


# Semi-Automated Metadata Detection for Assessing the Credibility of Map Mashup






**Sr. Dr. Nurul Hawani Idris**  
**Universiti Teknologi Malaysia (UTM)**


Commission No 3






## Co-Authors

 <p><b>Prof. Mike J. Jackson, FRICS</b> Nottingham Geospatial Institute, University of Nottingham UK</p>	 <p><b>Sr. Mohamad Nor Said (Assoc. Prof.)</b> Geoinformation Department Universiti Teknologi Malaysia</p>
 <p><b>Zamri Ismail</b> Geoinformation Department Universiti Teknologi Malaysia</p>	 <p><b>Mohamad Ghazali Hashim</b> Geoinformation Department Universiti Teknologi Malaysia</p>
 <p><b>Dr. Mohamad Hafis Izran Ishak</b> Infocomm Research Alliance Universiti Teknologi Malaysia</p>	 <p><b>Fathin Nazeera Nazri</b> MSc. Student Universiti Teknologi Malaysia</p>




FIG

# Introduction





- Web 2.0 and the release of Google Map APIs (2005) have made big impact on the culture of mapping
- Neogeography -
  - Neogeography consists of a set of techniques and tools that fall outside the realm of traditional GIS (Turner, 2006)
- Volunteered Geographic Information (VGI)
  - The widespread engagement of large number of private citizen, often with little in the way of formal qualification. They are largely untrained and their actions are always almost voluntary and the results may or may not be accurate (Goodchild, 2007)



FIG


# Introduction





**Official Air Pollutant Index (API) of  
Department of Environment**  
Ministry of Natural Resources and environment

Utama (Main)
Arkib
Agensi Remote Sensing Malaysia (ARSM)
Hotspot: (FORFIS)
Maklumat Tikik Panas (NOAA18)
FAQ



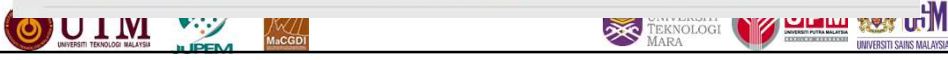
Map data ©2014 Google, MapIT Terms of Use

**04-04-2014 (Friday) - 06:00AM - 11:00AM**

> Bacaan Terkumpul Indeks Pencemaran Udara  
Accumulated Air Pollutant Index

> Maklumat Umum Indeks Pencemaran Udara  
General Info of Air Pollutant Index

> Guidelines for the Protection of Employees  
Against the Effects of Haze at Workplaces



# FIG Introduction

1 Malaysia Map  
A Geoinformation Viewer for Everyone

1malaysiamap.mygeoportal.gov.my

Places of Interest

Layer Visibility

- Educational Institution
- Accommodation
- Bank
- Retail and Services Outlet
- Recreational Place
- Tourism and Travel Destination
- Institutional Building
- Lot WPKL

Latitude: 3.154045 Longitude: 101.717639 Scale: 1:43118.98

UTM UNIVERSITI TEKNOLOGI MALAYSIA NUPEM MaCGDI UNIVERSITI TEKNOLOGI MARA UPM UNIVERSITI PUTRA MALAYSIA UAM UNIVERSITI SAHAB MALAYSIA

# FIG Introduction

## Timeline map: Tracking Malaysia Airlines Flight 370

4:24 p.m. CDT, March 24, 2014

March 08, 2014  
Plane crashes in the Indian Ocean

Satellite data shows the Indian Ocean is the plane's last position before it crashes.

Investigators will use the data to conclude that all lives were lost.

"This is a remote location, far from any possible landing sites," Malaysia's Prime Minister Najib Razak said. "It is therefore, with deep sadness and regret, that I

Plane crashes in the Indian Ocean

UTM UNIVERSITI TEKNOLOGI MALAYSIA NUPEM MaCGDI UNIVERSITI TEKNOLOGI MARA UPM UNIVERSITI PUTRA MALAYSIA UAM UNIVERSITI SAHAB MALAYSIA

**FIG** Introduction

mashable.com/2014/03/16/malaysia-airlines-runways/

**Mashable** 5.2k sources

**WNYC** Malaysia Airlines Flight 370: Runways in Range

On March 8, 2014, Malaysia Airlines Flight 370 disappeared from civilian radar. Amid **unusual** (some say **unlikely**) theories that the flight may have landed, can we figure out how many runways might be available?

Data from **X-Plane** provides coordinates for runways around the world. A Boeing 777 pilot is **quoted in Slate** as estimating a runway length requirement of 5,000 feet. A recent **Wall Street Journal** article quoted sources stating the flight could have continued for 2,200 nautical miles from its last known position.

The WNYC Data News team found 634 runways that meet these criteria, spread across 26 different countries, including such far-flung places as:

Gan Airport (Maldives), Dalanzadgad Airport (Mongolia), Yap Airport (Micronesia), Miyazaki Airport (Japan)

Logos: UTM, NPEM, MaCGDI, UNIVERSITI TEKNOLOGI MARA, UPM, UNIVERSITI SAINS MALAYSIA

**FIG**

www.malaysiacrime.com/recent/commented/

**Africans are selling drugs out of their homes in Malaysia**  
9 months, 3 weeks ago at **Puchong** reported by Markere // 2 comments

**Sad Sad Sad ,I will kill them all**  
1 year, 2 months ago at **everywhere** reported by never sleep at nite // 1 comment

**African criminals seek their dreams in Malaysia**  
1 year, 9 months ago at **Puchong** reported by Tamhi // 190 comments

**Car thief drives off with baby on board**  
9 months, 3 weeks ago at **Jalan Kuching** reported by dan // 1 comment

**Beware! Most Africans in Malaysia possess drugs.**  
10 months, 1 week ago at **Selangor** reported by Jackie // 4 comments

**fake money**  
1 year, 5 months ago at **kualumpurbukitbintan** reported by myanmarman // 23 comments

**Malaysia crime racist**  
1 year, 6 months ago at **Kelana jaya** reported by Malaysian // 13 comments

**FIG**

## Research Background

**BEFORE**

- Professional developers
- Data supplied by governments / commercial organisations
- Standardised procedure – capture, quality control, dissemination

**AFTER**

- Non-professionals / citizens
- Personal/localised data/various sources data
- Low cost tools
- Low programming skills
- Low experiences/prior knowledge – mapping design/data capture

Issues – truth, quality, accuracy !

Produce own maps

Logos: UTM, NIFEM, MaCGDI, Universiti Teknologi MARA, UPM, IISM

**FIG**


## Previous Studies

- Simply tagging metadata on map mashup –low influence on users' perceived credibility  
Nurul Hawani Idris, Mike J.Jackson & Robert J. Abrahart (2011), [Map Mashups: What looks good must be good?](#), GIS Research UK Conference (GISRUK), Portsmouth UK, 27-29 April 2011
- High influence of Credibility Labelling on users perceived credibility– Colour Coded Traffic Light labelling  
Nurul Hawani Idris, Mike J.Jackson & Robert J. Abrahart (2011), [Credibility Labelling in Map Mashup Design: Influence on Users' Judgement](#), EuroSDR/ISPRS Workshop on Web Cartography, Lund Sweden, 5th-6th May 2011
- [Colour Coded Traffic Light Labelling: An Approach to Assist Users in Judging Data Credibility in Map Mashup Applications](#). Nurul Hawani Idris, Mike J. Jackson, and Robert J. Abrahart (2011). In Cidalia C. Fonte, Luisa Goncalves & Gil Goncalves (Eds), Proceedings of the 7th International Symposium on Spatial Data Quality (pp. 201-206). Coimbra Portugal: INESC Coimbra


Logos: UTM, NIFEM, MaCGDI, Universiti Teknologi MARA, UPM, IISM

FIG


## Previous studies




- Successful implementation of ‘seal of approvals’ in other domains









This site complies with the HONcode standard for trustworthy health information: [verify here.](#)  
Search only trustworthy HONcode health websites:






**Multiple traffic light labelling**  
We've developed our traffic light labelling - an at-a-glance look out for it on front of our packs.  
[More on our multiple traffic light labelling](#)
- Visual quality indicators on GIS maps (Devillers et al. (2002)) - a response to the difficulty of communicating metadata to professional and lay users. Devillers et al. (2007) - a practical model to implement a quality rating system in GIS for use by experts to give advice about the quality of a dataset.











FIG









## Problem Statements



- Issue 1 : the insufficient of simply tagging metadata on map mash up
- Issue 2: A crucial need of a gatekeeper control scheme on online maps, particularly on map mashups













## Purpose of study

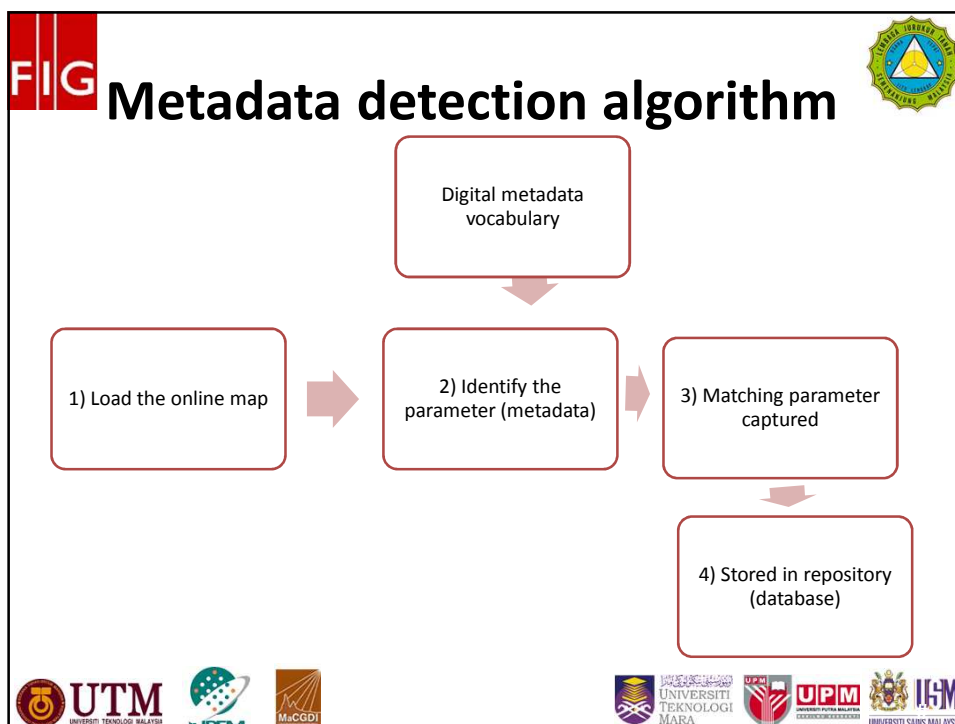
- To demonstrate the ability of web crawler (scraper) to detect metadata criteria towards the implementation of credibility labelling on map mashup

## Related studies

Metadata component

1. Rule based line classification – a set of rules to detect metadata concept (Wang and Richard 2007)
2. Supervised machine learning – similar technique to automate recognition and filtering spam emails. A good corpus of training data from a huge collection of metadata used in mapping application. Patterns of texts to specific categories of metadata used to detect and rate the indicators (Gaudinat et al., 2007)

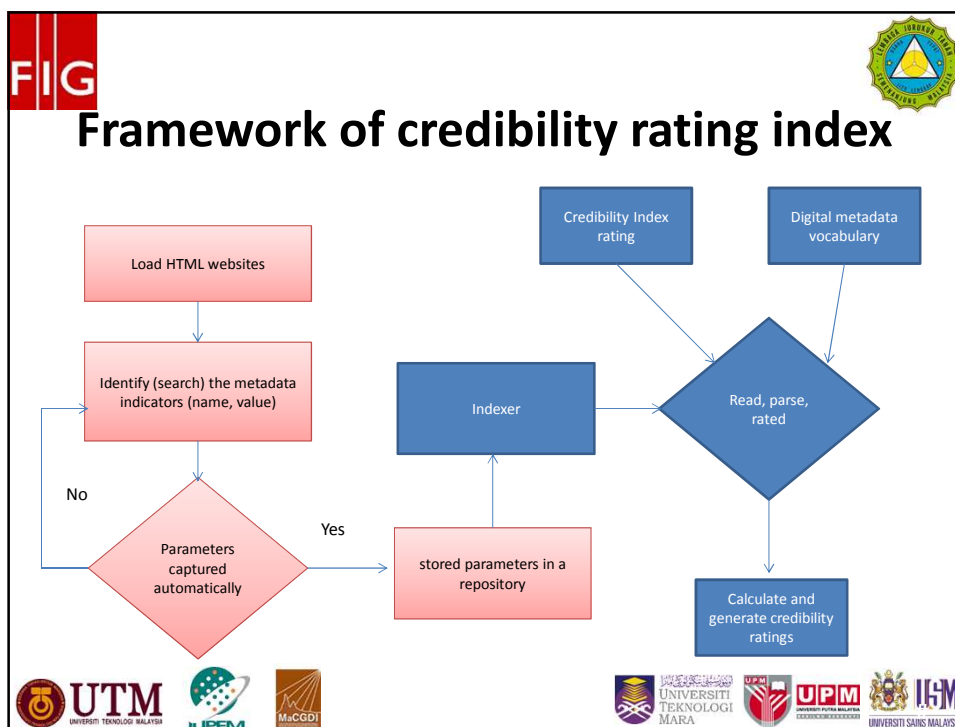


**FIG Digital Metadata vocabulary**

Name (metadata indicator)	Common code expressions	Possible values (keywords)
Currency – publication date	structured	Date, published data
Currency – last updated date	structured	Date, last updated date
Currency for the period under consideration	unstructured	Timeline, temporal, period
Disclose author/creator map mashup	structured	Published by, created by, developed by, author,
Contact information	structured/ unstructured	About us, email, organization, telephone (phone), fax number,
Disclose identity of background data	unstructured	Copyright
Disclose identity of foreground data supplier	structured	Source, supplied by, origin,
Reputation domain URL	structured	URL http://
Affiliation/association	structured	Hyperlinks
Seal approval	unstructured	Approved by,
Sponsorship	unstructured	Sponsored by, funded by,
Mission/purposes/ motive	unstructured	About us, Objectives, purposes, goal

The table lists various metadata indicators, their common code expressions (structured or unstructured), and their possible values or keywords. Logos for UTM, UPM, and IISM are visible at the bottom.

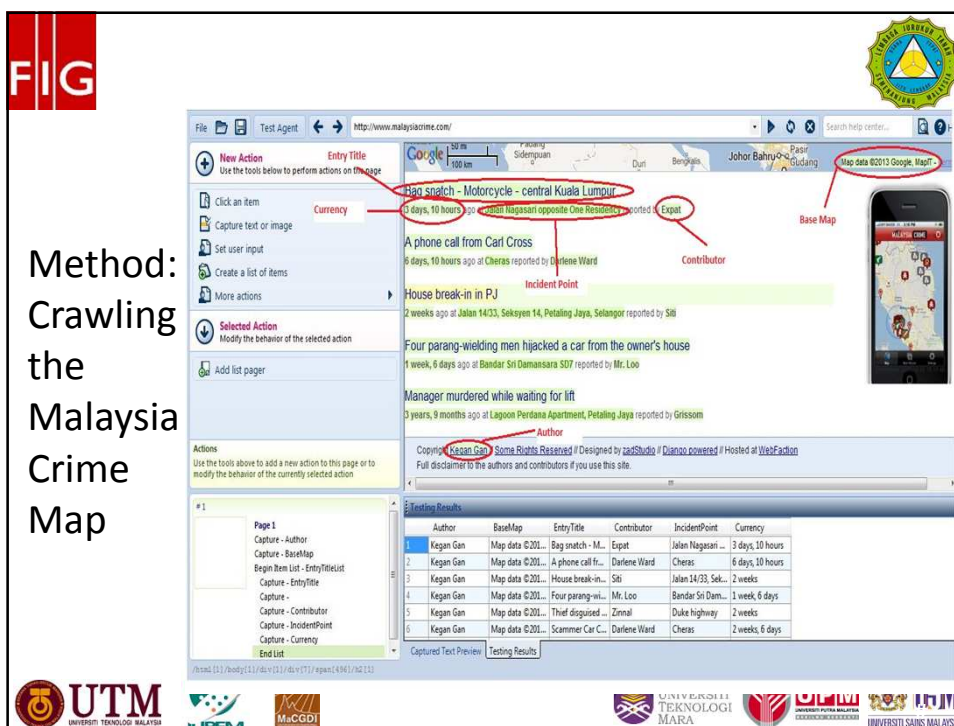
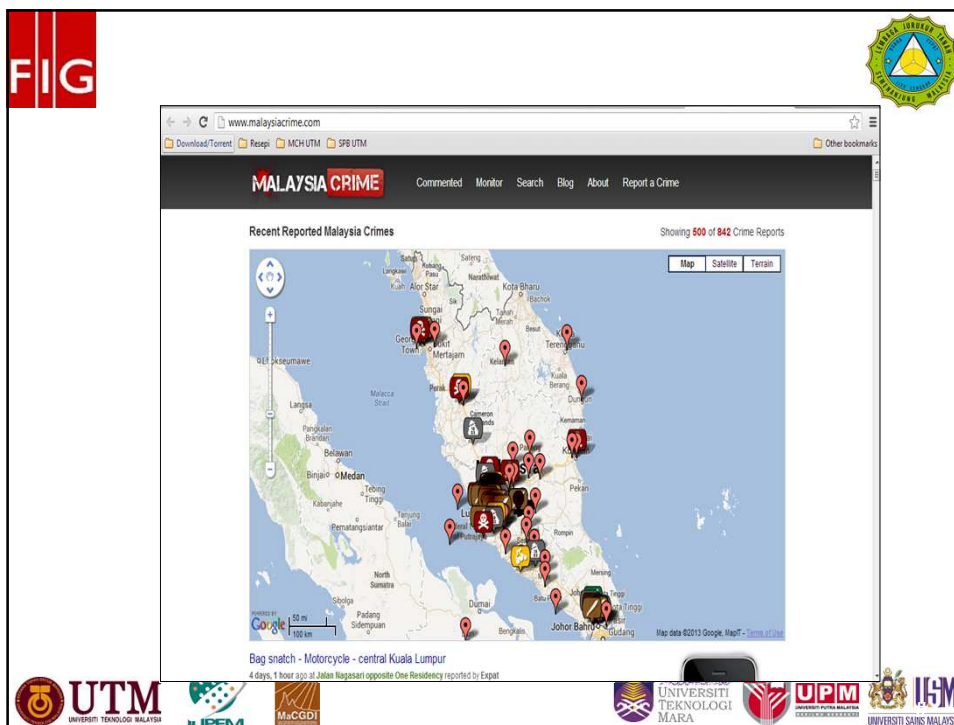






**FIG**

## Case study




Map title	Parameters values					
	Domain URL	currency	Identity of author	Disclose foreground data supplier	Disclose background data supplier	Affiliation / association
Malaysia Crime Map Mashup [1]	<a href="http://www.malaysia-crime.com/">http://www.malaysia-crime.com/</a>	age	Copyright	Reported by	Map data	-
Tropical Cyclone Phailin Map Mashup [2]	<a href="http://google.org/crisismap/2013-phailin">http://google.org/crisismap/2013-phailin</a>	Last updated	Published by	Source	Map Data	-
Mapping South Africa with dot distribution [3]	<a href="http://dotmap.adrianfrith.com/">http://dotmap.adrianfrith.com/</a>		By	Data from	Map data	-
2013 Colorado Floods Map Mashup [4]	<a href="http://google.org/crisismap/2013-boulder-floods">http://google.org/crisismap/2013-boulder-floods</a>	Last updated	Published by	Source	Map data	-
The Ellis Act Map Mashup [5]	<a href="http://www.antievictionmappingproject.net/ellis.html">http://www.antievictionmappingproject.net/ellis.html</a>	-	Created by	-	Powered by	-
London Cycling Census Map Mashup [6]	<a href="http://casa.oobrien.com/traffic/">http://casa.oobrien.com/traffic/</a>	Direct detection	Produced by	Source	Basemap data	-
UTM Campus Map [7]	<a href="http://web.utm.my/campus_map/">http://web.utm.my/campus_map/</a>	Since	Contact	Copyright	Imagery	Direct detection
Oak Mapper Map Mashup [8]	<a href="http://www.oakmapper.org">http://www.oakmapper.org</a>	Recent Submissions	By	-	-	Links
Typhoon Pablo (Bopha) [9]	<a href="http://google.org/crisismap/2012-pablo">http://google.org/crisismap/2012-pablo</a>	Last updated	Published by	Source	Map data	-
Operational UK Renewable Electricity Sites Map Mashup [10]	<a href="http://ukdataexplorer.com/renewables/">http://ukdataexplorer.com/renewables/</a>	Downloaded on	Published by	-	Powered by	-




## Results



Author (creator) of mashup	Point cases	Data Contributor	Currency	Base Map	Incident Point
<b>structured</b>	structured	structured	structured	Unstructured	structured
<b>Kegan Gan</b>	Bag snatch - Motorcycle -	Expat	3 days, 9 hours	Map data ©2013 Google, MapIT -	Jalan Nagasari
<b>Kegan Gan</b>	A phone call from Carl Cross	Darlene	6 days, 9 hours	Map data ©2013 Google, MapIT -	Cheras
<b>Kegan Gan</b>	House break-in in PJ	Siti	2 weeks	Map data ©2013 Google, MapIT -	Jalan 14/33, Seksyen 14

## Detection accuracies

Map ID [See Table 3 ]	Values (count)				
	currency	Identity of author	Disclose foreground data supplier	Disclose background data supplier	Affiliation / association
[1]	500/500	1/1	500/500	1/1	-
[2]	2/3 *one date not captured	1/1	6/7 *one source incorrectly captured	1/1	-
[3]	-	1/1	1/1	1/3	-
[4]	3/4 *one date not captured	1/1	8/9 *one source incorrectly captured	1/1	-
[5]	-	1/1	-	3/3	-
[6]	1/1	1/1	1/1	1/2*one data supplier not captured	-
[7]	1/1	1/1	1/1	1/1	1/7
[8]	-	1/1	-	-	5/5
[9]	1/3	1/1	14/14	1/1	-
[10]	1/1	3/3	1/1	1/1	-






## Remarks

This study demonstrates **the ability of a web crawler to detect metadata indicators** and values before it can be indexed to generate a credibility rating of a map mashup application.

The possible techniques to detect the structured and unstructured metadata by using a web crawler

The results indicate the efficiency of metadata indicators and values to be detected in an automated manner.


This could **reduce the cost of labor and additional tasks** required to generate a list of metadata indicators before it could be indexed to generate credibility ratings.

## Remarks


**The relevance of this study to Surveying Community**

- Current tools available for **the automated creation of metadata** for the assessment of credibility and trust **are not yet adequate** for professional purposes.
- As crowd-data is increasingly seen as providing a **significant data contribution** to authoritative data sets and especially spatial data infrastructures (SDI's)
- This study represent **a step towards a formal quality assessment** framework within which crowd-data may be accompanied by metadata









FIG

## Limitations




- The ideal way for the web crawler (harvester) to assess the credibility of map mashup is through a fully intelligent automated mechanism
- The current web crawler tools used in this study **still require human assistance.**
- The **less intelligent** of the web crawler tools used in this study to learn the similar expression of metadata











FIG


## Future works



- To develop a digital metadata vocabulary – large sample of dataset
- To develop a credibility rating indexer component in order to support the framework of assessing credibility of map mashup application in semi-automated manner.






**FIG** 

## Acknowledgement

- Financial assistance from Malaysian Land Surveyors Board (LJT).
- This project was funded by the Universiti Teknologi Malaysia (UTM) and Ministry of Higher Education Malaysia (MOHE) under Research University Grant vote no. 08J94.



**FIG** 

## Thank you for your attention!

Corresponding author:  
hawani@utm.my  
Geoinformation Department  
Faculty Geoinformation and Real Estate  
Universiti Teknologi Malaysia (UTM)

