistics when responding to flooding.

It can be obviously seen, that Indonesian community had been involved in OSM. Moreover, the government also supports and endorses the use of OSM, particularly for hazard mitigation. The attribute names that are given by community, which were always updated and giving local knowledge, could help the government in giving toponyms for their authoritative map.

2.3.2 OSM Tags

The information that given by communities is provided by three basic components of data primitives, also called elements. They are consisted of nodes, ways and relations (wiki.openstreetmap.org 2014). Node is a specific point on the map which defined the attributes latitude and longitude. It is used to define stand-alone feature. Second, an ordered of nodes which interconnected for describing linear feature named Way. It can represent the street, foot path and boundaries. The last element is Relation which defined as a grouping of elements (nodes, ways and relations).

The three elements of data primitives has a tag which orders to describe the meaning of particular elements. The tag tells map users what all the data primitives of real-world features are represented in OSM. A tag consists a pair of key and value. They are simple strings which contain any valid Unicode characters up to a 255 characters (Bennett 2010). The key details a broad class of features, while the value details the specific feature that was generally classified by the key (wiki.openstreetmap.org 2015). Key and value are equal in tag, it can be written as key=value for a fully qualified tag, or key=* for a key without no particular value. To enlighten, these are examples about tagging of key and value for a "way" element in OSM.

- The tag is "highway=residential". The key is highway while residential associated to value. It means that a "way" indicates a road along which people live.

- The tag is “name=City Road”. The key is name whilst City Road related to value. It defines a specific name of a way element.

Based on those examples, if we join them, “highway=residential”+ “name=City Road”, the information that we can get from the “way” element is “road in residential area which has name City Road”. Sometimes, a key has not been given a value, for instance an element of way given tag as “barrier=*”, which means closed ways are used to define barriers, such as hedges and walls, that go completely round a property (wiki.openstreetmap.org 2015).

The OSM tag name:value is related to toponym, since it gives names to geographical objects on a map. For instance, a contributor describes a “way“ as name:City Road. The tagging expresses a meaning of features (semantic) which is derived from contributor’s knowledge. So, when OSM contributors tag a “way”, both of key and value, then map users can obtain a textual information of features based on contributors’ perspective. Then, the question comes to seek how qualified that tag can be utilized for toponyms in authoritative map which has standardization in naming geographical objects.

2.4 Data Quality for Geographical Information

The terms of data quality concern to words “quality” and “data quality”. Jakobsson and Tsoulos (2007) said that “quality is defined as fitness for use, including both quality of design, conformance to the design (production oriented quality), customer satisfaction and the fulfillment of the needs of society or environment”. While ISO/CD 19157 Geographic Information – Data Quality (2010) defines as “degree to which a set of inherent characteristics fulfils requirements”.

The geographical data quality is distinguished by several major focus elements, which are positional/spatial accuracy, attribute/thematic accuracy, temporal accuracy, completeness, logical consistency, lineage, semantic accuracy and usage. Since this project is related to geographical name places, the element of data quality that should be concerned are completeness and attribute/thematic accuracy.

According to Harding (2006), “attribute accuracy is the accuracy to which attributes in the vector-data record information about real-world entities”. It includes attributes of feature and feature representation (feature classification, textual information and change history of attributes). However, Whilst ISO/CD 19157 (2010) used terms “Thematic accuracy”, which is defined as the accuracy of quantitative attributes and the correctness of non-quantitative attributes and of the classifications of features and their relationships. While completnes, based on ISO/CD 19157 (2010) defines as “the presence and absence of features, their attributes and relationships.” It is grouped by applicable data quality elements from the following list commission (excess data present in a dataset) and omission (data absent from a dataset).

The absence of data and the presence of excess geographical data take effect when it is comparable to another dataset. For instance, if a spatial feature (e.g. road feature) does not exist in OSM datasets, then we cannot utilize them for taking their textual information. In addition, the quantitative of textual information of their road feature also major element to be used for comparing them with other datasets. Lastly, the attribute of spatial features is also the major element to be

evaluated with other datasets. Their quality of OSM tags is a question in order to use them for toponyms.

The OSM geographical data is acquired from random contributors, either by professional cartographers and non-professional cartographers, the data input and the process could make the results are varied. Meaning, the contributors have an important part to bring quality of OSM. Contributors use intuitive approaches to create geographical data and also do not necessarily tie to accept standards or methodologies. For instance, in giving OSM tags, users labeled the geographical features by using their own knowledge. In addition, the standardization of toponyms do not apply when tagging OSM geographical objects. So, the data gathered from various volunteers may be not uniform.

3. DATA ANALYSIS AND RESULTS

3.1 Obtaining Datasets

The OSM datasets was taken from Geofabrik websites, was downloaded on 20 May 2015, whilst NLA Base Map was obtained from General Directorate of Survey and Mapping, National Land Agency of Indonesia (NLA). The roads feature are picked to investigate their toponym attributes and ESRI shapefiles was used on this project to simplify the analysis on the GIS software. Geofabrik shapefiles always carry the standard names, the one that given in OSM's "name" tag (Frederik Ramm 2014).

OSM shapefiles format are OSM data format which converted to vector as points, polylines and polygons. In this case, road feature which represents as "Highway" in OSM formats are converted to polyline. In addition, their tags are also transformed as attributes table related to road segments. For instance, a Highway have two tags (key:value), name:Jalan Hasan and name:residential. These mean Highway were converted as polyline that have attributes Jalan Hasan as its road name, and residential as its road classification. Thus, the attributes of "name"s in OSM shapefile would represent the toponym of roads segments.

3.2 Preparing Datasets

The preparing datasets are aimed to match datasets between NLA Map and OSM, and to edit the datasets in a way that both of them contain similar information for facilitating the analysis procedure. The datasets preparation have important role in this project, it makes both datasets valid, uniform the datasets and reducing the vagueness.

3.3 Establishing Object Correspondences

The challenge in comparing attributes of two datasets is to find correspondence in both of datasets. It is recognizing features represent same object in the real world (Anand et al. 2010). In order to correspond both datasets, segmenting technique was performed. The main idea of this technique is to join the attributes tables between NLA roads and OSM roads on polygons segments. So, each

polygon of road segments had two attributes, one from NLA roads and one from OSM roads. The output of attributes table on these polygons would be examined for comparing road names between two maps.

Id	URUT	Nama_Resmi	name	anoKd
0	5237	Jalan Pocut Baren	Jalan Pocut Baren	64322
0	1224	Jalan Peukan Bada		64322
0	1224	Jalan Peukan Bada	Ajun Jeumpit	64322
0	504	Jalan Pendidikan		64322
0	834	Jalan Pemancar		64322
0	488	Jalan Pattimura		64322
0	488	Jalan Pattimura		64322
0	448	Jalan Pasar induk Lambaro		64322
0	766	Jalan P.Nyak Makam	Jalan Panglima Nyaj Makam	64322
0	1712	Jalan Montasik		64322
0	137	Jalan Moh Hasan	Jalan Teuku Muhammad Hasan	64322
0	5223	Jalan Malikul Saleh	Jalan Wedana	64322
0	5223	Jalan Malikul Saleh	Jalan Sultan Malikul Saleh	64322
0	1962	Jalan Maimun Saleh		64322
0	5222	Jalan M.Thaher	Jalan Ir. H. Mohammad Thaher	64322
0	5233	Jalan Lhoknga	Jalan Banda Aceh Meulaboh	64322
0	4551	Jalan Lhok Buya		64322
0	791	Jalan Layang P.Nyak Makam	Jalan Prof. Ali Hasjmy	64322
0	1964	Jalan Lampuuk		64322
0	1926	Jalan LamLong		64322
0	442	Jalan Lambaro	Jalan Rel Kereta Api	64322

16 (1 out of 216 Selected)

Figure 2. Examples of Results of Establishing Correspondences

3.4 Road Name Existence

After corresponding datasets, the name of each road segment between two datasets can be compared. The completeness that will be assessed in this project is about existence and nonexistence of road names in both datasets which are resulted from corresponding datasets process.

To be known, not all the NLA roads segments have attributes, this is because there were no reliable information at the time of the toponym surveys. It is also occurred in OSM roads segments, the reason is that not all contributors give tags (name:value) on their OSM data primitives. Thus, there are four potential result of the existence and non-existence condition of road names. The term "Yes" used for the road that has name, while "No" represents that the road does not has name. 1) NLA Yes-OSM Yes; 2) NLA Yes-OSM No; 3) NLA No-OSM Yes; 4) NLA No-OSM No. The formula used to calculate the ratio of road names existence is:

Equation 1 . Ratio of road names existence

$$Ratio = 100\,x\,\frac{Condition}{Total\ Segment}$$

This formula was performed for both study area, hence Banda Aceh and Aceh Besar have their ratio of road names existence condition. Here are results of road name existence.

3.5 String Comparison

Road names attribute contain lexical information and it is normally represented as strings. The name of each roads have meaning in a natural language, either national language or local language. In order to compare the road names between NLA map and OSM, the string comparison technique is common way to use (Du et al. 2012).

Before comparing the strings of two datasets, it is needed to apply normalization. The purpose of normalization is to remove noise in data formats. (Euzenat & Shvaiko (2007) described the following normalization practices and useful for matching NLA and OSM datasets. They are case normalization, blank normalization, link stripping and punctuation elimination.

According to Pandiselvam et al (2014), string matching is a technique to find pattern from the specified input string. By using string matching algorithm, the similarity between pattern and specified string can be measured. One of the string comparison techniques is Levenshtein Distance (LD). This technique used in order to measure quantitative of attribute accuracy and also implements Levenshtein Distance to quantify the types of changes the attributes of spatial object in OSM.

3.6 Utilizing OSM Tags for NLA Maps

The method on this project was to create a scenario in order to utilize road names from OSM to NLA Map. As described on previous section, there are four results of existence and nonexistence road names after segmenting process between two datasets. On this section, each condition will be given a scenario to optimize them for NLA toponyms.

- *NLA Yes-OSM Yes,* if the road name on NLA Map and OSM were exist in one road segment, the string comparison would be applied.
- *NLA Yes-OSM No,* if an NLA road has attributes of road name while an OSM does not has, there is no action to be implemented on this condition.
- *NLA No-OSM Yes*, if there are no road names on NLA roads segments, while OSM have, the action is to take road name from OSM to fulfil attribute on NLA roads.
- *NLA No-OSM No,* if both geometry do not have road names, there is no action in this condition.

In addition, NLA Yes-OSM No condition, result of Levenshtein Distance (LD) would be varied from 0 to higher. In Levenshtein Distance, the zero (0) distance value means that strings between two datasets are identical, otherwise they are not same. So, we split it into two sub conditions, LD = 0 and LD > 0. Then, further scenario was created for these two sub conditions. At LD = 0, there is no action given on this sub condition. So, the road names from NLA would be remained as toponyms. However, when LD is larger than 0, semantic analysis would be implemented. It is purposed to seek whether there is a similarity meaning or no in both strings. Usually, the same meaning happened when Target string contents a part of Source string or vice versa, whereas the

different meaning generally are completely different. In order to simplify those terms, the same meaning and different meaning will be denoted as SM and DM respectively. For more details, the scenarios are described as on this table below.

Table 1 Scenario for Utilizing OSM Tags for NLA Maps

No	Existence Conditions		Rules	Actions	Method	Result
	NLA	OSM				
1	Yes	Yes	Strings Comparison (LD)			
			a)If LD = 0	No Actions		Road's Polyline taken from NLA Map, including the road names
			b) If LD > 0	Semantic Analysis	Manual Examinations	Road's Polyline taken from NLA Map, while road names taken from semantic analysis result
			i) SM			
				- Categorizing Misspelling		
				- Categorizing Name Completeness		
				- Categorizing Abbreviation		
				- Categorizing Title		
			ii) DM	Semantic Analysis	Manual Examinations	Road's Polyline taken from NLA Map, but one segment road has two names from NLA and OSM
				- Categorizing Local name		
2	Yes	No	-	No Actions		Road's Polyline taken from NLA Map, including the road names
3	No	Yes	-	Enriching NLA Map	Visual Checking	Road's Polyline taken from NLA Map, while road names taken from OSM.
4	No	No	-	No Actions	Visual Checking	Displaying roads geometry from NLA Map

4. RESULTS

In total, the number of road polygons segments that had been corresponded are 140 road segments in Banda Aceh, while Aceh Besar has 91 road segments. It is reasonable because Banda Aceh is an urban area that was rebuilt after the tsunami and it is also the capital of Aceh Province, whereas Aceh Besar is a rural area with population less than Banda Aceh. So, the government might be focus on constructing the urban area after tsunami disaster.

Table 2 Ratio of Road Names Existence

Existence Condition	Banda Aceh		Aceh Besar	
	Numbers	Ratio	Numbers	Ratio
NLA Yes-OSM Yes	38	27.14%	20	21.98%
NLA Yes-OSM No	13	9.29%	11	12.09%
NLA No-OSM Yes	31	22.14%	6	6.59%
NLA No-OSM No	58	41.43%	54	59.34%
Total Road Segments	**140**	**100.00%**	**91**	**100.00%**

The string comparison is only applied on NLA Yes-OSM Yes existence condition. At this case, there are 38 pair of strings in Banda Aceh and 20 pair of strings in Aceh Besar that would be inspected. The statistics analysis gives information that the average of LD values in Banda Aceh is slightly better rather than Aceh Besar with 10.24 and 12.15 respectively. This fact show that the standardization of naming roads is complex.

Table 3 Statistics of Levenistein Distance Values and Similarity Values

Statistics	Similarity Value		Levenshtein Distance Value	
	Banda Aceh	Aceh Besar	Banda Aceh	Aceh Besar
Maximum	100.00%	95.00%	24	21
Minimum	25.00%	29.41%	0	1
Mean	58.80%	48.49%	10.24	12.25
Standard Deviation	0.26	0.17	6.8	4.31

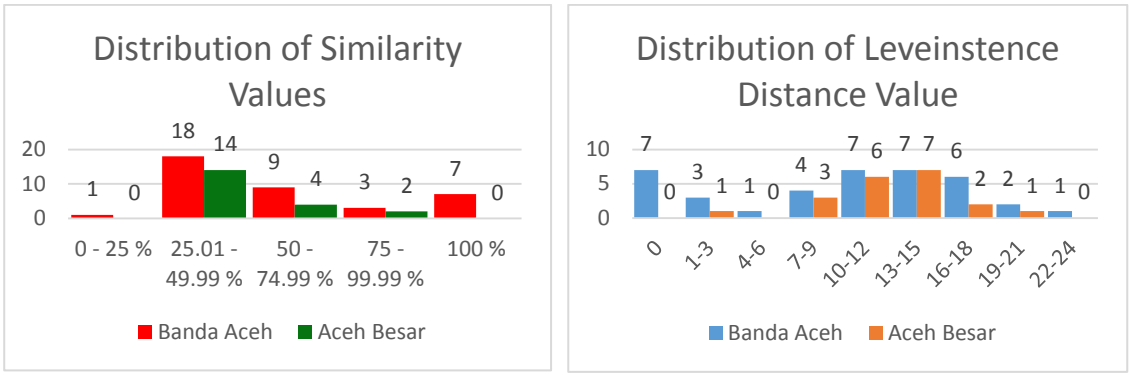
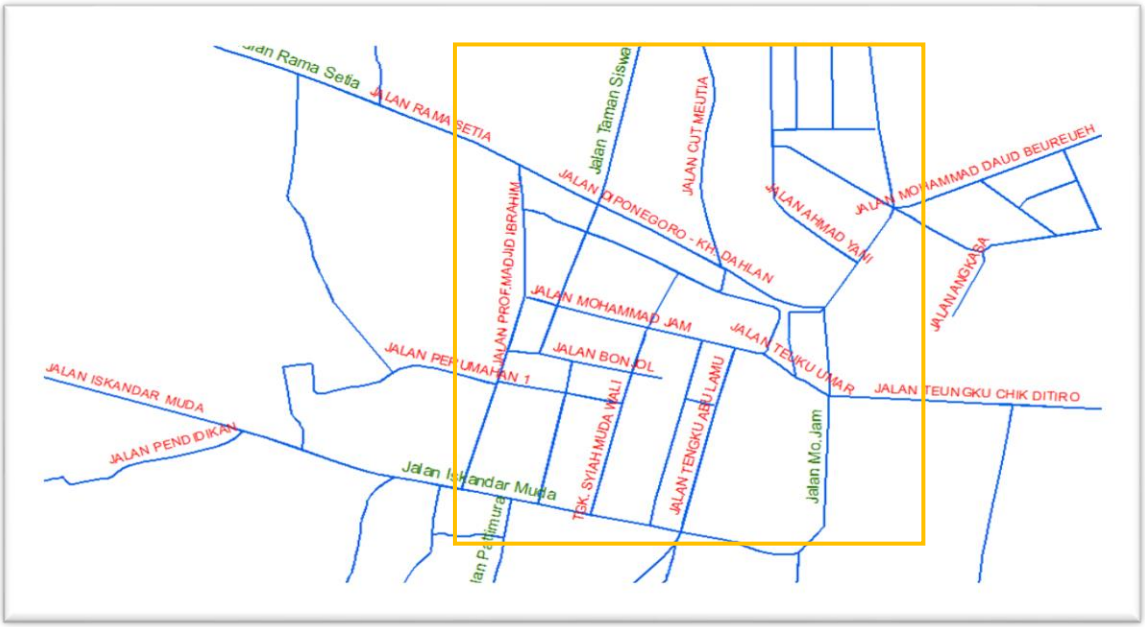


Figure 3 Distribution of Similarity Values and Leveinstence Distance Value

The NLA No-OSM Yes in Banda Aceh has around 22% ratio with 31 roads segments, whilst Aceh Besar has only 6% ratio with 6 road segments. The advantage of this condition is that OSM road names can be utilized to fulfill NLA road segments (conflation). However, the criteria of naming roads should be thought carefully with NLA's procedures such as how to write abbreviation, misspelling and name's title correctly. The visualization about conflating road names from OSM to NLA roads can be seen on figure 3. The red road names represents suggestion road names, while green words reflects NLA original name. This map shows that



OSM can fill the road segments that do not have road names (yellow rectangle).

Figure 4 Examples of Conflating Road Names for NLA Map in Banda Aceh

In Banda Aceh, there are 12 SM sub conditions while there are 6 SM sub conditions in Aceh Besar. From the SM results, 10 comparisons decided to use OSM road names due to several mistakes in naming the roads in NLA toponym. These mistakes are may be caused by mistyping by surveyors when they input the road names to GPS, lack of information from field area, and habitual to shortened people names. So, based on these examples, OSM tags can amend and complement NLA toponyms.

Table 4 Same Meaning Semantic analysis in Banda Aceh

No	FID	NLA NAME	OSM NAME	Categorize of Action	Explanations	String that will be used in NLA roads
1	14	JALAN CUT NYAK DIEN	JALAN CUT NYAK DHIEN	Misspelling/Mistyping	DIEN instead DHIEN; means Candle	NLA
2	4	JALAN P.NYAK MAKAM	JALAN PANGLIMA NYAJ MAKAM	Misspelling/Mistyping	NYAK instead NYAJ; means "Affection or Mom"	NLA
3	15	JALAN PROF. MADJID IBRAHIM	JALAN PROFESOR ABDUL MAJID IBRAHIM	Name Completeness;	added ABDUL;	NLA
				Abbreviation	PROF instead PROFFESOR	
4	7	JALAN TEUN	JALAN T. ISKANDAR	Abbreviation	TEUNGKU instead T. In addition, this is wrong shortened. TEUNGKU should be shortened	NLA

No	FID	NLA NAME	OSM NAME	Categorize of Action	Explanations	String that will be used in NLA roads
		GKU ISKANDAR			as “Tg” means Religious Leader	
5	28	JALAN M.THAHER	JALAN IR. H. MOHAMMAD THAHER	Abbreviation	MOHAMMAD instead M;	OSM
				Title	added IR and H	
6	10	JALAN DAUD BEUREUH(SIMPANG LIMA)	JALAN MOHAMMAD DAUD BEUREUEH	Name Completeness;	added MOHAMMAD	OSM
					delete SIMPANG LIMA, it is a POI name of road junction	
7	5	JALAN SOETA	JALAN SOEKARNO-HATTA	Abbreviation	SOEKARNO-HATTA instead SOETA, no abbreviation for road names from People Names	OSM
8	44	JALAN SUDIRMAN KOTA BANDA ACEH	JALAN SUDIRMAN	Name Completeness;	delete KOTA BANDA ACEH	OSM
9	8	JALAN MALIKUL SALEH	JALAN SULTAN MALIKUL SALEH	Name Completeness;	added SULTAN; means King in Islamic Emperor	OSM
10	1	JALAN MOH HASAN	JALAN TEUKU MUHAMMAD HASAN	Name Completeness;	added TEUKU; means peerage	OSM
				Abbreviation	MOHAMMAD instead MOH no abbreviation for road names from	

No	FID	NLA NAME	OSM NAME	Categorize of Action	Explanations	String that will be used in NLA roads
					People Names	
11	30	JALAN TENGKU CIK DI TIRO	JALAN TEUNGKU CHIK DITIRO	Misspelling/Mistyping	CHIK instead CIK; means Grandpa and TEUNGKU instead TEGUK; means Religious Leader	OSM
12	32	JALAN TEGUK IMUM	JALAN TEUNGKU IMUM LUENG BATA	Misspelling/Mistyping;	TEUNGKU instead TEGUK; means Religious Leader	OSM
				Name Completeness;	added LUENG BATA	

Table 5 Same Meaning Semantic analysis in Aceh Besar

No	FID	NLA NAME	OSM NAME	Categorize of Action	Explanations	String that will be used in NLA Map
1	23	JALAN M.THAHER	JALAN IR. H. MOHAMMAD THAHER	Abbreviation	MOHAMMAD instead M;	OSM
				Title	added IR and H	
2	29	JALAN SOETA	JALAN SOEKARNO-HATTA	Abbreviation	SOEKARNO-HATTA instead SOETA, no abbreviation for road names from People Names	OSM
3	28	JALAN MOH HASAN	JALAN TEUKU MUHAMMAD HASAN	Name Completeness;	added TEUKU;	OSM
				Abbreviation	MOHAMMAD instead MOH; no abbreviation for road names from People Names	
4	14	JALAN CUT NYAK DIEN	JALAN CUT NYAK DHIEN	Misspelling/Mistyping	DIEN instead DHIEN; means Candle	NLA

5	24	JALAN MALIKUL SALEH	JALAN SULTAN MALIKUL SALEH	Name Completeness;	added SULTAN; means King in Islamic Emperor	OSM
6	26	JALAN RAYA BANDA ACEH - MEULABOH	JALAN BANDA ACEH MEULABOH	Name Completeness;	added RAYA	NLA

DM sub condition semantic analysis results present 19 pairs of string in Banda Aceh and 14 pairs of strings in Aceh Besar. These are the highest ratio in utilizing OSM for NLA toponym. What we can consider why this is happen are the assumption of local road names. It means that local people in that place sometimes give a nickname for geographical place. Also, another hypothesis is there could be wrong in naming roads either by contributors of OSM or NLA surveyor. Nevertheless, it needs further research to prove these hypothesis. The suggestion to check and to compare with other resources datasets (such as Google maps or other authoritative maps from other agency like Geospatial Information Agency) should be taken.

Table 6 Different Meaning semantic analysis in Banda Aceh

No	FID	NLA NAME	OSM NAME	Categorize of Action
1	9	JALAN GAMPONG LAMJRUM	JALAN KRUENG RAYA-LAMTEUBA	Local name
2	19	JALAN GAMPONG NUSA	JALAN BANDA ACEH MEULABOH	Local name
3	1	JALAN GAMPONG WEU	JALAN JANTHO-TANGSE	Local name
4	17	JALAN KEUTAPANG	JALAN BANDA ACEH MEULABOH	Local name
5	6	JALAN KRUENG RABA	JALAN GOLF	Local name
6	25	JALAN LAMBARO	JALAN REL KERETA API	Local name
7	22	JALAN LHOKNGA	JALAN BANDA ACEH MEULABOH	Local name
8	18	JALAN PEUKAN BADA	AJUN JEUMPIT	Local name
9	0	JALAN PRADA UTAMA	JALAN TEUKU NYAK ARIEF	Local name
10	4	JALAN RAYA BANDA ACEH - MEULABOH	JALAN CUT NYAK DIEN	Local name

11	20	JALAN RAYA JANTHO	JALAN RAYA MEDAN-BANDA ACEH	Local name
12	30	JALAN SULTAN ISKANDAR MUDA	JALAN BANDARA	Local name
13	12	JALAN TEGUK IMUM	JALAN RAYA LAMBARO	Local name
14	15	JALAN RAYA BANDA ACEH - SIGLI	JALAN RAYA MEDAN-BANDA ACEH	Local name

Table 7 Different Meaning semantic analysis in Aceh Besar

No	FID	NLA NAME	OSM NAME	Categorize of Action
1	27	JALAN ANGSA	JALAN NYAK ADAM KAMIL I	Local name
2	11	JALAN H.SALEH	JALAN NYAK ADAM KAMIL III	Local name
3	20	JALAN ISKANDAR MUDA	JALAN PELABUHAN LAMA ULEE LHEUE	Local name
4	29	JALAN JAMBOTAPE	JALAN SYIAH KUALA	Local name
5	36	JALAN LAYANG P.NYAK MAKAM	JALAN PROF. ALI HASJMY	Local name
6	22	JALAN LHOKNGA	JALAN BANDA ACEH MEULABOH	Local name
7	18	JALAN MALIKUL SALEH	JALAN WEDANA	Local name
8	38	JALAN PRADA UTAMA	JALAN TEUKU NYAK ARIEF	Local name
9	19	JALAN PROF.MADJID IBRAHIM	JALAN ISKANDAR MUDA	Local name
10	47	JALAN TAMAN MAKAM PAHLAWAN	JALAN NYAK ADAM KAMIL III	Local name
11	45	JALAN TAMAN MAKAM PAHLAWAN	JALAN RESIDEN DANUBROTO	Local name
12	35	JALAN TAMAN MAKAM PAHLAWAN	JALAN SULTAN MALIKUSSALEH	Local name
13	41	JALAN TAMAN SISWA	JALAN KH AHMAD DAHLAN	Local name
14	42	JALAN TEGUK IMUM	JALAN RAYA LAMBARO	Local name
15	16	JALAN TENGKU DIL HONG	JALAN TENGKU DILHONG 2	Local name
16	24	JALAN TENGKU DIL HONG	JALAN WEDANA	Local name
17	6	JALAN TENGKU DIL HONG	JALAN KEMUNING	Local name
18	23	JALAN TEUNGKU NYAK ARIEF	JALAN MOHAMMAD DAUD BEUREUEH	Local name
19	43	JALAN ULEE LHEUE	JALAN BANDA ACEH MEULABOH	Local name

5. CONCLUSION

One of the duties from NLA is to create good land administration system. Like any other The Impacted Areas, Aceh Province needs good land administration system to reduce the social and

ownership conflicts. This paper intends to improve the quality and coverage of toponyms of NLA Map, and can determine the benefits and limitation of OSM data in Aceh Province related to Rehabilitation Action for Post Tsunami Disaster. NLA as one of the ministries which has responsibility for developing Aceh Province after Tsunami, is in charge for mapping Base Map with large scale area, cadastral mapping and land registration and one of part in producing Base Map is toponyms survey. By taking OSM tags post-tsunami reconstructions, NLA can use them as initial information before doing toponyms survey. Also, OSM could give better understanding about local language toponyms in the surveyed area. So, NLA toponyms resulted from survey are expected to be accurate and it can accelerate the survey process.

The results shows that OSM tags in urban area are quantitatively better than rural area. It can be seen from quantitative comparisons that there are 4 results. First, the road names existence condition of NLA Yes-OSM Yes presented around 27% ratio and around 21% ratio in Banda Aceh and Aceh Besar respectively. Second, comparing road names existence condition of NLA No-OSM Yes in Banda Aceh and Aceh Besar with 22.14% ratio and 6.59% ratio. These two road names existence conditions are types which could optimize NLA toponyms due to their OSM tags existence. Third, there are 7 LD values which are same in both datasets in urban area, while there are no similar strings in rural area. Lastly, LD standard deviation of Banda Aceh is lower than Aceh Besar which has differences nearly 2. Third and fourth results have given a correlation that many OSM users are interested in mapping urban area, so that made the completeness of OSM in urban area is better than rural area. Hopefully, OSM also increases in rural area of Indonesia considering the government is still mapping whole country.

In utilizing toponym, the same meaning (SM) sub condition which was taken from string comparisons give some advantages for complementing and correcting NLA toponym. OSM tags showed that they can follow the road naming regulation of NLA rather than NLA toponyms itself. It can help to complement an incomplete road name, incomplete title or correcting misspelling that derived from mistakes during toponyms survey. In addition, the result of different meaning (DM) semantic analysis could give another perspective about the road names, but it needs to compare other authoritative datasets or commercial datasets like google maps. Also, NLA No-OSM Yes existence condition is a merit for enriching NLA roads which do not have attributes. Appearing mostly in Banda Aceh, this condition can be good for urban area particularly in city center. By considering these results, OSM tags in the study area can give benefits for NLA toponyms.

However, the completeness of road's OSM tags is the limitation for NLA toponyms, mainly in rural area. The road names existence conditions are from around 9% to 12% ratio for NLA Yes-OSM No, and from 41% to 60% ratio for NLA No-OSM No. These conditions make the comparison between them can be implemented, and they could also not fulfill NLA road segments which do not label with road names.

Finally, as OSM is the leading crowdsourced mapping in the world, government could involve the communities in taking spatial information and textual information. Thus, it is suggested to share the regulation of toponyms that the government made to OSM communities. By socializing their

regulation in naming geographical places, Indonesia can get more complete geographical data in whole country, particularly to speed up and to reduce toponyms survey. Also, this may help government to obtain uniformity in naming geographical places in Indonesia.

REFERENCES

Anand, Suchith et al. 2010. "When Worlds Collide: Combining Ordnance Survey and Open Street Map Data."

"Belajar Bahasa Aceh [Learning Aceh Language]." 2015. *Belajar Bahasa Aceh*. https://bahasaaceh.wordpress.com/2010/02/14/nama-asli-aceh/ (August 1, 2015).

Bennett, Jonathan. 2010. *OpenStreetMap*. 1st ed. ed. Mithun Sehgal. Birmingham: Packt Publishing.

Du, Heshan et al. 2012. "Geospatial Information Integration for Authoritative and Crowd Sourced Road Vector Data." *Transactions in GIS* 16(4): 455–76.

Euzenat, Jérôme, and Pavel Shvaiko. 2007. Booksgooglecom *Ontology Matching*. Berlin: Springer. http://book.ontologymatching.org/.

Frederik Ramm. 2014. "OpenStreetMap Data in Layered GIS Format." http://www.geofabrik.de/data/geofabrik-osm-gis-standard-0.6.pdf.

GFDRR. 2014. *Understanding Risk The Evolution of Disaster Risk Assessment*. Washington. https://www.gfdrr.org/sites/gfdrr.org/files/publication/_Understanding_Risk-Web_Version-rev_1.7.3.pdf.

Haklay, M., Antoniou, V., Basiouka, S., Soden, R., and Mooney, P. 2014. *Crowdsourced Geographic Information Use in Government Report to GFDRR*. London. http://www.preventionweb.net/english/professional/publications/v.php?id=38130.

Harding, Jenny. 2006. "Vector Data Quality: A Data Provider's Perspective." In *Fundamentals of Spatial Data Quality*, eds. Rodolphe Devillers and Robert Jeansoulin. London, UK: ISTE, 141–58.

InaSAFE. 2015. "About InaSAFE." http://inasafe.org/en/general/index.html (August 15, 2015).

"ISO/CD 19157 Geographic Information – Data Quality." 2010. : 155. http://www.isotc211.org/protdoc/211n2959/.

Jakobsson, A, and L Tsoulos. 2007. "The Role of Quality in Spatial Data Infrastructures." *Eurogeographics*: 10. http://hostmaster.icaci.org/files/documents/ICC_proceedings/ICC2007/documents/doc/THEME 7/oral 3-2/7.3-2.1 (НЕТ 1 ФОТО) THE ROLE OF QUALITY IN SPATIAL DATA INF.doc.

Neis, Pascal, and Alexander Zipf. 2012. "Analyzing the Contributor Activity of a Volunteered Geographic Information Project — The Case of OpenStreetMap." *ISPRS International Journal of Geo-Information* 1(2): 146–65.

Pandiselvam, P, T Marimuthu, and R Lawrance. 2014. "A Comparative Study on String Matching Algorithms of Biological Sequences." In *International Conference on Intelligent Computing*, New York: Cornell University Library, 1–5. http://arxiv.org/ftp/arxiv/papers/1401/1401.7416.pdf.

Perdana, a. P., E. Hendrayana, and W. E. Santoso. 2012. "The Important of Toponym in the Middle of Maps and Imagery for Disaster Management." *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives* 39(September): 523–27. http://www.scopus.com/inward/record.url?eid=2-s2.0-84924335929&partnerID=tZOtx3y1.

Rais, Jacub. 2005. “Arti Penting Penamaan Unsur Geografi Definisi , Kriteria Dan Peranan PBB Dalam Toponimi [The Importance of Naming Geographical Objects: Definition, Criteria and the Role of the UN in Toponymy].” : 1–25.
Statistics Indonesia. 2014. *Trends of Selected Socio-Economic Indicators of Indonesia*. 03230.1401 ed. Jakarta: BPS-Statistics Indonesia. http://www.bps.go.id/aboutus.php?booklet=1\n.
wiki.openstreetmap.org. 2014. “Elements.” http://wiki.openstreetmap.org/wiki/Elements (December 10, 2014).
———. 2015. “Tags.” http://wiki.openstreetmap.org/wiki/Tags (August 20, 2015).
Wikipedia. 2014. “OpenStreetMap.” https://en.wikipedia.org/wiki/OpenStreetMap.
———. 2015. “Toponymy.” https://en.wikipedia.org/wiki/Toponymy.

CONTACTS

1. Indra R. Hutabarat
   Institution : Inspectorate General of Ministry of Agrarian and Spatial Planning/National Land Agency
   Office Address : Jl. H. Agus Salim No. 58 Jakarta
   Position : Surveyor, GIS Analyst, Planner, Researcher
   Email : indra.rohara.hutabarat@gmail.com

2. Memby U. Pratama
   Institution : Inspectorate General of Ministry of Agrarian and Spatial Planning/National Land Agency
   Office Address : Jl. H. Agus Salim No. 58 Jakarta
   Position : Auditor, Land Management Consultant
   Email : membyup@2009.jukuin.keio.ac.jp