

The Route Towards *The Shawshank Redemption*:  
Mapping Set-jetting with Social Media

Mengqian Yang

A Thesis  
in  
The Department  
of  
Geography, Planning & Environment

Presented in Partial Fulfillment of the Requirements  
for the Degree of Master of Science (Geography, Urban and Environmental Studies) at  
Concordia University  
Montreal, Quebec, Canada

August 2015

© Mengqian Yang, 2015

**CONCORDIA UNIVERSITY**  
**School of Graduate Studies**

This is to certify that the thesis prepared

By: Mengqian Yang

Entitled: The Route Towards *The Shawshank Redemption*:  
Mapping Set-jetting with Social Media

and submitted in partial fulfillment of the requirements for the degree of

**Master of Science (Geography, Urban and Environmental Studies)**

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final Examining Committee:

Dr. Pascale Biron      Chair  
\_\_\_\_\_

Dr. Thierry Joliveau      Examiner  
\_\_\_\_\_

Dr. Christian Poirier      Examiner  
\_\_\_\_\_

Dr. Sébastien Caquard      Supervisor  
\_\_\_\_\_

Approved by \_\_\_\_\_  
Chair of Department or Graduate Program Director

\_\_\_\_\_  
Dean of Faculty

Date

2015  
\_\_\_\_\_

## **ABSTRACT**

The Route Towards *The Shawshank Redemption*:

Mapping Set-jetting with Social Media

Mengqian Yang

With the development of the Web 2.0, more and more geospatial data are generated via social media. This segment of what is now called “big data” can be used to further study human spatial behaviors and practices. This project aims to explore different ways of extracting geodata from social media in order to contribute to the growing body of literature dedicated to studying the contribution of the geoweb to human geography. More specifically, this project focuses on the potential of social media to explore a growing tourism phenomenon: set-jetting. Set-jetting refers to the activity whereby people travel to visit shooting locations that appear in movies. The case study presented here focuses on the Mansfield Reformatory (Ohio, USA), which was used as the shooting location for the film *The Shawshank Redemption* (Dir. Frank Darabont, 1994). Through the analysis of georeferenced data mined from Twitter, Flickr, and Tripadvisor, this project presents and discusses the differences and similarities between the use of these three platforms by set-jettors to share and access geodata associated with an alternative tourist destination. The results demonstrate the complementarity of each of these applications to studying set-jetting at different scales. While Twitter appears more appropriate to study this phenomenon at a global scale, Tripadvisor provides more relevant information at the regional level and Flickr can be mobilized to study the movements of set-jettors at a very local scale. Overall, beyond the methodological and technological issues associated with the use of these social media in studying the geography of set-jetting, these applications offer new perspectives for the tourism industry and open new research areas for academics as well.

## **Acknowledgements**

I would like to express my sincere gratitude to the following people. This thesis could not have been realized without their help.

First of all, I would like to thank my supervisor Sébastien Caquard. He showed endless patience, guided me in the right direction, gave me valuable comments and suggestions, and helped me so much during the entire process of completing the thesis.

Second, I would like to thank my friends for their help and encouragement; Taien Ng-Chan helped me to edit the final version of my thesis, and the other members of the Geomedia lab gave me a lot of valuable ideas and suggestions during the past three years.

In addition, I would like to thank my thesis committee members for their time and advices. I would also like to thank all the staffs and teachers of the department of Geography, Planning and Environment at Concordia for their help and support.

Finally, I would like to thank my family for giving me so much love, support and encouragement.

# Table of Contents

<b>Chapter 1: Introduction and Problem Statement .....</b>	<b>1</b>
<b>1.1 Introduction.....</b>	<b>1</b>
<b>1.2 Objectives and Research Questions .....</b>	<b>2</b>
<b>1.3 Organization of the Thesis .....</b>	<b>3</b>
<b>Chapter 2: Literature Review.....</b>	<b>4</b>
<b>2.1 Overview .....</b>	<b>4</b>
<b>2.2 Geodata mining from social media .....</b>	<b>4</b>
2.2.1 Web 2.0.....	4
2.2.2 Crowdsourcing.....	5
2.2.2 Mashup .....	7
2.2.4 Neocartography and Neogeography .....	8
<b>2.4 Geodata from social media.....</b>	<b>13</b>
<b>2.5 Natural Language Processing (NLP) &amp; Geoparsing.....</b>	<b>16</b>
2.5.1 Natural Language Processing (NLP) .....	16
2.5.2 Geoparsing .....	17
<b>2.6 Set-jetting .....</b>	<b>17</b>
<b>2.7 Case Study .....</b>	<b>21</b>
2.7.1 A case study of the Shawshank Trail.....	21
2.7.2 The Ohio State Reformatory.....	23
2.7.3 The cinematic landscape of <i>The Shawshank Redemption</i> .....	24
<b>Chapter 3: Methodology.....</b>	<b>27</b>
<b>3.1 The social media application.....</b>	<b>28</b>
<b>3.2 Data mining methods .....</b>	<b>33</b>
<b>3.3 Data Extraction methods.....</b>	<b>41</b>
<b>Chapter 4: Data analysis and Discussion .....</b>	<b>44</b>
<b>4.1 Overall data collection .....</b>	<b>44</b>
<b>4.2 Spatial analysis.....</b>	<b>46</b>
<b>4.3 Tracking the movement of visitors.....</b>	<b>60</b>
4.3.1 Tracking Tweets .....	62
4.3.2 Tracking places mentioned in the reviews in Tripadvisor .....	62
4.3.3 Tracking photos from Flickr .....	64
<b>4.4 Summary .....</b>	<b>65</b>
<b>Chapter 5: Conclusion and Future Work.....</b>	<b>67</b>
<b>References .....</b>	<b>70</b>
<b>Appendix I.....</b>	<b>78</b>

<b>Appendix II .....</b>	<b>80</b>
<b>Appendix III .....</b>	<b>83</b>

## List of Figures

Figure 1. Screen capture of the area near Mansfield Reformatory as it appears in OpenStreetMap .....	7
Figure 2. A map containing keyword “Shawshank” generated from geotagged images from Flickr.com .....	11
Figure 3. Interactive Map of Movie London .....	21
Figure 4. Plot in <i>The Shawshank Redemption</i> .....	23
Figure 5. The west cell block of the Ohio State Reformatory .....	25
Figure 6. The overall view of Ohio State Reformatory .....	25
Figure 7. Research Design Process .....	27
Figure 8a. Example of a geotagged tweet with a place name appearing in the body of the text from Twitter .....	29
Figure 8b. Example of a geotagged tweet with an embedded picture of the Ohio State Reformatory from Twitter .....	29
Figure 9. Ohio State Reformatory webpage on Tripadvisor.....	32
Figure 10. REST request .....	37
Figure 11. Streaming API request.....	38
Figure 12a. The distribution of geotagged photos from Flickr containing “Shawshank + Prison/Jail” at four different scales .....	48
Figure 12b. The distribution of geotagged tweets containing “Shawshank + Prison/Jail” at four different scales .....	48
Figure 13a. Satellite view of Mansfield reformatory.....	50
Figure 14a. Geolocated photos mined from Flickr near Mansfield Reformatory.....	52
Figure 14b. Geolocated Tweets mined from Twitter near Mansfield Reformatory .....	53
Figure 15a. West cell block, Photo taken by David Perl .....	53
Figure 15b. Entry, photo taken by Michael Savino.....	54
Figure 15c. Overview, photo taken by Doug Butchy.....	54
Figure 16a. & 16b. Annual number of reviews about the Mansfield Reformatory posted on Tripadvisor (16a) and number of photos uploaded on Flickr (16b) .....	56
Figure 18a. The origins of the visitors to Mansfield Reformatory on Flickr, Twitter, Tripadvisor at 3 different scales .....	58

**Figure 18b. The hot spot of the origins of the visitors to Mansfield Reformatory on Flickr, Twitter, Tripadvisor in national scale .....59**

**Figure 19. Tracking routes of Twitter users .....61**

**Figure 21a, b & c. Places mentioned in Tripadvisor in reviews talking about the Mansfield Reformatory .....63**

**Figure 20a. & b. Tracking routes of Flickr users .....64**



## List of Tables

<b>Table 1. Impact of the Film on visitors .....</b>	<b>20</b>
<b>Table 2. Description of each of the Search API returned parameters .....</b>	<b>36</b>
<b>Table 3: Explanation of Search API parameters.....</b>	<b>39</b>
<b>Table 4. Extracted results from Twitter, Flickr and Tripadvisor .....</b>	<b>42</b>
<b>Table 5. Overall view of the mined data from Flickr, Twitter, Tripadvisor .....</b>	<b>45</b>

# Chapter 1: Introduction and Problem Statement

## 1.1 Introduction

Since the 1960s, digital technologies have percolated into people's lives. This percolation has accelerated towards the end of the 20<sup>th</sup> century with the development of the Internet and the emergence of what is now called the "information age." In terms of information technology, the end of the 20<sup>th</sup> century corresponds also to the beginning of the Web 2.0. The Web 2.0 can be briefly defined as the interactive version of the original Web in which people can actively contribute through the production and distribution of content. **Social media** (such as Twitter, Facebook, Flickr, etc.) is one of the most archetypical forms of the Web 2.0. Social media can be defined as "collaborative websites and applications" that enable users to "create and share content and participate in social networking" (Oxford dictionary, 2012). Social media has transformed our society into a more closely connected world (Tsou & Leiter, 2013).

The development of the Web 2.0 has also impacted geography. In the Web 2.0 era, maps have become more prevalent due to the development of delivery mechanisms for geographic information and maps (Haklay et al., 2008). This has had several impacts on people's lives. Thanks to the increasing availability of appropriate Internet technologies and mobile applications, as well as the democratization of online mapping services such as Google Maps, maps play a more and more central role in our daily activities (Faby & Koch, 2010). Maps can serve to locate a store, and to find the shortest route to get to a place, as well as to share personal experiences related to places. In virtual communities, people use different mapping platforms (e.g. Google Maps) to publish and share their whereabouts to friends as well as to collect, evaluate and disseminate interesting information linked to certain places (Fischer, 2012). Therefore, a huge

amount of place-related data is generated and distributed on a daily basis via the Internet, and maps are used more and more often to navigate this data. Maps are also considered as navigational platforms in this ocean of data (November et al., 2010).

The use of social media by a growing number of individuals around the world to characterize places offers a great opportunity to study our relationships to places through the filter of social media. Geographers can now trace, monitor, and map the spread of social movements, protests, disease outbreaks, natural hazards, elections and political campaigns in cyberspace by digitally collecting social media and online content (Tsou & Leitner, 2013).

In this context, I am interested in further exploring how social media can be mobilized to study the relationships between fictional and real places. I do so by exploring the potential of social media to study a growing phenomenon: set-jetting. Set-jetting is the term used to define the activity whereby people visit movie shooting locations after having watched a movie (Joliveau, 2009). In this study, set-jetting serves as a case study to further explore the potential of social media in human geography, as well as to further explore film-induced tourism.

## **1.2 Objectives and Research Questions**

More specifically, this project seeks to address the following questions:

1. How do people use different social media to share their spatial experiences in the context of set-jetting?
2. How can we better use tools offered by the Web 2.0 to extract and compile geospatial data related to set-jetting from social media?
3. How can we use this data to study movements and spatial patterns?

### **1.3 Organization of the Thesis**

In order to address these questions, the thesis is organized as follows: Chapter 2 sets the historical and conceptual contexts for this project and introduces the case study. Chapter 3 presents the methodological choices made to address the research questions. In Chapter 4, the results are analyzed at different scales and discussed. Finally, the limits and the potential of the applications are discussed in the conclusion section.

## **Chapter 2: Literature Review**

### **2.1 Overview**

The objective of this section is to help the reader to understand the theoretical and technological foundation of this project. In order to do so, I will start by defining some of the main concepts and practices associated with the Geoweb. This contextualization will be followed by a review of the potential of the Web 2.0 to generate geospatial data and of different techniques developed to mine and retrieve these data. Finally, I will also introduce the practice of set-jetting with a specific focus on my case study: set-jetting generated by the movie *The Shawshank Redemption* (dir. Darabont, 1994).

### **2.2 Geodata mining from social media**

#### **2.2.1 Web 2.0**

The term “Web 2.0” was first coined by Tim O’Reilly in 2005 and used to describe the trends and technologies that enable individuals and communities to create, develop and share information through the Internet (O’Reilly, 2005; Haklay et al., 2008). A typical Web 2.0 site can allow users to interact with others in a virtual community. As described in Wikipedia (one of the archetypal Web 2.0 sites), Web 2.0 applications include “social networking sites, blogs, wikis, video sharing, sites, hosted services, Web application, mashups and folksonomies” (“Web 2.0”, 2014, para.2). Social media sites such as Facebook, Twitter, YouTube and Flickr are part of the Web 2.0 ecosystem.

According to Anderson (2007), six of O'Reilly's principles can be used to define the Web 2.0 environment (Anderson, 2007; O'Reilly, 2005; Batty, et al, 2010):

- ◆ Individual production of user-generated content;
- ◆ Harnessing the power of the crowd;
- ◆ Data is collected and available on an epic scale;
- ◆ Architecture of participation;
- ◆ Network effects; and
- ◆ Openness.

The Web 2.0 is then used to generate and share data including geographic data. The merging of the Web with geospatial technologies is now called the Geoweb (Herring, 1994). Geoweb applications are tightly connected to the Web 2.0 (Roche et al., 2013). Online mapping services such as Google Maps, OpenStreetMap, and Bing Maps are the most popular Geoweb applications. The success of these applications and of the Geoweb in the past few years has been accompanied by the emergence of new practices and concepts, such as crowdsourcing, mashup, Neocartography, Neogeography, big data and geodata. In the following pages I will describe and discuss each of these terms.

### **2.2.2 Crowdsourcing**

The term "crowdsourcing" is derived from the concept of outsourcing business operations. It has been originally used to describe the process of outsourcing businesses to remote cheaper locations (Friedman 2006). In the context of the Web 2.0, crowdsourcing is now defined as

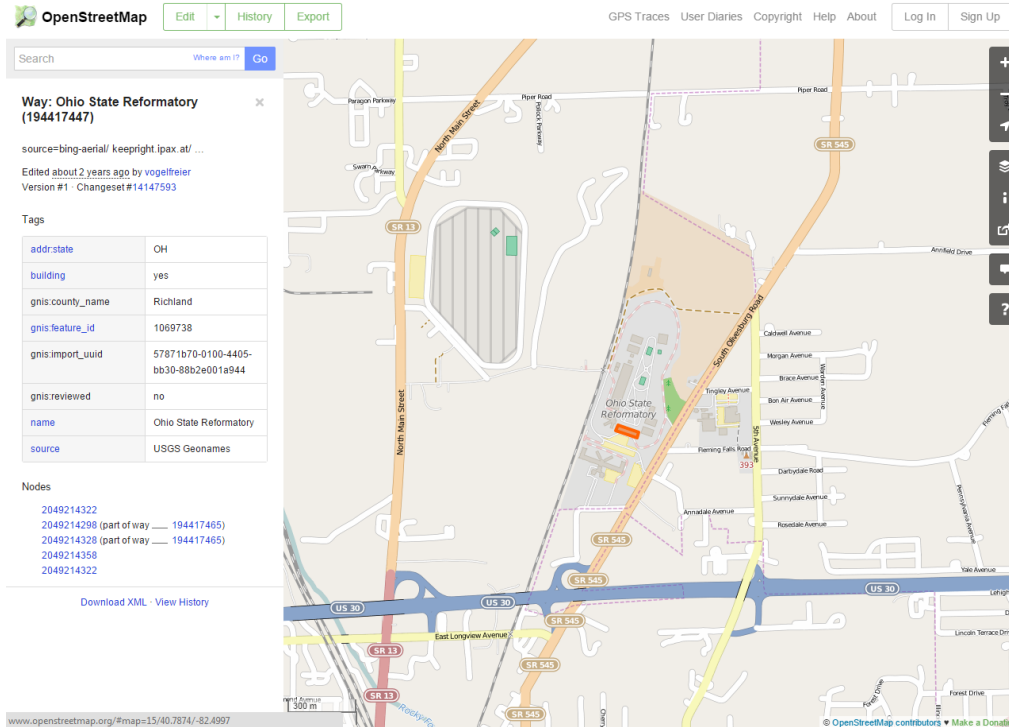
"a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge,

heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task” (Estellés-Arolas & González-Ladrón-de-Guevara, 2012, p.9).

Crowdsourcing has been proven to be one of the most significant outcomes of the Web 2.0 (Haklay et al., 2008). This phenomenon was further explored by Michael Goodchild through the concept of Volunteered Geographic Information (VGI, Goodchild, 2007) that he described as follows:

“the widespread engagement of large numbers of private citizens, often with little in the way of formal qualifications, in the creation of geographic information. They are largely untrained and their actions are almost always voluntary, and the results may or may not be accurate. But collectively, they represent a dramatic innovation that will certainly have profound impacts on geographic information systems (GIS) and more generally on the discipline of geography and its relationship to the general public.” (Goodchild, 2007, p.212)

OpenStreetMap and Google (via Google Map Maker) rely on geocrowdsourcing for the development and continuous updating of their base maps. OSM (OpenStreetMap) is often considered the most successful crowdsourcing map project (Hudson-Smith et al., 2009).



**Figure 1.** Screen capture of the area near Mansfield Reformatory as it appears in OpenStreetMap (source: <http://www.openstreetmap.org/> date: 2014 July)

## 2.2.2 Mashup

The term “mashup” – originally coined by DJ DangerMouse’s “The Grey Album” (Hoff, 2006) – first appeared in 2004 to describe the mixing and combination of different music tracks. The term now refers to websites that weave data from different sources into new integrated user services (Hoff, 2006). A “map mashup” refers to the plotting on a base map provided by one source (e.g. Google maps) of geodata from another source (e.g. personal data). The idea of mashup gives more possibilities and potential to maps. Through the map mashup anybody with limited technological knowledge can use online mapping services such as Google Maps and OpenStreetMap to hash and map his/her own data. Since the emergence of Google Maps in January 2005, map mashups have become extremely popular. Data can be plotted on the map, and different types of media such as photos and videos can be geotagged and mapped as well. According to the statistics from the programmable web website



(<http://www.programmableweb.com/mashups>), in 2008 more than one third of 7034 mashup websites tracked in that project use Google Maps APIs (Application Programming Interface), and 26% of these websites are tagged with the term “mapping.” Maps have become the major form of mashup (Marjit & Jana, 2009; Batty, 2010).

#### **2.2.4 Neocartography and Neogeography**

In the context of the Web 2.0, the term Neocartography is currently used to characterize the emergence of new forms of mapping technologies and new types of applications (Faby & Coch, 2010). Neocartography has been defined as

“[...] neo-geographic characteristics with ubiquitous cartography and geo-media techniques. Beside a time and space independent access to maps and modification of geospatial data, neo-cartography takes the characteristics of transmitting media, the impact of information-content and user needs for the presentation of geospatial information into account. The new aspects of neo-cartography indicate the possibility to directly access mental imagery by using user inputs. The ubiquitous existence of maps and a public participation develop a social imagery of space that should be used for the abstracted and simplified presentation of space” (Jobst 2009, p.214).

In short, Neocartography captures the idea that users are now able to develop and self-publish their own maps via the Web 2.0 (Cartwright, 2012).

The combination of geography with Web 2.0 technologies has become known as “Neogeography” (Hudson-Smith et al., 2009). This derived term of “Neogeography” is used to define the use of Web 2.0 techniques to geolocate spatial data available on the Web. Turner (2006) has described Neogeography as a subject that combines the complex techniques of cartography and GIS and

makes it more reachable for the users and developers (Turner, 2006). According to Eisnor on [palatial.com](http://palatial.com) (2006), Neogeography can be defined as

“a diverse set of practices that operate outside, or alongside, or in the manner of, the practices of professional geographers. Rather than making claims on scientific standards, methodologies of Neogeography tend towards the intuitive, expressive, personal, absurd and/or artistic, but may just be idiosyncratic applications of ‘real’ geographic techniques.”

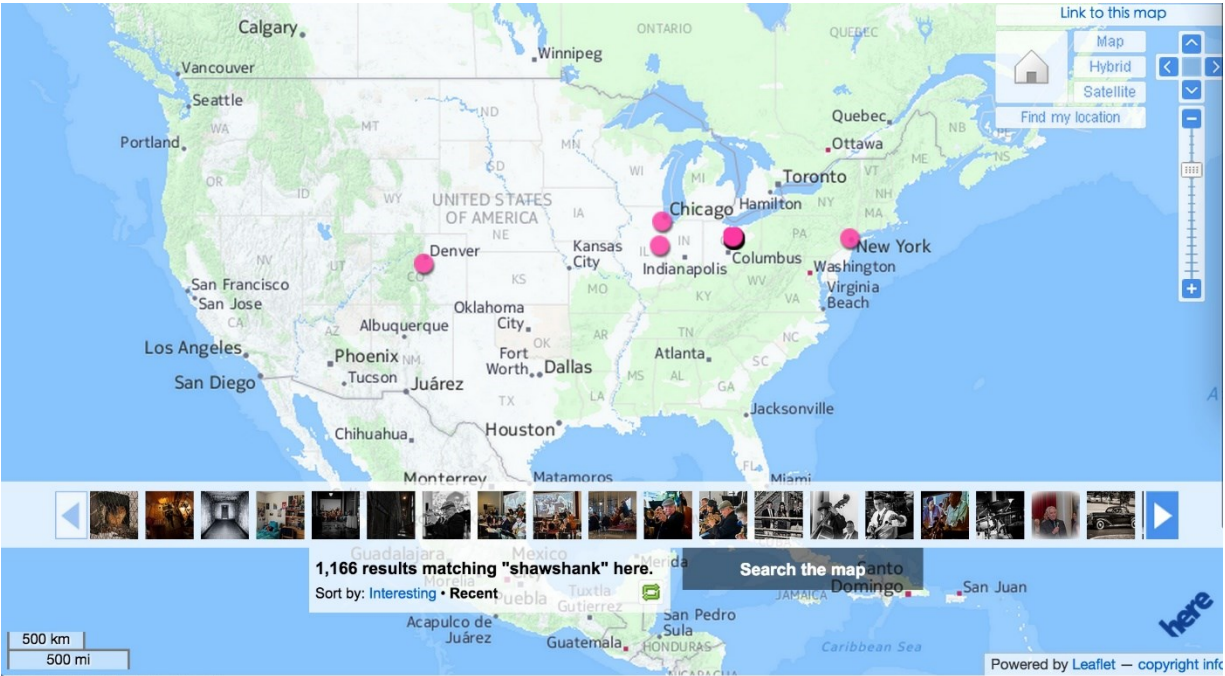
In other words, Neogeography can be synthesized as the use of geospatial technologies provided by the Web 2.0 by non-expert users for a range of purposes (Goodchild, 2009).

Due to the fast development of online mapping tools as well as of the Web 2.0, there has been a growing need for standardization of data formats. While Open Geospatial Consortium (OGC, [www.opengeospatial.org](http://www.opengeospatial.org)) has become a significant force in developing standards for Web mapping, including Web Mapping Service specification (OGC, 2000) and Web Map Tile Service (OGC, 2010) (Haklay & Weber, 2008, Batty et al., 2010), private sector companies have also imposed their open data standards such as Google with its KML and KMZ files. The advent of the Web 2.0 technologies and standards allows interoperability of data across different platforms and tools.

One of the keys to the success of Neogeography is the possibility offered by the Geoweb’s geotag media. Geotagging (also written as GeoTagging) can be seen as a process of adding geographical identification metadata to various media, such as photographs, video, websites, SMS messages, as well as geoinformation to QR Codes (abbreviated from Quick Response Code) or RSS (Rich Site Summary) feeds (Luo et al., 2011). One clear example of the power of geotagging is the online photosharing platform Flickr. Flickr is not only used to upload and share

photographs, but also to retrieve and map spatially these photos. Different authors for different studies have retrieved geotagged images from Flickr. For instance, Fischer used the retrieved photos to identify places in cities attracting mainly local photographers versus places that attracted mainly tourists (Fischer, 2012).

Flickr is not only a photo sharing platform but also an example of a Web 2.0 website that uses geotagging and mapping to store the location data of photos and images that allows the mapping of the locations of the photos. When users upload photos and images to Flickr, they can also tag them with geographic coordinates. People can then plot the geolocated images on a Nokia basemap and create personal maps with these collectively shared images (Haklay et al, 2010) (see Figure 2). This kind of Web 2.0 website contributes a large amount of spatially coded data that can be mapped and analyzed geographically. The Geoweb is thus made of massive amounts of data collected through a process called “Crowdsourcing.” The geotagged portion of this data can be mapped more and more easily with a range of mapping services (e.g. Google Maps) creating what is called mashups. These maps can be generated and interpreted by anyone with a basic understanding of the Internet. These tools, data and practices are part of what is now called Neocartography and Neogeography.



**Figure 2.** A map containing the keyword “Shawshank” generated from geotagged images from Flickr.com (source: <http://www.Flickr.com/> retrieved date: 2014 December)

In conclusion, the Web 2.0 can be seen as the framework enabling new forms of developments and interactions with geospatial data through maps. Neogeography could then be seen as an extension of the discipline of geography, interested in further exploring the potentialities offered by combining the Web 2.0 with geographic information and geospatial technologies.

Geocrowdsourcing characterizes the process of collaboratively collecting geospatial information via the Web 2.0, while map mashups characterize the plotting of this data through online mapping services. Finally, Neocartography characterizes the changes that the discipline of cartography is experiencing with all of these new technologies and practices. In this project, I aim to explore how these tools, data and practices can be mobilized to study set-jetting.

**2.3 Social media**

Social media is an outcome of the Web 2.0 era. Today, millions of people are connected via social media. In cyberspace, people use social media to create, share and exchange different

kinds of information. There is no formal definition of social media. Kaplan and Haenlein envision social media through their relationships with the Web 2.0 and User Generated Content (UGC). These last two concepts are often mentioned together. The Web 2.0 is considered as the platform of social media; it provides the technological foundation of social media. UGC is usually “used to describe the various forms of media content that are publicly available and created by end-users,” and can be seen as the “sum of all ways in which people use social media content” (Kaplan & Haenlein, 2009, p.61). UGC is often used for crowdsourcing. Thus, social media can be generally understood as "a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content" (Kaplan & Haenlein, 2009, p.61).

The number of social media applications and users has been increasing dramatically with the emergence of popular sites such as Myspace (2003) and Facebook (2004). The history of social media can be traced back to “Usenet” (1979), a worldwide discussion system, and “Open Diary” (1998), a social networking diary sharing community (Kaplan & Haenlein, 2009).

By January 2014, the number of active users on Facebook was already over 1.23 billion (“Facebook passes 1.23”, 2014). The Pew Research Center’s Internet & American Life Project conducted a survey in late 2012. There were 1,802 respondents, and the data showed that 67% of all participants had previously used a social networking site, and adults between 18 and 49 were the main group using social network websites. Main social media websites such as Twitter and Facebook are mainly used by young adults (Age 18 to 29), while other ages are interested in different websites and services (Duggan & Brenner, 2013).

Due to the development of new Web 2.0 technologies, the format of information available via social media is multiple. Messages are no longer limited to simple words, and other formats of

information such as pictures and videos are also easily transferred through different kinds of social media websites. It is difficult to find an accurate figure of the number of social media websites since everyday there are new ones created. It is also difficult to categorize different types of social media websites because of the evasive definition of social media. All of today's most popular websites, such as Wikipedia, Twitter, Facebook, and Flickr are part of the social media ecosystem.

## **2.4 Geodata from social media**

Some people argue that the Web 2.0 era can be also called the Data 2.0 era (Brown et al., 2011). Social media is one prominent source that generates large amounts of data, and the term "Big Data" is often associated with social media. In 2012, Big Data became a well-known term since the Obama Administration invested 200 million dollars in the Big Data Research and Development Initiative (White House, 2012).

According to Wikipedia, Big Data can be defined as

“a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications. The challenges include capture, creation, storage, search, sharing, transfer, analysis and visualization “(“Big Data”, 2012, para. 2).

In this definition, the important word is not only “large” but also “complex.” Big data includes data from various sources and different formats. The term “Big Data” first appeared in the open source project “Nutch” of apache.org, and was used to describe the big data set for updating and analyzing search engines (Krishnan, 2013). In 2011, Mckinsey and Company anticipated the coming of the “Big Data era.” They reported that effectively collecting, storing, and analyzing Big Data raises a range of new issues that have been synthesized by Manovich (2011, p.39):

“Big data changes the definition of knowledge; claims to objectivity and accuracy are misleading, bigger data are not always better data, taken out of context, big data loses its meaning, just because it is accessible doesn’t make it ethical, limited access to big data creates new divides.”

The term Big Data is not only used to describe the data that comes from the Internet, but also to describe the massive data coming from a variety of fields, such as biology, geography, and chemistry, and has become a hot topic in different research domains (Manyika et al., 2011).

The booming of social media data also raises challenges for geographers and cartographers. With more and more people owning GPS enabled smartphones, “location” becomes a key element in many social media activities. It works as a central organizing theme engaging with users’ online interaction and communication (Caverlee et al., 2013). For example, the Foursquare location sharing service has enabled over 4.5 billion “check-ins” whereby users can share their presence, notes, and photographs to a particular venue. The mobile image sharing service, such as Instagram, allows users to attach the latitude-longitude coordinates to their photograph. Flickr and a lot of other services such as Pinterest and Tumblr also provide similar geotagged image-sharing services. 500 million Tweets are published every day on Twitter, of which around 5 million (1%) are tagged with latitude-longitude coordinates (Twitter, 2013). As emphasized in a recent Pew Research Center report (Duggan & Smith, 2013) location is now an increasingly central part of the social media experience.

Although geotagged data makes up a very small proportion of the total data accumulated daily over the Internet, the overall numbers are so vast that geotagged ones have become a meaningful source of data. This growing development of geotagged content from social media has stimulated

growing interest in academia in the last few years, including in the field of geography.

Geographers have tackled this new material from two main perspectives: firstly by focusing on geotagged content from conventional social media websites, for instance, Twitter and Flickr. The researchers address a variety of topics, for example, mapping Flickr photos to analyze and reshape the city cores of five main cities in the United States (Hollenstein & Purves, 2015), or studying predictable travel behavior using Flickr geotags (Clements et al., 2010). The range of research based on geotagged tweets is even broader, from crisis management (MacEachren et al., 2011) to the construction of simulated movement of zombies (Behuniak, 2011). Secondly, geographers explore how social media are changing the way people perceive and interact with it. For instance, applications such as Foursquare and Tripadvisor contribute to the attractiveness as well as the repulsiveness of certain sites, locations and businesses.

Mobilizing this data for geographic analysis raises a range of issues. For instance, according to Crampton (2013), the data mined from social media for geographic study is limited in its explanatory value, no matter how big it might be. Tsou and Leitner (2013) have addressed the issue of scale in the relevance of big data to study geographic phenomena. They have compared two studies that used big data at different scales: a first study done by Li, Goodchild and Xu (2013) proposes an overview of the use of a huge database from Twitter and Flickr to study the spatial pattern of the data; and a second one by Kent and Capello (2013) focuses on studying the potential of social media to analyze wildfires at a very local scale. This comparison of the two studies demonstrated that the data mined from social media can be used for both “micro” and “macro” scale research frameworks (Tsou & Leitner, 2013).



## **2.5 Natural Language Processing (NLP) & Geoparsing**

### **2.5.1 Natural Language Processing (NLP)**

Another major issue that has been tackled is the extraction of meaningful geospatial information from this ocean of data through the use of method Natural Language Processing (NLP).

The geographic information embedded in social media can be extracted in different ways. The easiest one is when the data is geotagged with geographic coordinates (i.e. lon/lat). Data can then be easily searched and mined based on these coordinates. Another way to extract geographic data from social media is through place names. Place names can be geotagged to different media (e.g. the name of a city associated to a picture). Place names can also appear in the body of text, for instance in a tweet or a Facebook page. Names of cities, neighborhoods, rivers as well as addresses can be identified through a capturing process known as Place Name Recognition (PNR). PNR is an instance of the studies of Named Entity Recognition (NER), which has been extensively studied in the domain of the Natural Language Processing (NLP) community (Freire et al., 2011). NLP is a field combining computer science, artificial intelligence and linguistics to help human-computer interaction. It can be seen as a bridge between computer and human language (Cunningham et al., 2002).

The study of NLP can be traced back to the 1950s. In 1950, Alan Turing started to develop the Turing test to make a creation of intelligence. The study of translating natural language into machine language evolved quickly in the 1950s. However, in the 1960s, due to the failure of many experiments, the period of stagnation of NLP began. In the 1980s, the increase of computer power stimulated the redevelopment of NLP study (Gazdar, 1996). Recent research focused on unsupervised and semi-supervised understanding of algorithms. More recently, Natural Language Processing has been used with the goal of improving the extraction of data based on locations (Zong, 2008).

### **2.5.2 Geoparsing**

The process of identifying and geocoding place-names from digital documents is known as geoparsing. In the Natural Language Processing field, it is referred to as Named Entity Recognition (NER), and it is the central core of text processing applications such as Information Retrieval (IR) and Information Extraction (IE). Geoparsing can be defined as the process of identifying place names in a text and recognizing near geographic context. This process connects the unstructured textual realm with the structured realm of geo-information (Leidner, 2008). Once a text has been geoparsed, the place names identified can then be sent to a gazetteer to be associated with geographic coordinates. Geoparsing is most often used to automatically analyze collections of text content. There are a number of companies producing commercial geoparsing products, such as Attensity Server (Attensity), Digital Reasoning (GeoLocator), ThingFinder (Inxight), and SRA (NetOwl). These products have the capability of extracting places with different entities, such as time, persons, money, etc. (Abascal-Mena & López-Ornelas, 2010).

Geographers are obviously interested in geoparsing. For instance, Xu, Wang and Yang (2013), have applied NLP to analyze and evaluate Twitter users' geographical awareness at different scales by using data mining skills. Another example is provided by Gelernter and Mushegian (2011), who studied how to geoparse crisis tweets and create disaster mapping after the earthquake in Christchurch, Canterbury, New Zealand (2011). In the context of this master project, I am interested in studying how geoparsing technologies and tools could be mobilized to mine social media context related to set-jetting.

### **2.6 Set-jetting**

The history of cinema started with the invention of the first movie camera in the late 1880s. With the spread of films in the 20<sup>th</sup> century, people's normal lives became more and more closely

intertwined with cinema. It can be said that to some degree, our contemporary life in the early 21<sup>st</sup> century was formed by films (Wenders, 2001). Cinema has a far-reaching influence on our habits, our perspectives, as well as our minds, and cinema has changed our way of thinking, moving and seeing things (Wenders, 2001).

Films, as an influential medium, can convey ideas, information and perspectives, and motivate people's behaviors. The specific characteristics of films can attract people's attention, which has been accomplished through the use of props, costumes, and food (Hudson & Ritchie, 2006). Films can act as an advertisement support and can contribute to making a market with huge economic potential.

Films have a strong relationship with places. First and foremost, because places often play a key role in films, the relationship between films and places go beyond artistic and narrative. Films are more and more often used as promotion vehicles for places such as cities. Films can be beneficial to certain locations not only because of the money they generate throughout the production and post-production processes, but also because they contribute to promoting the image of the places where the action unfolds.

As described by the writer J. B. Jackson (1979), the landscape of films can work as a "theater" (Landscape Magazine, 1979). In other words, locations play a significant role in creating an environment that "places" viewers "into" a film (Jackson, 1979). Jackson argues that films "*emphasized the visual, the spectacle in the sense of dramatic production with a well-defined space, an organization of place and time; and a coherent action*" (p. 4). It is the human control and design that makes the landscape of films work as a theater to stage social and artistic productions, bringing us to see ourselves (Jackson, 1979; Lukinbeal, 2005). The term *cinematic landscape* is used to define the fundamental relationship between films and places.

Since the 1950s, especially during the last 20 years, geographers have become interested in film locations (Kennedy & Lukinbeal, 1997; Lukinbeal 2012; Caquard & Cartwright, 2014; Hallan & Roberts, 2012). Their studies focus on a variety of subjects, such as the construction of meaning and place in films (Kennedy, 1994), film-induced touristic phenomena (Jones & Smith, 2005), film-related industry districts (Coe, 2000) and the history of cinema-going in different places around the world (Hallan & Roberts, 2014).

Shooting locations is one of the subjects that geographers have investigated, including the more specific relationships between geospatial technology and set-jetting (Joliveau, 2009). Set-jetting can be defined as the process of visiting places where movies were filmed (Joliveau, 2009). This concept is related to “*on-location tourism that follows the success of a film made (or set) in a particular region*” (Beeton, 2005, p.9), which is also called a “film-induced tour.”

Due to the globalization of the film industry and the democratization of world travelling, set-jetting has become a hot leisure activity with important economic impact (Bolan & Williams 2008; Tzanelli, 2004). Blockbusters such as *The Lord of the Rings* (Dir. Jackson, 2001), *The Da Vinci Code* (Dir. Howard, 2003) and the Harry Potter movies have attracted lots of tourists respectively in New Zealand, Paris and London (Joliveau, 2009). These tourists want to walk in the steps of their heroes and to visit the places they have seen in their favorite movies. This type of behavior is synthesized by Bolan and his colleagues (2011, p.35): “We have seen it in the movie, let’s see if it is true.” Every year millions of people travel to different destinations with the intent of visiting those set locations (Macionis, 2004). Riley and colleagues (1998) studied 11 films and found that the peak of interest appears after the release of the film with an important 50% increase per year in visitation during at least five years after the release. The study by Hudson and Ritchie (2006) of the attractivity of 10 movies distributed globally emphasizes the different

impacts of film-induced tourism on the shooting locations as well as the duration of this impact (see table 1).

**Table 1.** Impact of the Film on visitors (Source: Hudson & Ritchie (2006a))

Film	Location	Impact of visitor number
Braveheart	Wallace Monument, Scotland	300% increase a year after release
Captain Corelli's Mandolin	Cephalonia, Greece	50% increase
Field of Dreams	Iowa	35,000 visits in 1991 and steady increase every year
Four Weddings and a Funeral	The Crown Hotel, Amersham, England	Fully booked for at least 3 years
Harry Potter	Various locations in UK	Increase of 50% or more in all filmed locations
Mission Impossible 2	National Park, Sydney	200% increase in 2000
Notting Hill	Kenwood House, England	10% increase in one month
Pride and Prejudice	Lyme Park, England	150% increase
Sense and Sensibility	Saltram House, England	39% increase
The Beach	Thailand	22% increase in youth market in 2000
Troy	Canakkale, Turkey	73% increase

In another study, Riley and his colleagues (1998) have analyzed the changes in visitor numbers at 10 former film sites in the United States. Busby & Klug (2001) conducted a study at the visitor profiles in Notting Hill, a famous setting location, in London. Connell (2005) reports a survey of tourism business operators on Mull, Scotland to study the “toddler tourism” caused by a children's television programme called *Balamory*. Iwashita (2006) surveyed how UK popular television impacts Japanese visitors coming to the UK. O'Connor (2008) and his colleagues aimed to create a sustainable brand for Northern Ireland through film-induced tourism. In addition, Frost (2010) analyzed how 22 fictional-feature films set in the Outback, Australia, have influenced viewers' perceptions. These studies emphasize the fact that not only does cinema affect our imaginary of places through films, but that films also impact the reality of place by attracting tourists and modifying the spatial structures and dynamics via these flows of tourists.



**Figure 3.** Interactive Map of Movie London (Source: [www.movielondon.net](http://www.movielondon.net))

Geospatial technologies have been used to promote set-jetting. For example, the website VisitBritain provides a movie map highlighting over 200 shooting locations of films across Britain (see Figure 3). This map does not only show the locations but also the videos and photos corresponding to the scene, which means that visitors could easily follow in the footsteps of their favorite characters (Demetradi, 1996). As pointed out by Joliveau (2009), new geospatial technologies can also be used to further study the geography of set-jetting.

The goal of this project is to study how social media can be mobilized to further study set-jetting. This question is addressed through the analysis of set-jetting related to the movie *The Shawshank Redemption* (Dir. Darabont, 1994).

## 2.7 Case Study

### 2.7.1 A case study of the Shawshank Trail

*The Shawshank Redemption* is a 1994 American drama film written and directed by Frank

Darabont. It is an adaptation of the novella “Rita Hayworth and Shawshank Redemption” by Stephen King (1982). It tells the story of Andy Dufresne (Tim Robbins), a young and successful banker who is wrongfully convicted and sentenced to two consecutive life terms for the murder of his wife and of her lover. The film is set in the 1940s, and it shows how Andy, with the help of his friend and fellow prison entrepreneur Red (Morgan Freeman), successfully escapes from the prison, and never let prison crush his spirit.

This film appears to be a good example for studying set-jetting for two main reasons. First, it is a very popular movie that has the potential to attract a lot of set-jettiers; it is ranked first in the IMDB (Internet Movie Data Base) Top 250 Movies (IMDB Charts, 2014). Second, it takes place in a very unique location: the Ohio State Reformatory in Mansfield, USA (see Figure 4). Indeed, to study set-jetting, it is important to take into consideration places that are clearly identified but that are not associated to existing tourist destinations. This is important to ensure that the touristic activity under study is directly related to set-jetting and not to a range of other touristic attractions. In the case of the Ohio State Reformatory in Mansfield, it is unlikely that tourists would visit this place for other reasons than because of the movie.



**Figure 4.** Scene in *The Shawshank Redemption*

(source: <http://www.npr.org/2011/08/04/138986482/on-location-mansfield-ohios-Shawshank-industry>)

(Note: Figure 4 shows a scene from *The Shawshank Redemption*, starring Tim Robbins (right) and Morgan Freeman (left), which was shot at the abandoned Mansfield Reformatory. Today, the buildings in the background are torn down, but local authorities have preserved the prison's main building (“On location: Mansfield”, 2011))

### **2.7.2 The Ohio State Reformatory**

In opposition to many Hollywood films, *The Shawshank Redemption* was not shot in a film-making studio in Los Angeles, New York, or Toronto. The director chose a unique place – a real prison - as the shooting location, which might have contributed to the incredible success of the movie. The realism of the landscape gives the viewer a deep sense of internment. That said, the shooting location was different than the story location.

While the movie was shot in Ohio, in the Ohio State Reformatory, the action was supposed to take place in a reformatory located in Portland, Maine. The Ohio State Reformatory was then used as the shooting location for the Shawshank State Penitentiary in rural Maine.



According to Wikipedia, The Ohio State Reformatory, also known as the Mansfield Reformatory, is a historic prison located in Mansfield, Ohio, in the United States. It was built at the end of the 19<sup>th</sup> century and remained in operation until 1990. This prison has been used in a lot of films, TV shows and music videos since its fame as the Shawshank State Prison in the film *The Shawshank Redemption* in 1994 (“Mansfield reformatory,” 2014). Although visitors come to this site for specific events such the a huge Halloween festival, extreme ghost hunts and murder mystery dinners, visitors come more specifically to visit the site of the 1993 filming of the extremely popular film *The Shawshank Redemption*. These set-jettors have an important economic impact on this area. According to the local newspaper *Post-Gazette*, in 2013, they have brought about \$10 million in tourism to the Mansfield area (“Former prison in”, 2014).

### **2.7.3 The cinematic landscape of *The Shawshank Redemption***

*The Shawshank Redemption* has turned what is often considered a dirty, creepy and marginalized place (Figure 5) into a popular tourist destination. Although the prison looks gorgeous with its three architectural styles –Richardsonian Romanesque, Victorian Gothic and Queen Anne (see Figure 6) – without the release of *The Shawshank Redemption* it would have remained just another prison and may have even been demolished. The use of the Mansfield Reformatory as a shooting set has transformed a real place into something else: a hybrid place made of real elements (the Mansfield Reformatory) and a fictional one (Shawshank Reformatory) that became a tourist attraction as well as a perfect case study for the potential of social media to influence set-jetting.



**Figure 5.** The west cell block of the Ohio State Reformatory

(Source: <http://www.npr.org/2011/08/04/138986482/on-location-mansfield-ohios-Shawshank-industry>)

(Note: For the movie, an exact replica of this block was built (mainly of wood) at an old warehouse of Westinghouse in Mansfield. This is because the wooden set was easier for the filmmakers to manipulate. Now it opens for visits and attracts people continually (“On location: Mansfield”,2011.)



**Figure 6.** The overall view of Ohio State Reformatory

(Source: <http://www.npr.org/2011/08/04/138986482/on-location-mansfield-ohios-Shawshank-industry>)

(Note: Figure 6 shows a view of the Mansfield Reformatory, where the film *The Shawshank*

*Redemption* (1994) was shot. However, the buildings, which “once formed the perimeter of the Reformatory “— and which can be seen in the film — has been demolished, but the Reformatory and its massive east and west cell blocks are still remained for visiting (“On location: Mansfield”, 2011)

## Chapter 3: Methodology

In this chapter, the research design process, and the different technologies and theories mobilized to collect and analyze the data are presented and discussed. This chapter starts with a description of the different social media applications selected in this project followed by a presentation of the data mining methods mobilized. This overall research design process is synthesized in Figure 7.

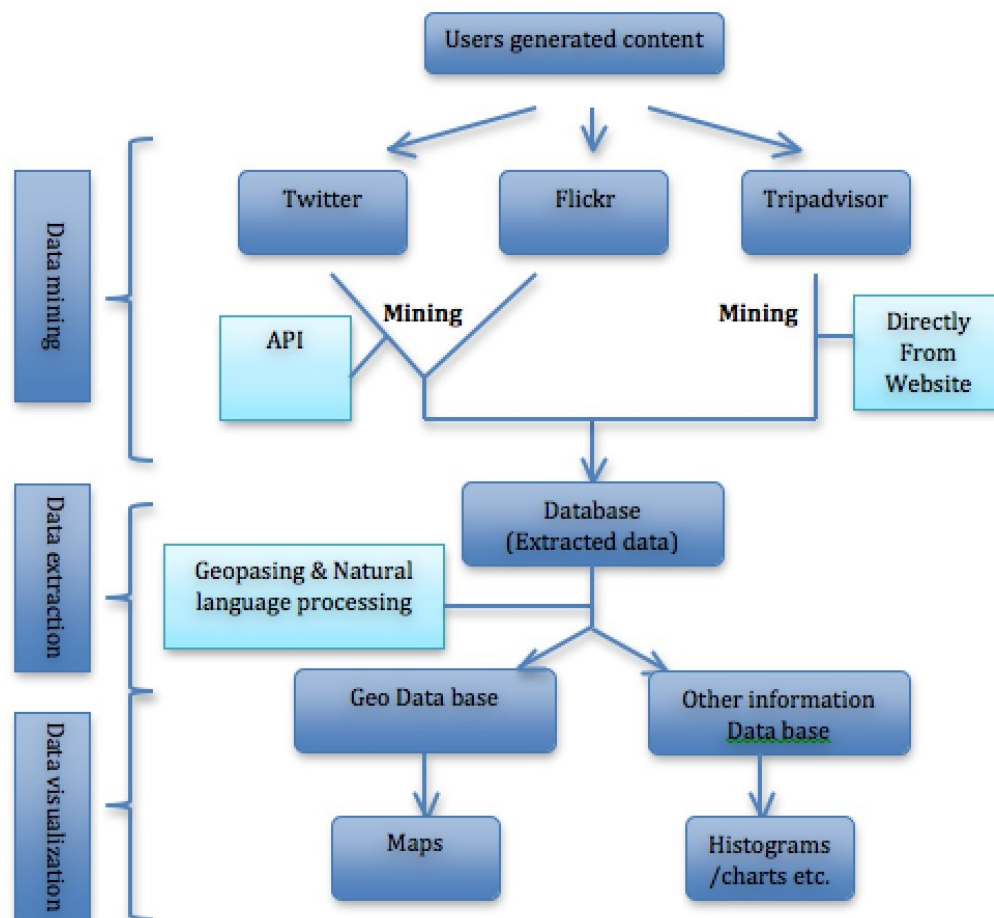


Figure 7. Research Design Process

### 3.1 The social media application

In this project, three social media websites have been selected: Twitter, Flickr, and Tripadvisor. This selection has been made for different reasons.

**Twitter** is a social networking website created in 2006 in which users are able to post, read, and share short messages (140-character maximum). As previously mentioned, Twitter is one of the most popular websites with more than 500 million users, including 284 million active users as of March 2014 (Twitter Help, 2014).

I have selected Twitter for the following reasons:

Firstly, it is available and used globally since 80% of the accounts are located outside United States (Twitter Help, 2014). Although the Shawshank Reformatory is located in the US, I am interested to see the global extent of set-jetting for this movie. Twitter might be useful for this.

Secondly, tweets can be geolocated in two different ways. First, by enabling Twitter to geocode the location from where a tweet is posted using latitudes and longitudes. Second, by parsing place names that appear in the body of the text of a tweet (see figure 8a & 8b).

Thirdly, the Twitter API can be used to mine not only tweets, but also the Twitter account users. These capabilities have been used for instance by Zook and Graham (2012) to develop the Crisis Map of the Haitian earthquake. For my project, mining data from Twitter will allow me to study how set-jettors use Twitter to share impressions and information about their trips.



**Figure 8.a.** example of a geotagged tweet with a place name appearing in the body of the text from Twitter (Note: when clicked, the short link 4sq.com/19lqtl shows the specific coordinates here, Source: <https://Twitter.com/>)



**Figure 8.b.** example of a geotagged tweet with an embedded picture of the Ohio State Reformatory from Twitter (Source: <https://Twitter.com/>)

**Flickr** is an image hosting website created in 2004, and currently owned by Yahoo. It enables users to upload and share their photographs, and it also offers an online forum, which enables users to interact with each other.

For my project, Flickr presents the following different assets:

Firstly, it is the most popular image sharing website, with currently 92 million users, spread across 63 countries as of Flickr Report (Flickr Fact Sheet, 2014). Everyday there are 1 million

photos shared (Techcrunch, 2014).

Secondly, the Flickr photos can be shared via other social media services such as Twitter, Facebook and Tripadvisor.

Thirdly, as discussed in the literature review, Flickr enables users to geotag their photos. In addition, there is a mapping interface embedded in Flickr that allows users to easily mine photographs based on location. Just like Twitter, Flickr also provides two ways to geolocate photos either by retrieving the geographic coordinates of geotagged photos or by using place names used to tag photos.

Fourthly, Flickr provides the possibility for developers to easily mine photos and associated information from its database, such as the serial number of the camera, the date and time the picture was taken and the information about the owner of the Flickr accounts. In other words, Flickr appears to be potentially a great source of information to study where people have been traveling within and around the Shawshank Reformatory.

**TripAdvisor** is a website designed for travellers. It provides the information and reviews about travel-related content by travellers and consumers. It was founded in February 2000 and is a typical website of user-generated content. Users of this website can post reviews of attractions, rate hotels and restaurants, and interact with other travellers. It is free for the users because it is supported by advertising revenues.

I have selected Tripadvisor for the following different reasons:

Firstly, it is one of the most popular social media websites related to travel (Xiang & Gretzel,

2009). On Tripadvisor, there are more than 190 million reviews and opinions from travellers around the world, and its database contains more than 4.4 million businesses and properties in 145,000 destinations around the world (TripAdvisor Fact Sheet, 2014).

Secondly, it is available in 45 countries and in 28 languages.

Thirdly, Tripadvisor provides sample information related to the sites described: reviews made by visitors, nearby restaurants, rating and ranking number, and maps of the destinations. For instance, as illustrated in Figure 9, the Ohio State Reformatory (Mansfield Reformatory) ranks #1 of overall activities in Mansfield, Ohio, based on 181 reviews about this site.



# Ohio State Reformatory

100 Reformatory Road, Mansfield, OH 44906-8625  
419-522-2644 [Improve this listing](#)

The screenshot displays the TripAdvisor page for the Ohio State Reformatory. At the top, the title "Ohio State Reformatory" is followed by the address "100 Reformatory Road, Mansfield, OH 44906-8625" and the phone number "419-522-2644". A "Save" button is visible. Below the title is a large photo of the reformatory building, with a "213 visitor photos" link underneath. To the right of the photo, a green badge indicates a "93%" rating and "Ranked #1 of 8 Activities in Mansfield". Below this, it lists "Certificate of Excellence 2014" and "181 Reviews". The "Type" is "Historic Sites" and "Activities" include "Group tours/walking tour". A "More attraction details" link is also present. To the right of the main content, a section titled "Hotels near Ohio State Reformatory" lists three options: "Holiday Inn Mansfield Conference Center" (88 Reviews, 3.1 km), "Merit Inn" (5 Reviews, 3.2 km), and "Quality Inn & Suites" (49 Reviews, 4.3 km). Each hotel listing includes a "Show Prices" button. Below the main content, a section titled "181 reviews from our community" features a "Write a Review" button and a "Visitor rating" bar chart. The chart shows: Excellent (123), Very good (46), Average (9), Poor (2), and Terrible (1). Below the chart, reviews are sorted by "Date" and "Rating", with a language filter set to "English first". A prominent review from a "TripAdvisor Member" is displayed, titled "Found Mansfield Reformatory to be very interesting" (4 stars, reviewed 28 July 2004). The review text describes the reformatory as a beautiful, castle-like structure with a lot of history, but notes that it has been vacant for fourteen years and is in disrepair. It mentions that volunteers are working to restore it but are short on funding, and that tours will take visitors through the largest iron... Below the review, there is a "More" link and a "Was this review helpful?" section with "Yes" and "13" responses. To the right of the reviews, a "Viewed attractions" section shows the Ohio State Reformatory with 181 reviews. Below that, a "Browse nearby" section includes a map of the area around the reformatory, showing "Richland Correctional Institution" and "Brace Ave". A "Which Mansfield hotels are on sale?" section is also visible, with date pickers and a "See hotels" button.

Figure 9. Ohio State Reformatory webpage on Tripadvisor

(Source: [http://www.tripadvisor.ca/Attraction\\_Review-g50599-d260123-Reviews-or40-Ohio\\_State\\_Reformatory-Mansfield\\_Ohio.html#REVIEWS](http://www.tripadvisor.ca/Attraction_Review-g50599-d260123-Reviews-or40-Ohio_State_Reformatory-Mansfield_Ohio.html#REVIEWS))

Overall, Tripadvisor will be used in my project to study how people describe the Ohio State Reformatory in their reviews and to compare these descriptions with the way this place is presented in Twitter and Flickr.

### **3.2 Data mining methods**

The data has been mined consistently from each of these social media applications to produce a geographic overview of set-jetting activities associated with these places, as well as to compare the use of these different sites for set-jetting purposes and touristic comments in more general terms.

The mined data includes the text from the tweets, the descriptions of photos from Flickr, and the reviews from Tripadvisor. It also includes the geolocation of tweets and photos as well as the origins and profiles of the selected contributors from these three websites. Beyond this general approach, specific mining methods have been developed for each website. For Twitter and Flickr, several relevant queries have been written to extract the data. The exact same queries have been applied to both tools in order to make the results comparable. For TripAdvisor, in addition to the reviews written about the Ohio State Reformatory, all the references to this place have also been mined as well as the profiles of the users who have generated these comments.

Now that I have introduced these general methodological considerations, I can discuss them in more detail. To mine data from Twitter and Flickr, I used a selection of key words.

First, the keyword “Shawshank” was used to query the content from both sources. This was done to compare the percentage of geotagged tweets and photos containing this keyword as well as to study their spatial distribution.

Secondly, a set of keywords were used during a second round of queries: “Shawshank jail OR Shawshank prison” and “Mansfield Reformatory OR Ohio State Reformatory”. The first set of keywords (Shawshank jail or Shawshank prison) captures mainly comments about the fictional place used in the movie, while the second capture mainly comments about the real place. These 2

sets of keywords were used to identify connections between the fictional and the real corresponding places.

Thirdly, the geo-query “40.78764,-82.50286, radius: 1km” was used with both Twitter and Flickr to capture comments and images that were expressed and taken at a distance of less than 1km from the Mansfield reformatory.

### **Flickr**

Flickr has an open Application Programming Interface (API). People can use it easily to mine data made public by their users, such as photos, author information, tags, and geolocations. Many tools have been developed to mine these data. Most of them require an authenticated function. Flickr uses the OAuth standards, which enables applications to authenticate users and interact with Flickr on their behalf. The OAuth standard is an industry standard, which provides a secure way for people to sign-in into their accounts with different kinds of websites that Flickr supports. Flickr’s OAuth flows work for web applications, desktop applications and mobile apps as well (Flickr, 2014).

The Flickr API works in two essential steps. First, the user sends a request with the detailed information about what s/he would like to do. This request can be made in different formats (e.g. REST, XML-RPC, or SOAP format), and it is realized by building an URL. Once Flickr receives the request, it returns a response to the user. The format of this response can be REST, XML-RPC, SOAP, JSON or PHP. After receiving the data included in the response the user can extract the relevant data. For instance, in this project, the data was mined from Flickr using the XML (REST) format. This format was chosen because the XML(REST) is simpler, clearer, runs faster, and uses less bandwidth.

There is a tool called API Explorer provided by Flickr that can help users with limited programming knowledge to play around with the API method and get some data back. In this project, I used the Flickr.photos.search methods, with a radial geo-query to search the Flickr database. The geo-query was set at the center of the Mansfield Reformatory with a 1km radius to ensure that a broad area was covered in the search. This URL of this request was as follows:

[http://api.Flickr.com/services/rest/?method=Flickr.photos.search&api\\_key=15fef5c1b147e48187b6257930a45bc2&min\\_upload\\_date=01%2F01%2F2006&min\\_taken\\_date=01%2F01%2F2006&accuracy=11&lat=40.782393&lon=-82.503354&radius=1&extras=geo%2Ctags%2Cdate\\_taken&per\\_page=500&format=rest&auth\\_token=72157636586988785-96e99945399320a2&api\\_sig=8dc48381ac5d04cfe4ec1f654a74a6cd](http://api.Flickr.com/services/rest/?method=Flickr.photos.search&api_key=15fef5c1b147e48187b6257930a45bc2&min_upload_date=01%2F01%2F2006&min_taken_date=01%2F01%2F2006&accuracy=11&lat=40.782393&lon=-82.503354&radius=1&extras=geo%2Ctags%2Cdate_taken&per_page=500&format=rest&auth_token=72157636586988785-96e99945399320a2&api_sig=8dc48381ac5d04cfe4ec1f654a74a6cd)

The response was returned quickly and included 1684 relevant results which were structured as follows:

```
<photo id="6284192490" owner="65231003@N06" title="Ohio State Reformatory"
latitude="40.783287" longitude="-82.504277" accuracy="14" context="0"
place_id="6etHccBTVr5vvHFG" woeid="2444995" tags="old ohio storm cold abandoned clouds
dark cleveland rusty haunted creepy falling prison jail oh ghosts mold stale redemption apart
Shawshank Reformatory" />
```

**Table 2.** Description of each of the Search API returned parameters (Flickr API documentation, 2013c).

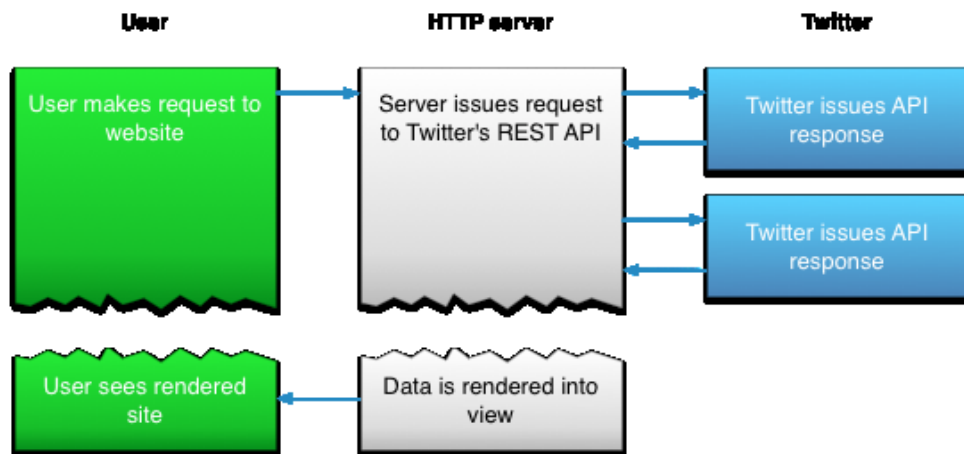
Photo_id	The ID of the photo whose WOE (Where On Earth) location is being corrected.
owner	The NSID (User ID) of the user whose photo to search. If this parameter isn't available, then everybody's public photos that match the search requirements will be searched.
title	The title of the photo.
tags	A comma-delimited list of tags. All of the tags added by user will be returned. You can include several tags separated by “,”, or exclude results that match a term by prepending it with a “-“ character.
Place_id	A Flickr Places ID (While optional, you must give either a valid Places ID or a WOE ID).
woeid	A Where On Earth (WOE) ID (While optional, you must give either a valid Places ID or a WOE ID).
latitude	The valid range of latitude is from -90 degree to 90 degree. The maximum decimal is 6
longitude	The valid range of longitude is from -180 degree to 180 degree. The maximum decimal is 6.
accuracy	Recorded accuracy level of the location information. The range is from 1 to 16. World level is 1, Country is ~3, Region ~6, City ~11, Street ~16. Default value is 16
context	Context is a numeric value, and it representing the photo's specific geolocation beyond the coordinates.  The current context IDs list is: <b>0</b> not defined; <b>1</b> indoor; <b>2</b> outdoors.

## Twitter

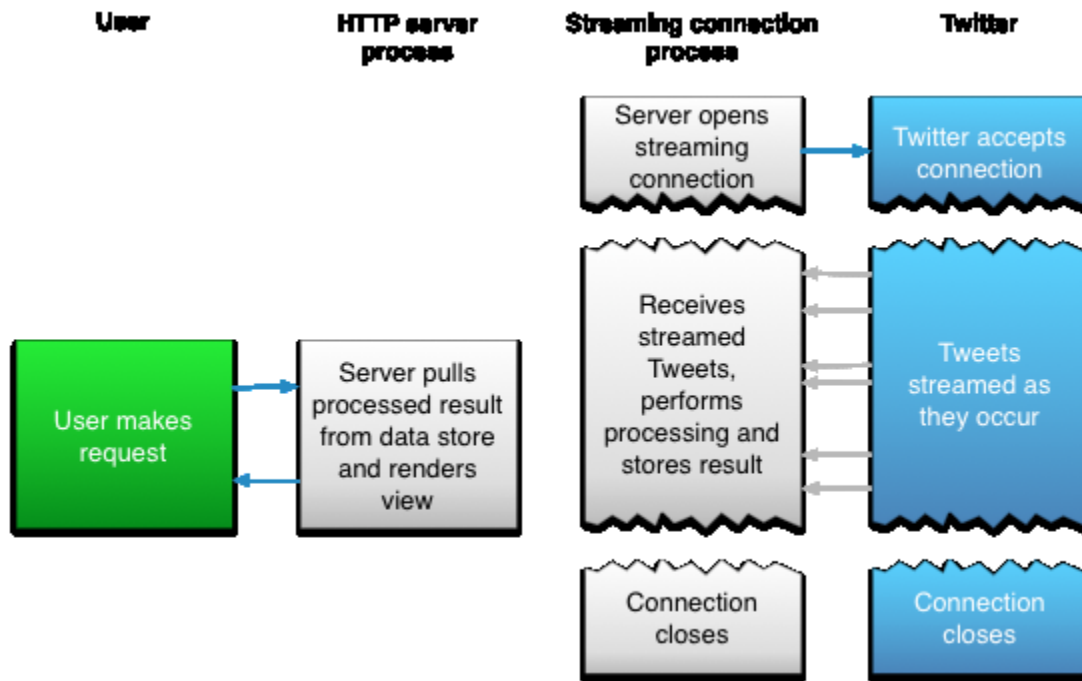
The timeframe for mining the data from Twitter was between September 1<sup>st</sup>, 2013 and March 1<sup>st</sup>, 2014 (6 months). Two kinds of tweets were collected: geotagged tweets in one specific place and geotagged tweets that referred to that specific place (e.g. tweets including the keyword “Mansfield Reformatory”). I then collected geotagged tweets that were posted near the

Mansfield Reformatory as well as those non-geotagged tweets mentioning the Mansfield Reformatory. In addition, I tracked the location of a selection of users mentioning their visit to the Mansfield Reformatory in their tweets.

For the purpose of this research, I used the Search API and Streaming API. Similar to Flickr, Twitter has its own open API. The Twitter Search API is part of Twitter’s v1.1 REST API. This REST API provides simple interfaces for most Twitter functionalities. The search API works very similarly as the Flickr one (Figure 10). The Streaming API works slightly differently from the Search API. It does not establish a connection in response to a user request. Instead, the connection and the HTTP requests are separated (Figure 11).



**Figure 10.** REST request (source: Twitter Developer, 2013c)



**Figure 11.** Streaming API request (source: Twitter Develop, 2013c)

The search API returns a collection of relevant tweets matching a specified query (Twitter Developer, 2013). The method I used here to mine tweets is named GET search/tweets. There are two main methods for extracting tweets. The easiest one is by using Twitter Explorer console. This console allows users with limited programming knowledge to test simple search requests. The use of this console app requires granting access to the Twitter database when you make a Twitter API request. The user is then directly authorized to access Twitter through recognizing authentication.

As mentioned before, after choosing the “search” method, the console requires the user to provide a query. The query I used here is q = “Mansfield Reformatory,” q = “Ohio State

Reformatory” or q= “Shawshank,” and I set the geocode parameter as “40.782393,-82.503354, 1km.” More details about these parameters can be seen in Table 3.

**Table 3:** Explanation of Search API parameters (Twitter API documentation, 2013c)

q	This value is usually the keyword that people prefer to search. It is a “UTF-8, URL-encoded search query of 1,000 characters maximum, including operators”. Queries may additionally be limited by complexity.
geocode	<p>“Returns tweets by users located within a given radius of the given latitude/longitude. The location is preferentially taken from the Geotagging API”.</p> <p>The parameter value used here is "latitude,longitude,radius", and the radius units here must be specified as either miles (mi) or kilometers (km).</p>
lang	Restricts tweets to the given language, e.g. "en" representing English, which is given by an ISO 639-1 code.
result_type	<p>The type is Optional. Specifies what type of search results people would prefer to receive. The default value is "mixed." Valid values include:</p> <p>“-mixed: Include both popular and real time results in the response.</p> <p>-recent: return only the most recent results in the response</p> <p>-popular: return only the most popular results in the response.”</p>

**Appendix II** shows one example of the results returned from the Twitter API. From the results, we can extract the user name, time, content, and geo-information of this tweet. However, this is not the best way for users to structure the extracted results. To structure the data, I used a related



Python library named Tweepy. Tweepy is a great library for using Twitter API. Through this library, I can realize searching and streaming Twitter simultaneously. Additionally, the returned results allows for more customization. The steps to use Tweepy is as follows: Installing Python (above version 2.7.0), setting the system environment, installing the Tweepy package (above version 2.0.0), and then the Tweepy library can be run through Python. In this case, the search function from search.api and the filter function from streaming.api are used. The parameters I used here (see Table 3) are the same as I used through the Twitter console. **Appendix III** shows the codes I programmed for searching by keywords and specific location through the Tweepy library, and **Appendix I** shows the results returned by using this method.

Currently, the search API does not provide access to all tweets, but only to tweets published recently between 6-9 days (Twitter Developers, 2013). To alleviate this problem, I have mined Twitter data once a week for 27 weeks. Appendix I shows the searched results of tweets mined during one week (from November 3rd, 2013 to November 9th, 2013).

### **Tripadvisor**

Several key words have been tested to mine as much relevant data as possible related to the Ohio State Reformatory. More specifically, I used the following sets of words “Mansfield Reformatory,” ”Ohio State Reformatory,” and “Shawshank Reformatory,” which returned a total of 260 reviews. Among these 260 reviews, 193 reviews directly refer to the Mansfield Reformatory. These 193 reviews were copied and pasted into a database that includes the places visited, the content of the reviews, the visit time, as well as the rate given to Mansfield Reformatory on a scale from 1 to 5. In addition, the profiles of selected visitors were also recorded, including their personal information (e.g. origins, gender, and date of birth).

### 3.3 Data Extraction methods

Once these data were extracted from these three websites, it was time to filter them in order to extract some meaningful information. I used several methods to do so.

Three types of geospatial information were extracted from these three databases: geospatial data based (1) on x and y coordinates; (2) on place names; and (3) on the location of the owner of the accounts under study.


(1) X and y coordinates were extracted from geolocated entries from Flickr and Twitter coordinates.

(2) Place names were identified using the GeoName extraction module. The GeoName extraction module is based on an integration of applying linguistic heuristics and gazetteer queries. The working principle of this module is that it can mine the lexical context such as preposition words (e.g. “in”, “at”, “near”) and spatial relationship words (e.g. “south of”, “left”, “northwest of”). These kinds of words are used to provide clues for location names (Shi & Barker, 2011). GeoDoc and CLIFF are two example applications of this module.

GeoDoc is a geotagging tool. It helps to automatically identify place names in texts and anchors each name to the place in the maps. It offers a web interface that allows the users to highlight the place names quickly and correctly, and can export the geotagged text as XML OR HTML formats (GeoDoc Guide, 2014). CLIFF is an application used to parse news articles and to identify places mentioned as well as people, organizations, etc. It has been developed to “focus on getting at what place an article is really about (as opposed to all the places it mentions).” In this case, I used it to pull out the long-length reviews from Tripadvisor. It will automatically generate a separate result (CLIFF, 2014).

(3) the information about the location of account owners was also obtained. For Twitter, the selected user’s information is obtained from the code (for example see Appendix I) through the previously mentioned geo-extraction application. For Flickr, this data is extracted directly with the Flickr API. For Tripadvisor, all of the users’ information has been copied and pasted.

**Table 4.** Extracted results from the Twitter, Flickr and Tripadvisor

<i>Source</i>	<i>Content</i>	<i>Locations</i>	<i>Coordinates</i>
<i>Twitter</i>	<p><b>Mike</b> @L_Am_PsychMike · 11月1日            @Zak_Bagans @AaronGoodwin @NickGroff_ at the <b>Ohio State Reformatory</b>.            This place is giving me the creeps</p>	<p><i>Ohio State</i>  <i>Reformatory</i></p>	<p>40.7853 ,            -82.5050</p>
<i>Flickr</i>	<p><b>Tom Bower</b>  <b>Dueling Nikons.....</b>            My new friend Nick and I were having a bit of flash fun shooting each other.            ...at the Mansfield Reformatory, Mansfield, Ohio USA</p>	<p><i>Ohio State</i>  <i>Reformatory</i></p>	<p>40.7823,            -82.5034</p>
<i>Tripadvisor</i>	<p><b>“Order the Shawshankwich!”</b>   Reviewed 14 July 2012            After our tour of the Ohio State Reformatory, we made a few stops along the Shawshank Trail including this cafe. It was a clean and inviting place that offered cafeteria styled selections and enjoyed the shawshankwich and bag special. After lunch, run across the street for some sweets at the Squirrel's Den.</p>	<p><i>Ohio State</i>  <i>Reformatory;</i>  <i>Squirrel’s</i>  <i>Den</i></p>	<p>40.7853 ,            -82.5050;            40.7593,            -82.5160</p>

The results acquired through the process of extracting the geo-information from the different content can then be mapped and analyzed (see Table 4)

In the following section, I will introduce, analyze, present and discuss the data retrieved and

extracted from the collecting process. More details about the challenges of the collecting process will be discussed further.

## **Chapter 4: Data analysis and Discussion**

This chapter presents the main results from this study. The data collected from the three websites are analyzed and compared in different ways in order to answer the research questions raised in chapter one.

### **4.1 Overall data collection**

Although I tried to be as consistent as possible in terms of data collection, there are still some discrepancies between the data collected from the three social media. In terms of timeframe, the data from Flickr and Tripadvisor have been collected since the first relevant record identified (2004 for Tripadvisor and 2009 for Flickr), while for Twitter, the data collection period covers only half a year (from October 1st 2013 to March 31st 2014). The following table (Table 5) provides an overview of the data collected.

**Table 5.** Overall view of the mined data from Flickr, Twitter, Tripadvisor

Map	Twitter	Flickr	Tripadvisor	Scale	Goal of the comparison
<b>No maps (not enough data from Twitter)</b>	Geolocated tweets with keywords (Mansfield reformatory or Ohio State reformatory) <b>Geotagged tweets: 15</b> <b>Total tweets: 933</b>	Geolocated photo with keywords Mansfield reformatory or Ohio State reformatory <b>Geotagged photo: 2729</b> <b>Total photos: 9234</b>	/	local national global	Comparing how people talk about real places on social media
<b>Figure 12.a &amp; 12.b</b>	Geolocated tweets with keywords (Shawshank prison/jail) <b>Geotagged tweets: 1137</b> <b>Total tweets: 232,006</b>	Geolocated photos with keywords (Shawshank prison/jail) in two weeks <b>Geotagged photos: 911</b> <b>Total photos: 5132</b>	/	local national global	Comparing the distribution of how people talk about fictional places on social media
<b>Figure 14a &amp; 14b</b>	Geolocated tweets within 1km from Mansfield Reformatory <b>Total:143</b>	Geotagged photos within 1km from Mansfield Reformatory <b>Total: 1684</b>	/	local	Comparing the spatial distribution of how people use social media near Mansfield Reformatory
<b>Figure 18a &amp; 18b</b>	Twitter users with completed information <b>Number: 100</b> <b>Total: 143</b>	Flickr users with completed information <b>Number: 93</b> <b>Total: 134</b>	Tripadvisor users with completed information <b>Number: 128</b> <b>Total: 146</b>	local national global	Comparing the origins of the visitors of Mansfield Reformatory
<b>Figure 19 a,19b &amp;19c</b>	10 users' tracking routes from Twitter	10 users' tracking routes from Flickr	10 users' tracking routes from Tripadvisor	local national global	Studying the potential movements of users of each application

All of the compared results are shown through a variety of methods. Besides the use of traditional GIS applications, such as Arcgis, I also used online mapping applications such as CartoDB to map and visualize the results, as well as atlascine to map and present dynamically the movements of set-jetting.

## **4.2 Spatial analysis**

In this section, I propose to study the spatial distribution of the mined data at different scales: from global to very local (i.e. at the scale of the Mansfield Reformatory).

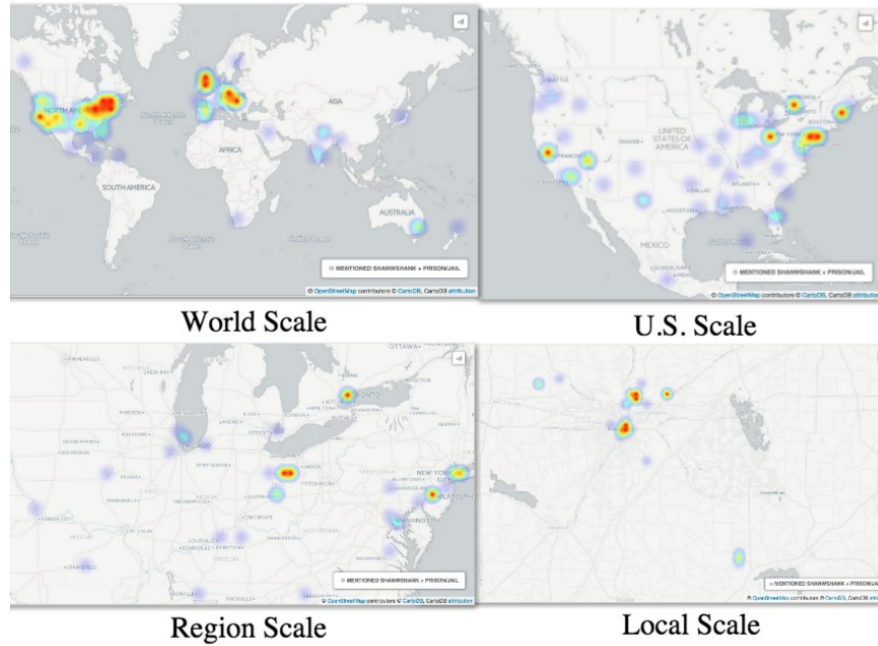
A first analysis involves comparing the overall spatial distribution of geotagged tweets from Twitter and photos from Flickr (Note: data from Tripadvisor were not integrated at this stage since the content from Tripadvisor is not geotagged). Among the 933 tweets containing the keywords “Mansfield Reformatory” or “Ohio State Reformatory,” only 15 tweets have been geotagged, which corresponds to 1.6% of the total number of tweets. Using the same keywords, I retrieved 9234 photos from Flickr, among which 2729 were geolocated (18.9% of the total, see Table 5).

To collect more tweets and Flickr photos, I also used “Shawshank + prison/jail” as keywords. There is a humongous amount of tweets generated on a daily basis with these keywords. In fact during a period of 2 weeks (Feb 2<sup>nd</sup> to Feb 17<sup>th</sup>, 2014), 232,006 tweets were published with these key words. In order to reduce the size of our database and to make Twitter data comparable to Flickr data, I have selected only the tweets from this two weeks period. Among these 232,006 tweets, only 0.5% (1137) has been geotagged. Meanwhile the number of photos in Flickr tagged with the same keywords for a period of 5 years (Jan 30<sup>th</sup>, 2009 to March 31<sup>st</sup>, 2014) is 5132, among which 19.4% (911) have been geotagged. These results confirm 2 things: First there are

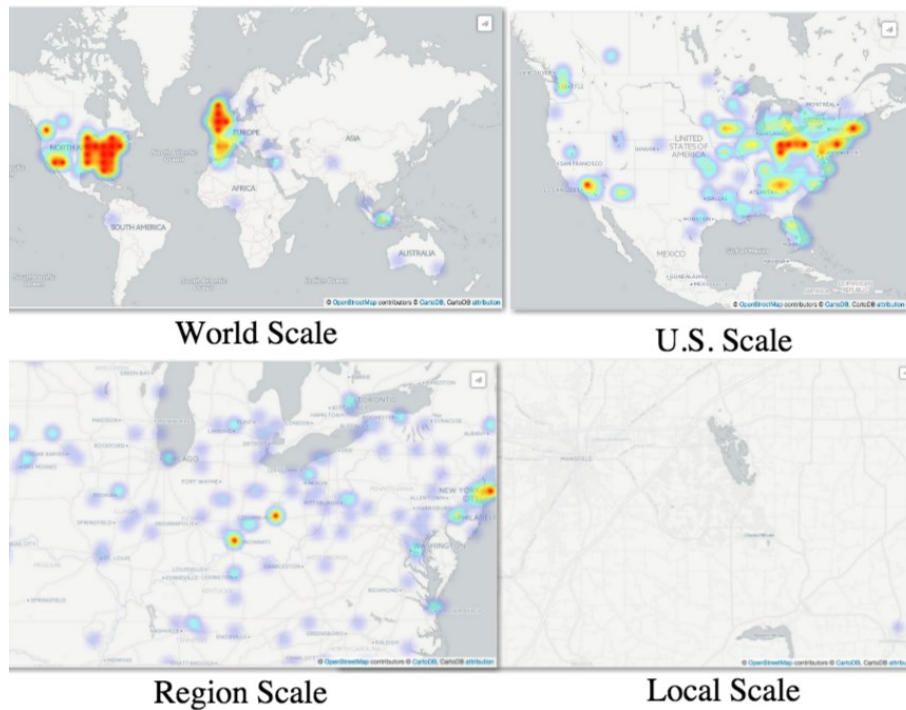
much more data generated via Twitter than via Flickr but the ratio of geotagged photos in Flickr is much higher than the ratio of geotagged tweets. These results also confirm that geotagged tweets represent about 0.5% of all the tweets (Twitter, 2014). Overall, this reveals that if geolocated photos could be considered as representative of all the photos available in Flickr, geolocated tweets are far from being representative of the content of the tweetsphere. Since they just represent a fraction of the tweets, the results of the analysis of these geotagged tweets can not be generalized to the entire tweetsphere. Furthermore, the data from Twitter for this part of the analysis were collected on a short period of 2 weeks, which is not representative of the overall year. Although these geolocated tweets are not necessarily representative of all the tweets referring to *The Shawshank Redemption*, they might help us to identify some spatial dynamics associated with the use of this media in the context of set-jetting.

What the results also emphasize is that photos from Flickr refer proportionally much more often to the real Ohio state reformatory (i.e. Mansfield Reformatory) than tweets. Twitter users often mention the fictional place (i.e. Shawshank/prison) but barely mention the real place. Twitter seems to be more intensively used to talk about fictional places, while Flickr seems to be more relevant to study real places, as we will see in more detail in this section.





**Figure 12.a.** The distribution of geotagged photos from Flickr containing “Shawshank + Prison/Jail” at four different scales (Mapped through CartoDB, online address: <http://cdb.io/1MqENGp>)



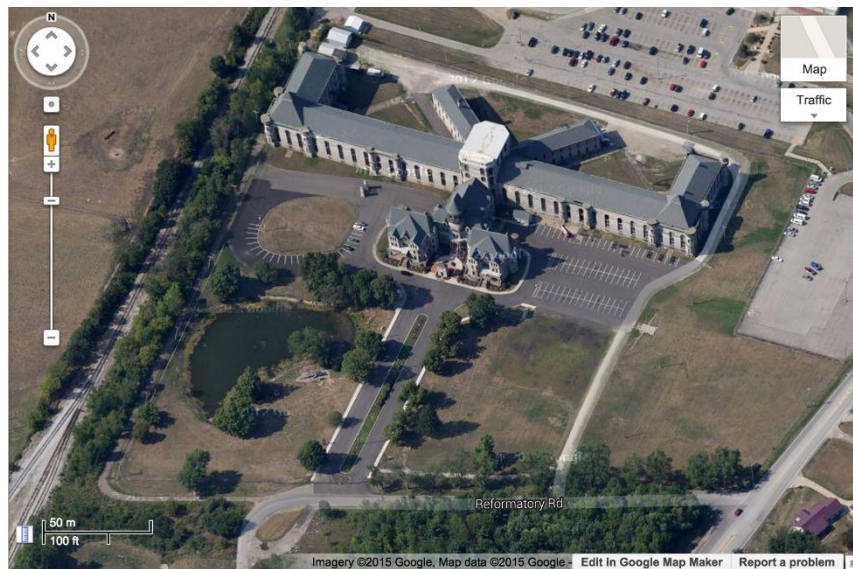
**Figure 12.b.** The distribution of geotagged tweets containing “Shawshank + Prison/Jail” at four different scales (Mapped through CartoDB, online address: <http://cdb.io/1MqCcMD>)

By comparing these 2 first sets of maps (Figure 12a &b), we initially see some similarities. Geotagged tweets and Flickr photos come mainly from Europe and the US and more specifically from the UK and the East coast of the US. What is a bit more surprising is that geotagged tweets are overwhelmingly posted from the UK (773 from a total of 1137). The reason for this overrepresentation of geotagged tweets from the UK can be explained by the fact that there is one famous British Twitter user who has a lot of followers and who wrote one tweet containing the key words “Shawshank + Prison”. This tweet has been retweeted about 600 times. The spatial distribution at regional scale in the US shows the different spatial structures between Twitter and Flickr. This emphasizes both the overwhelming influence of a few people on Twitter as well as the impact of retweeting on my data. Indeed, while most of the photos are concentrated near the Mansfield Reformatory, tweets appear to be much more randomly distributed and do not show any spatial pattern. Not surprisingly, photos (and Flickr) have a stronger connection with place than tweets (and Twitter).

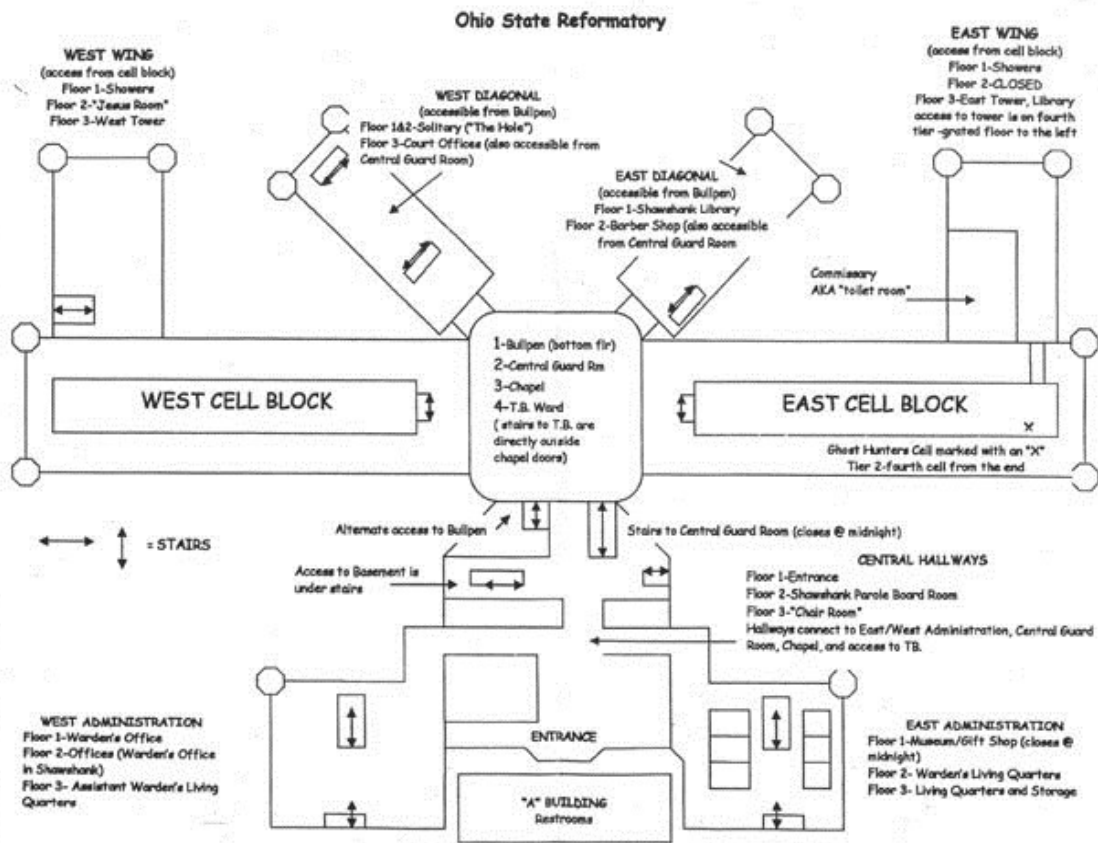
This first comparison between Flickr and Twitter illustrates the different geographic potentials of these two social media. While Twitter can be used from anywhere to talk about anything, Flickr is much more place-dependent, since photos are more likely related to places. While Twitter gives us a sense of the interest raised by an event (e.g a movie) in different parts of the world, geotagged photos from Flickr give us a better sense of local activities associated to this event. Although Twitter can be place related when associated to a specific event confined in time and place (e.g. a demonstration), both social media appear to be complementary for studying film-induced tourism.

The analysis of the content of the tweets reveals that among all of the tweets containing the key word “Shawshank + Prison/Jail” (i.e. referring to the fictional place), less than 1% also refer to the real place (i.e. Mansfield Reformatory or Ohio State Reformatory). On the other hand, a

much higher percentage of tweets (33%) referring to the real place (i.e. Ohio State Reformatory) refer at the same time to the fictional place (i.e. Shawshank prison). This means that the real place is much more often associated to the fictional place than the way around. This element also appears clearly in the Tripadvisor reviews. Indeed, 80% of the long reviews related to the reformatory also mention the movie *The Shawshank Redemption*. While the fictional place can exist independently from the real one, the real location is clearly connected to the fictional one. The way these two places appear in social media confirms the idea that the real Mansfield Reformatory exists mainly as a materialization of the fictional Shawshank prison.



**Figure 13.a** Satellite view of Mansfield reformatory (source: Google map)



**Figure 13.b** Inner map of Ohio State Reformatory  
 (Source: [http://www.hauntedhocking.com/Ohio\\_State\\_Reformatory.htm](http://www.hauntedhocking.com/Ohio_State_Reformatory.htm))

The Mansfield Reformatory currently includes a set of buildings surrounded by a park and two parking lots (see Figure 13a). The high level of accuracy of the geolocation of tweets and photos allows for a previous location of where these digital elements have been produced on the site as well as within the buildings (see Figure 13b). First, it is possible to identify hot spots of where pictures have been taken around the Mansfield Reformatory. The main hot spots appear to be inside the main building in the center of the Reformatory in the west cell block and in the west diagonal (see Figure 14a & 15a). While similar hot spots appear in the use of Twitter (see Figure 14b), secondary locations where pictures have been taken were also identified. For instance, there are two secondary hot spots outside the buildings. One is on the path in front of the main

entrance which provides an overview of the façade of the building (see Figure 15b), and the other one is at the intersection of the south Reformatory road and the Reformatory road outside the building that also provides a good overall view to the reformatory (see Figure 15c). Another point worth mentioning is that people would tweet from the parking lot located on the southeast side of the building but would not take pictures from there (see Figure 14a & 14b).

Again, we can see some differences in the spatial use of both social media and in their spatial reference. Here the use of Twitter on the side appears to be more spread out spatially (with a few hot spots) while the use of Flickr is much more structured around clusters determined by the shooting location towards the main building. Although some people tweet from a very specific location associated to a detail or to a picture, they mainly tweet from the site in general terms. The results related to the use of Twitter need to be taken with caution given the small percentage of geolocated tweets.



**Figure 14.a** Geolocated photos mined from Flickr near Mansfield Reformatory (From Jan 30<sup>th</sup>, 2009 to March 31<sup>st</sup>, 2014)



**Figure 14.b** Geolocated Tweets mined from Twitter near Mansfield Reformatory (From October 1<sup>st</sup>, 2013 to March 31<sup>st</sup> 2014)



**Figure 15 a.** West cell block, photo taken by David Perl  
 (Source: <https://www.flickr.com/photos/64712918@N06/7906487516>)



**Figure 15 b.** Entry, photo taken by Michael Savino  
(Source: <https://www.flickr.com/photos/101401721@N04/9688584467/>)



**Figure 15 c.** Overview, photo taken by Doug Butchy  
(Source: <https://www.flickr.com/photos/dougbutchy/7230572904/>)  
(Note: Figure 15a, b & c are examples of pictures taken from the different hot spots)

The geolocated data collected from these social media can also be useful to study the origin of the people visiting the Ohio State Reformatory. I studied the origin of the content providers based on Flickr, Twitter and Tripadvisor. I used Flickr API to extract the user ID and to acquire

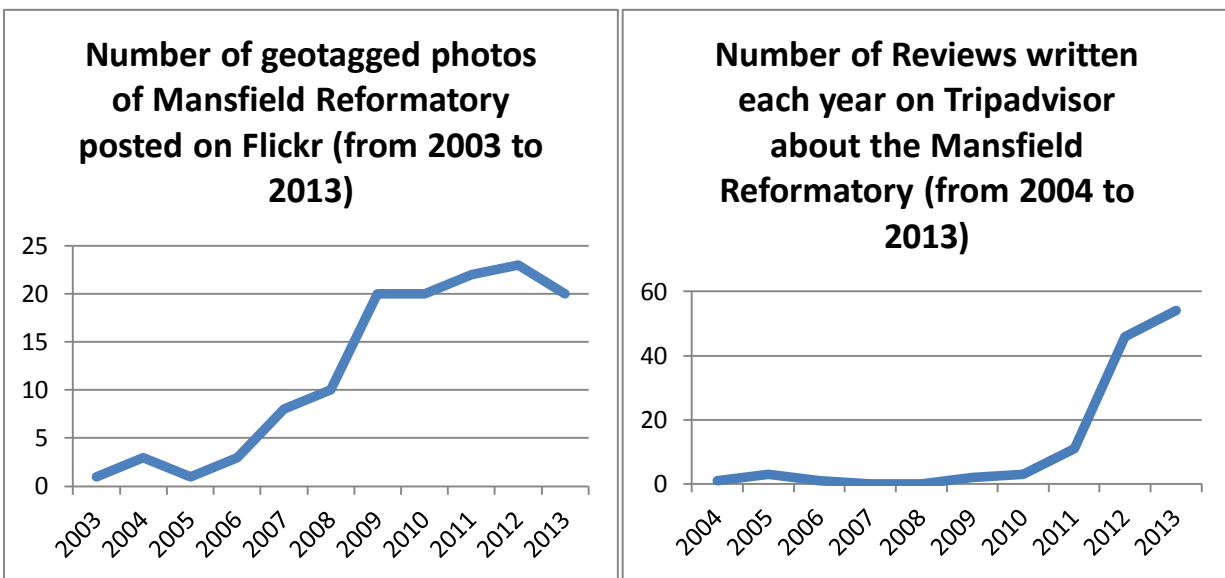
the personal information of the owners of the accounts used to upload the 1684 mined photos. I was able to retrieve 134 accounts from which 93 had complete information about the origins of the owners. From Tripadvisor and Twitter I extracted manually the origins of the users based on the profile account information. I retrieved 100 users from Twitter who posted geolocated tweets related to the reformatory and 128 users from Tripadvisor to which I had access to the location of the account owner as well.

In terms of the origins of the visitors, the results (Figure 18a. & 18c.) show some similarities and differences between different social media. Let's start with the similarities. First, most of the accounts used to post geotagged data are located in the U.S, especially in the north-eastern part of the country. Only a few identified accounts come from abroad: 8 from Europe (2 from Twitter, 3 from Flickr and 3 from Tripadvisor), 5 from Asia (1 from Twitter, 2 from Flickr and 2 from Tripadvisor), and 7 from other areas (2 from Twitter, 3 from Flickr, and 4 from Tripadvisor). Second, for all of the three social media, the main cluster for the accounts is in Ohio where the Mansfield reformatory is located. Most visitors come from the surrounding area of the Mansfield Reformatory (about 60% in Ohio state). The Mansfield Reformatory is firstly a Local/Regional tourist attraction.

In terms of the differences between the three social media, the origins of the Twitter accounts are more spread out than the origins of the Flickr and Tripadvisor accounts. This means that Twitter users overall come from further than the visitors using Flickr and Tripadvisor. This difference is captured by the average distances between the origins of visitors and the destination: 982 km for Twitter, 876 km for Flickr, and 718 km for Tripadvisor. This difference as well as the Local/Regional attractiveness of the site appears more clearly with the median distance: Twitter is 263 km, Flickr is 190 km, and the Tripadvisor is 185 km.

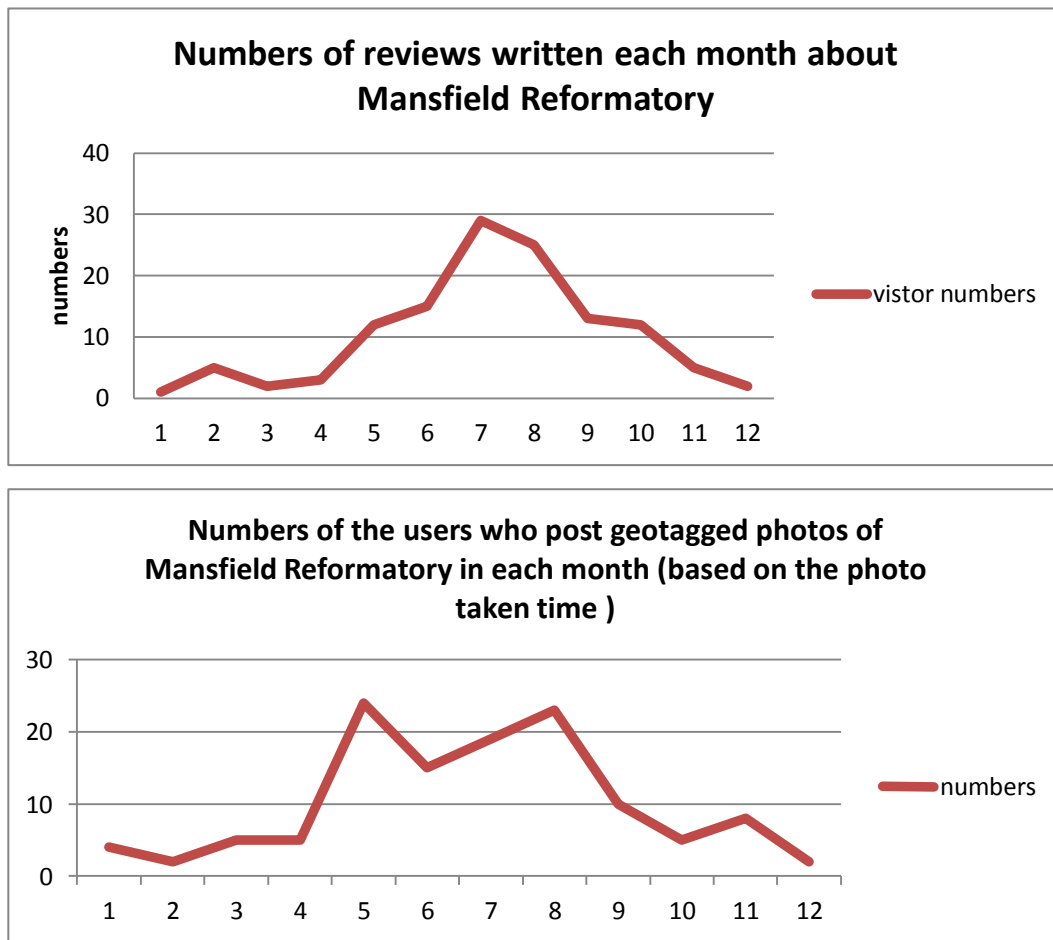


The number of Internet users who have been writing reviews on Tripadvisor has increased dramatically since 2011 (see Figure 16a and 16b). This can be explained by two exponential developments of Tripadvisor since 2011 (Tripadvisor, 2012). From 2011 to 2014 the number of registered users has increased from 20 to 70 million users, and the total number of reviews has also increased from 60 to 190 million. More and more people are willing to share their travelling experience via Tripadvisor. This increasing popularity characterizes the increasing popularity of social media in general. For instance between 2009 and 2013, the total number of uploaded photos on Flickr also increased steadily from 4 to 8 billion. The increasing number of reviews on Tripadvisor about the Mansfield Reformatory can also be explained by the operational organization of the Mansfield Reformatory, which has been actively promoting the site as a tourist destination and adding different thematic tours for visitors. Social media can be used to monitor the tourist activity, not only selectively, but also temporally.



**Figure 16a. & 16b.** Annual number of reviews about the Mansfield Reformatory posted on Tripadvisor (16a) and number of photos uploaded on Flickr (16b)

Based on the monthly use of these social media, the peak tourist activity is during the summer months, which is expected. When we compare the results obtained through the two websites we can see some discrepancies. Flickr shows two peaks (May and August) where Tripadvisor shows only one peak (in July, see Figure 17 a & b). The attractiveness of the reformatory during Halloween (end of October) appears to be a marginal phenomenon.



**Figure 17a. & 17b.** Monthly number of reviews about the Mansfield Reformatory posted in Tripadvisor (17a) and number of photos uploader of Flickr (17b)

Although more than 1,000 photos have been uploaded to Flickr, the photos taken near the Mansfield Reformatory are contributed by only 93 people, which characterizes the fact that 90%

of Flickr users upload more than one photo. This is the opposite for Twitter. 90% of Twitter users posted only one tweet near the Mansfield Reformatory. Since people tweet less often than they take pictures, it is difficult to follow their movements at a very fine scale around the site using Twitter. Flickr appears to be much more relevant to track people's movements at the site level.

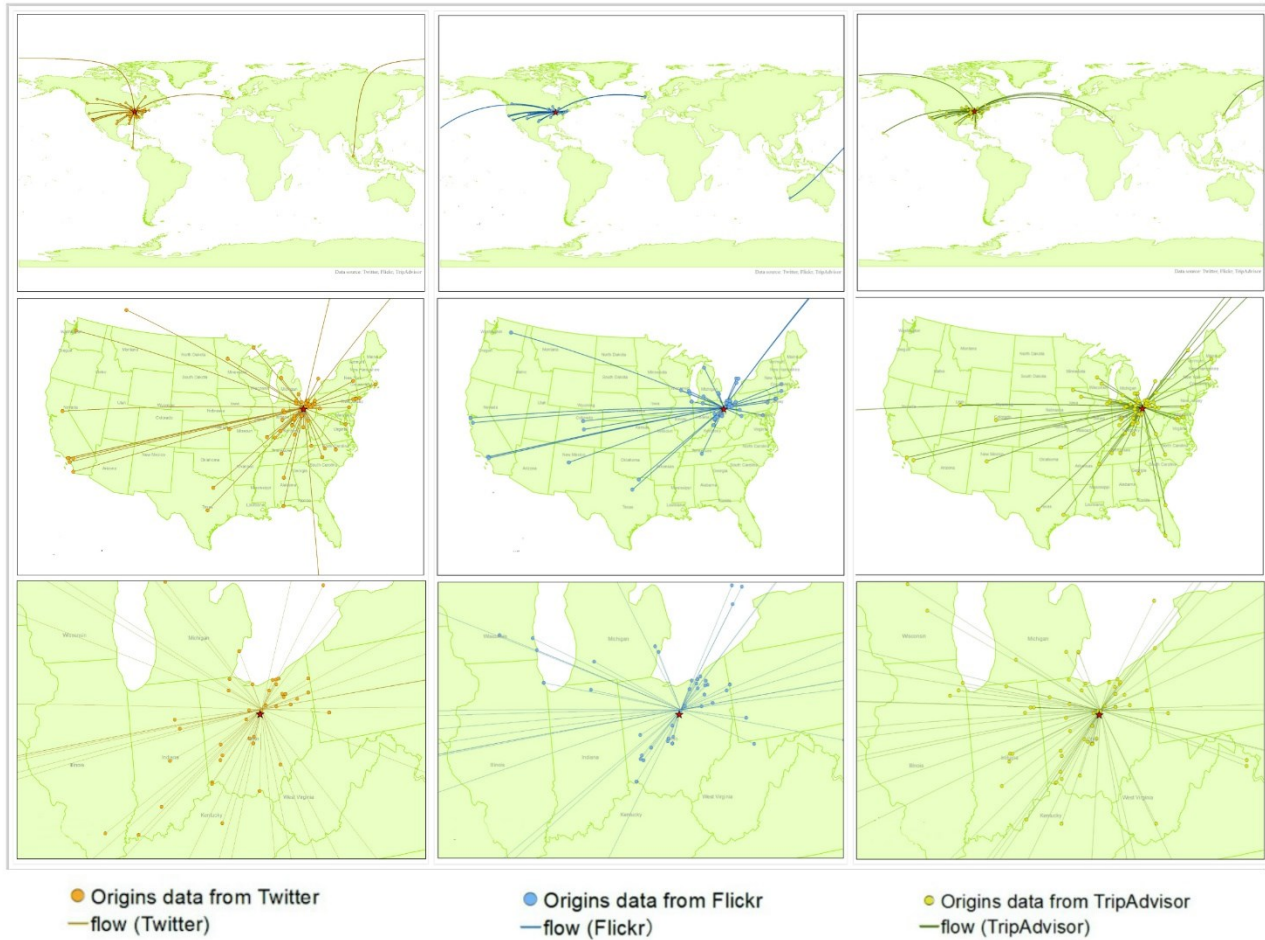
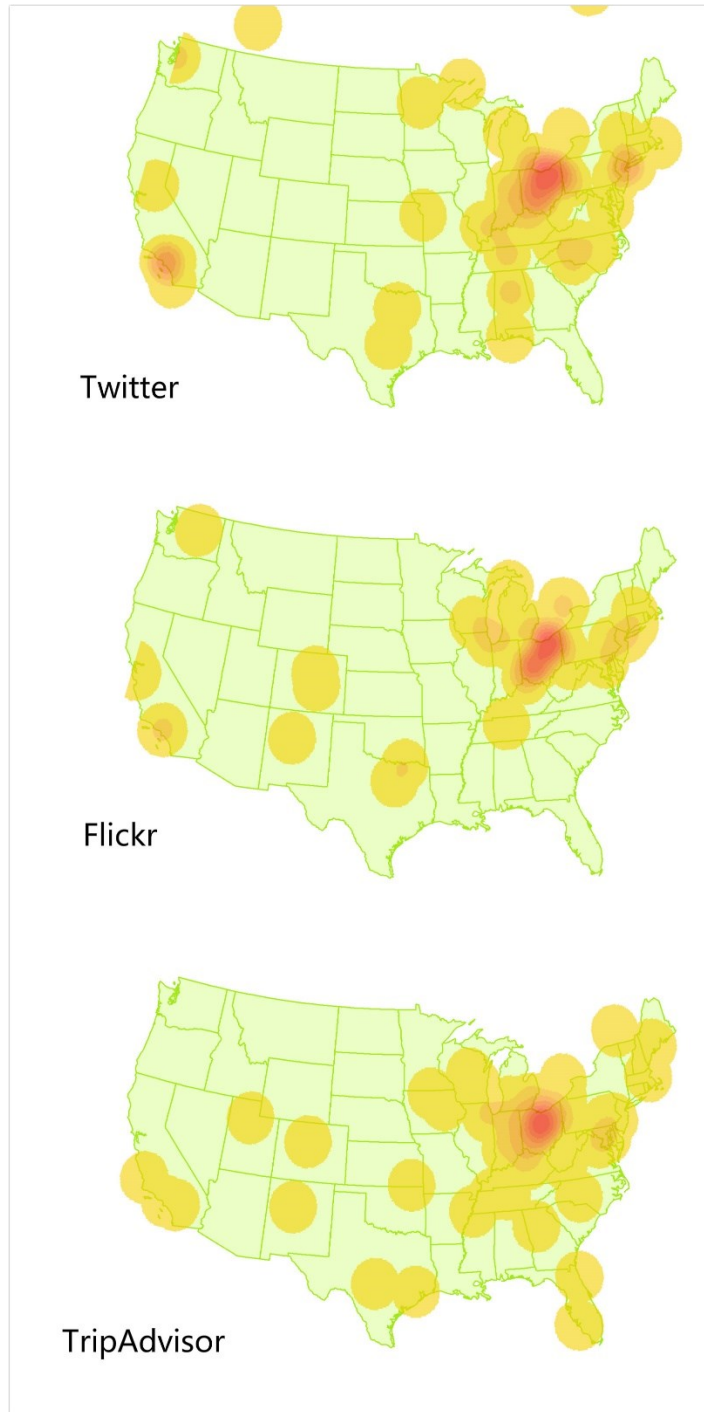


Figure 18a. The origins of visitors to Mansfield Reformatory on Flickr, Twitter, Tripadvisor at 3 different scales (world, national, regional)

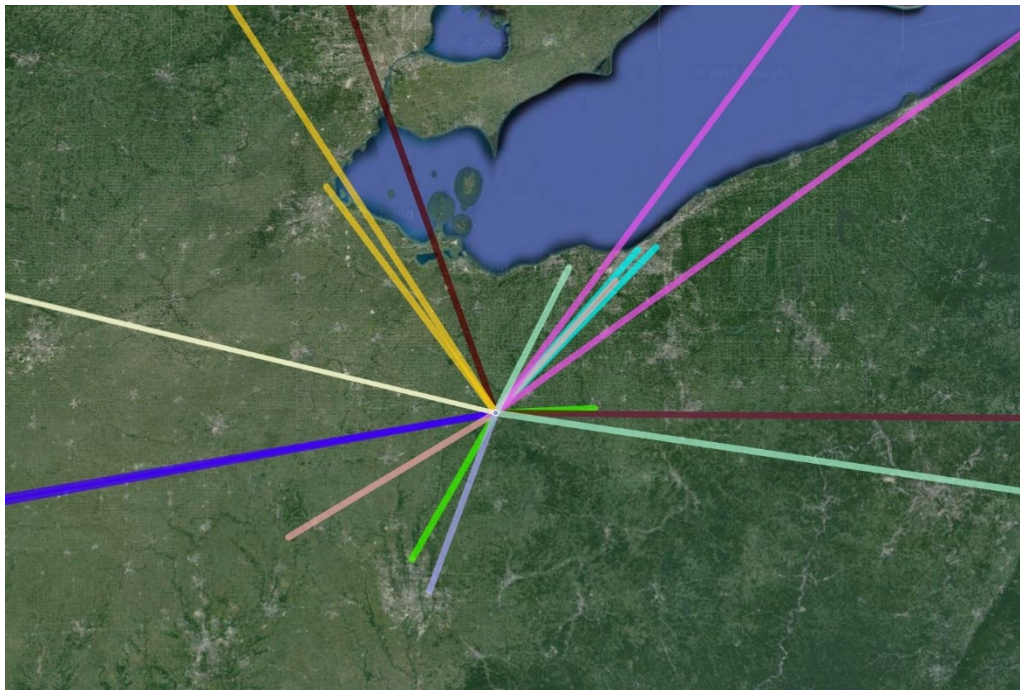
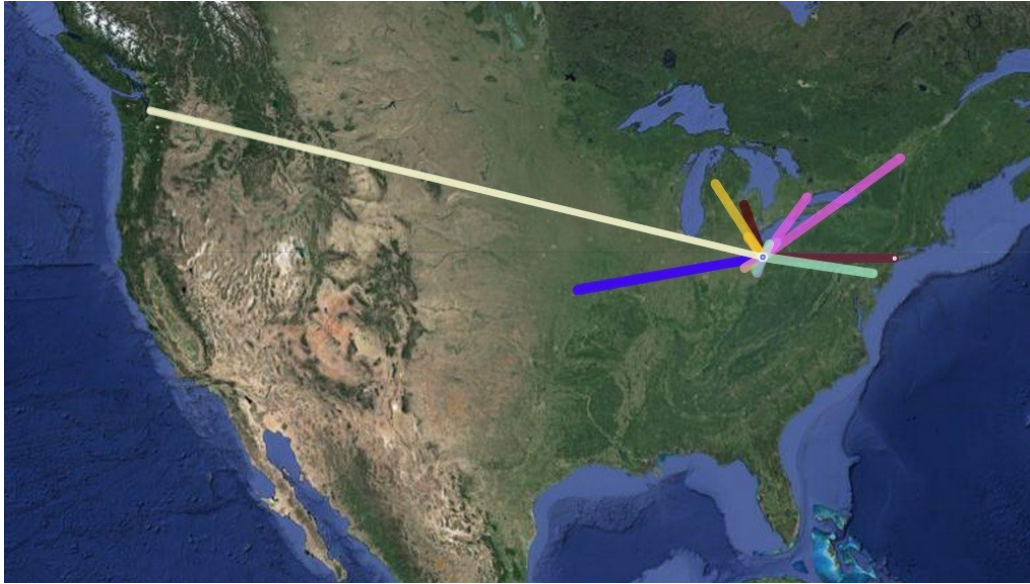


**Figure 18 b.** The hot spots of the origins of visitors to Mansfield Reformatory on Flickr, Twitter, Tripadvisor at national scale

### **4.3 Tracking the movement of visitors**

As mentioned in chapter 2, people like to use social media to share their experience about places. Social media can then be used to track the movement of their contributors. Tracking the movement of people within and around a touristic site could be useful for the site managers as well as for local planners and decision makers (Vu et al, 2014). Tracking these movements can also serve to highlight the potential of different social media to follow the displacement of people at different scales.

This section aims to identify the potential of each of these three social media to follow the paths and routes of visitors at different scales. In order to better understand this potential, I have randomly selected 10 users for each social media. I have then mapped the paths followed by these users using different methodologies for each social media. With Twitter, I have tracked the locations of the tweets posted near the reformatory as well as the location of the tweets posted right before their visit and the location of the tweet posted right after their visit. This method was chosen to identify the trajectories of the owners of these accounts since a large majority of them only tweeted once at the reformatory site as discussed previously (see Figure 19). For Tripadvisor, I mined all the locations that appeared in the long reviews to identify the places mentioned in these reviews that are also associated to the reformatory by the 10 selected users. To mine these places, I used GeoDoc and Cliff as mentioned in the methodology section. However, some of these locations could not be extracted automatically, such as the names of a restaurant or of a hotel, so I had to identify these names manually, and I was then able to create a map connecting all of the mined places with the reformatory (see Figure 20). Finally, since Flickr users produced more data on site, I have decided to map all the photos taken by the selected 10 Flickr users on the site of the reformatory to identify the path followed by these visitors (see Figure 21).



**Figure 19 a & b.** Tracking routes of Twitter users (National /Regional scale)

These lines connect each of the 10 selected geolocated Tweets that were posted from the Mansfield reformatory with the tweet posted right before and right after from each of the 10 selected Twitter accounts (each account is represented by a different color). This map was designed with AtlasCine and is available online:

<http://atascine3.classone-tech.com/index.html?module=1kvJSNVQUpnuaPY0lOQ8T7knd6UiJ9ya-KZiMfKG9cwI>

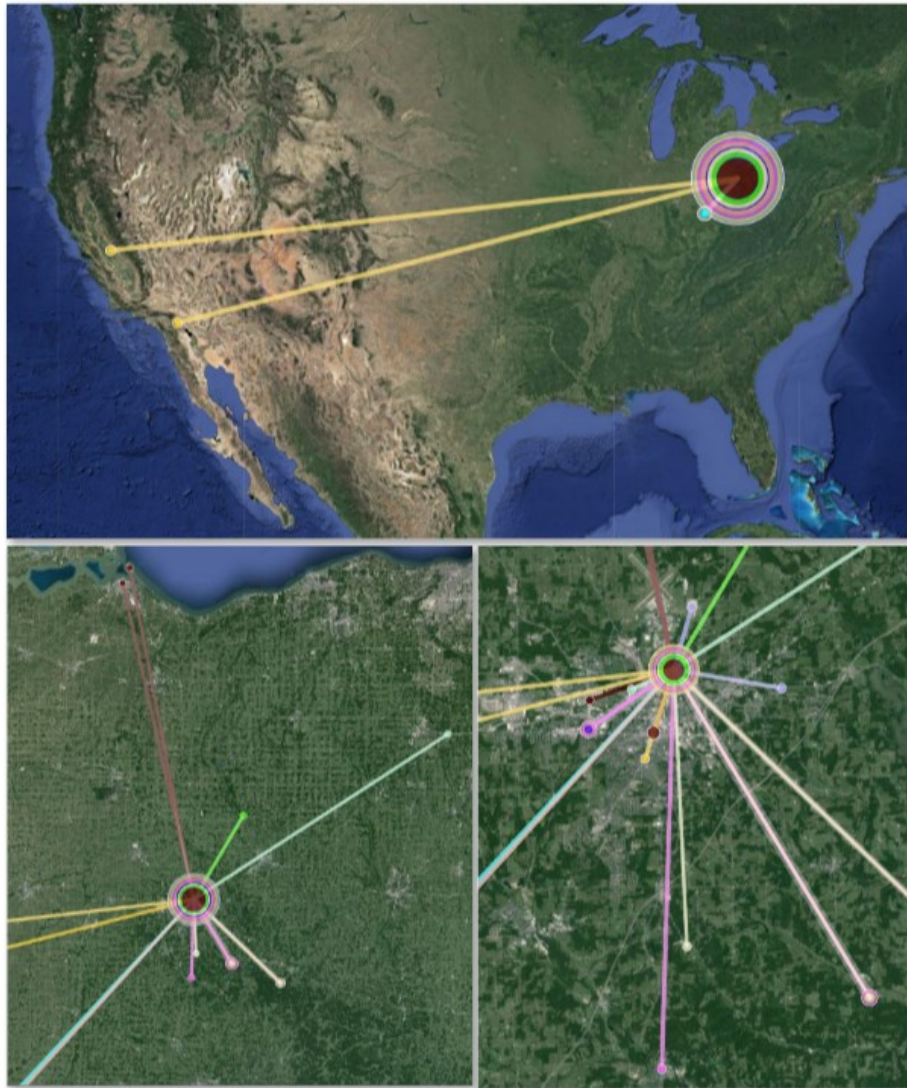
### **4.3.1 Tracking Tweets**

As mentioned previously, Twitter is not really relevant to tracking the local displacements associated to set-jetting. There are too few geolocated tweets and set-jettors most often only tweet one time when visiting the site. For these reasons it doesn't appear to be that relevant to use Twitter to try to keep track of set-jettors at the local scale. On the other hand, at the regional scale, the mapping of these tweets shows an interesting element. For most of the users (6/10), the tweets proceed as follows: the tweets sent before and after the one from the reformatory was sent from their hometown. Among the 10 selected users, 4 have their previous and following tweet sent from their hometown, which means that they did not tweet from any other places during their trip. This is furthermore surprising since some of these users had travelled a long distance to get to the Reformatory (i.e. one is located in Seattle, another one in Kansas city, a third one in Frankenmuth and the fourth one in Cleveland). Although it is not possible based on these data to affirm that these people only travelled to the Reformatory during their trip (e.g. there is a gap of 134 days between the 3 Tweets retrieved for the user from Seattle), it emphasizes that the visit to the Reformatory was definitely the only place that was worth Tweeting about during the trip. The Mansfield Reformatory is not only a fictional and real place, it is also a virtual one that has its own existence in social media.

### **4.3.2 Tracking places mentioned in the reviews in Tripadvisor**

Tripadvisor is much more interesting at the regional scale. Indeed, by connecting the reformatory with the different places mentioned in the comments, it is possible to highlight some spatial relationships. For instance, there are different places that visitors associate to the Mansfield Reformatory, such as hotel and restaurant. Most of these places are close by the Mansfield Reformatory with a special mention. For example, Restaurant Squirrel's Den is famous for its Shawshank sandwiches and is mentioned in one review. Although applying Natural Language Processing to extract information from Tripadvisor can allow a better understanding of what kind

of tourist facilities (i.e. hotels) are used by set-jettors, the process of automatically retrieving all the place names and geolocating them is not yet fully operational and still requires some more interpretation.



**Figure 20a, b &c.** Places mentioned in Tripadvisor in reviews talking about the Mansfield Reformatory

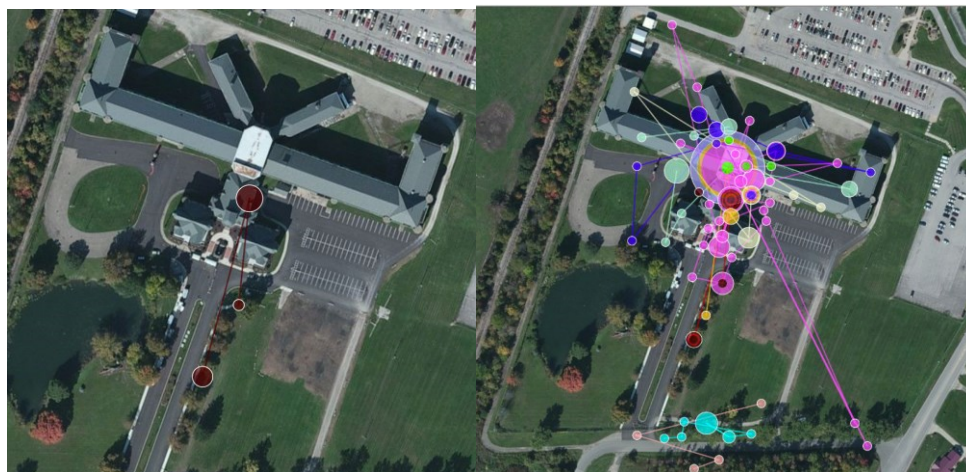
These lines connect the Mansfield reformatory with the different places mentioned in the 10 selected Tripadvisor reviews (each review is represented by a different color). This map was designed with AtlasCine and is available online:

<http://atlasCine3.classone-tech.com/index.html?module=1GAhIEjVHgfVTVsEynvQTZlo11LNvcqJv710qILrjDT0>



### 4.3.3 Tracking photos from Flickr

Finally, Flickr provides a good sense of the movement of set-jettors at the local scale. The map of the routes from Flickr (see Figure 21.a) shows not only the places where people like to take more photos as emphasized previously, but also the connections between these places. Based on these results, it seems that visitors usually start to take photos of the outlook of the building and then enter into the reformatory. We can even follow the movement of these visitors within the buildings, highlighting the potential of geotagged photos from Flickr to follow the path of visitors at a very fine scale. Although these potentialities have only been identified in this project they could be mobilized in a more systematic way to identify patterns of movements associated with the taking of pictures in touristic sites. This could be useful for merchandization as well as to better showcase some cultural artifacts and historical information at the site level. This map (Figure 21a & b) on the left shows the path of one visitor based on his/her Flickr photos while the map on the right shows all of the locations and connections of the 10 selected accounts.



**Figure 21 a & b.** Tracking routes of Flickr users (Local scale)

Figure 21a shows where photos were taken by one Flickr user, as well as the order in which these photos were taken (lines) and the number of photos taken at each location (size of the point). Figure 21b shows the same information for the 10 selected Flickr accounts (each account is represented by a different color). This map was designed with AtlasCine and is available online: [http://atascine3.classone-tech.com/index.html?module=1ds\\_UkJTJ\\_VTaj6O543VzQvoBIZzzC4PF6cJnYoezOQb8](http://atascine3.classone-tech.com/index.html?module=1ds_UkJTJ_VTaj6O543VzQvoBIZzzC4PF6cJnYoezOQb8)

#### 4.4 Summary

Overall, this analysis emphasizes the spatial complementarity of the three social media platforms to study set-jetting. At the fine scale of the site, Flickr provides valuable information about the main places of interest in terms of pictures, hot spots, and the movements of the visitors within the site. Although there are a few issues in terms of accuracy for some of these photos, these issues are minor and can be identified as discussed previously. Flickr appears to be an interesting tool to study the spatial movements of the visitors as well as their favourite places to take pictures. A high percentage of pictures downloaded via Flickr is also geolocated, which makes it an extremely useful platform to study set-jetting and tourist activities in more general terms at a local scale. Through the sharing of these photos, Flickr also contributes to the producing of imagined places associated with the site.

Tripadvisor appears to be more relevant to mining at a slightly broader scale, since it provides a range of information about nearby places. Tripadvisor offers a range of possibilities for the tourism industry through the connections people make between different places. It can help to identify where people stay, where they travel before and after visiting the site, what they enjoy and dislike, where they come from, etc. Tripadvisor is certainly an interesting source of geographic information but its mining requires more time and more sophisticated geoparsing tools not only to identify geographic places, but also to characterize them.

Finally, Twitter seems more relevant to studying the importance of set-jetting at a more global scale. Given the fact that most of the people tweet only once when they are on the site, that Twitter is mainly used by an elite minority and that only about 0.5% of the tweets are geotagged, it is irrelevant to use it to study set-jetting at the local scale. Nevertheless, Twitter presents a few assets to studying set-jetting: the data is available at the global scale, and the volume of data produced is impressive. These data can then be mobilized to address related questions: how do

people travel to get to the site? What are the other sites they tweet about? Where do they go once they have visited a site? That information could be highly relevant for studying the phenomena of set-jetting at the global scale. That said, the main challenge raised by the use of Twitter data is how to make sense of this big set of disparate data. Turning data from social media into meaningful information about places and spatial behaviours is probably the main challenge that researchers working with these data are facing at the moment.

## Chapter 5: Conclusion and Future Work

With the development of Web 2.0 technologies, the number of users of social media has been increasing dramatically. On a daily basis, an increasing number of users around the world generate a humongous amount of data with social media applications such as Twitter, Flickr or Tripadvisor. These social media are more and more envisioned as extended sources of geospatial information. The goal of this thesis was to review and assess some of the potentialities these social media offer to neogeographers, with a specific focus on studying a spatio-cultural phenomenon: Set-jetting.

In this paper, different kinds of approaches for mining, extracting, and analyzing geodata from social media have been proposed and compared. This comparison allowed to further explore the ways people use social media in the context of set-jetting in order to see how different social media platforms could be mobilized differently by different actors involved in this activity, such as researchers, site managers, local decision makers and the tourism industry.

Throughout the analysis of the data mined with these three social media, the spatial complementarity of these social media to study set-jetting has been emphasized. Twitter appears mainly relevant to study set-jetting at the global scale. It allows identifying who is talking about the real and the fictional locations of movie scenes. Where are these people and where do they travel before and after getting to the real site of the movie shooting? Although the data from Twitter are relatively easy to mine, they present some serious flaws. Indeed, mining data from Twitter requires using a few processes and different pieces of software to access data that are only available for the last 6-9 days for the public. Mining data from Twitter requires setting up a

rigorous process. But the more challenging part is to make sense of the data mined. Although the use of different key words can serve to differentiate the tweets referring to the fictional place that appeared in the movie (i.e. Shawshank + Jail/prison) from the tweets referring to the real shooting locations (i.e. Mansfield Reformatory), these references to the real location are not sufficient to conclude that these tweets are directly linked to a visit (past, present or future) of the site. Mobilizing Twitter to study set-jetting requires relying on a small percentage of tweets that are georeferenced (i.e. about 0.5%). Furthermore, since most (more than 95%) of the Twitter users post one tweet at the site visited, this platform is not appropriate to studying the geography of set-jetting at a local scale.

Flickr appears to be much more relevant to studying set-jetting at the local level since about 95% of the people uploading photos from the site visited upload more than one photo, and since about 29.5% of these photos are geotagged. Flickr is highly relevant to following people's movements and more specifically tourist movements as already highlighted in different studies (Vu et al., 2014). The process of mining data from Flickr is relatively easy. However most of the photos do not have a detailed description and some of them are not geotagged directly through GPS equipment but manually, which influences the accuracy of the geolocation. Although the textual data available from Flickr is minimalistic, photos could be extremely rich in terms of visual information. Even if we are not able to extract much information at the moment from these images, it is clear that they will soon be an extremely valuable source of geographic information.

Tripadvisor appears to be an interesting source of geographic data about people's subjective experiences related to their site visit and their set-jetting experience. In terms of its geographic relevance, this source of data could be useful to studying set-jetting at a more regional scale. Tripadvisor provides a rich geographic database about set-jetting in the extensive reviews it provides. However, mining textual data requires mobilizing geoparsing techniques, which can be

automated up to a certain point. Although geoparsing can easily identify place names and associate them to geographic coordinates via a gazetteer, there are a lot of places that are not necessarily included in gazetteers (e.g. names of hotels and restaurants) and there are a lot of nuances associated to these places that cannot be identified systematically yet with geoparsing processes. This is another growing area of research that is attracting a lot of interest given the increasing volume of text produced on a daily basis with social media.

At the end of this project it appears that social media could certainly be mobilized to study set-jetting for different purposes at different scales. Although a lot of time and energy has been mobilized so far to mine these data (not only for this thesis, but in more general terms), the upcoming challenge is to find better ways to make sense of these data (textual and visual) in a systematic way in order to better understand not only how the use of social media is changing our relationships with places, but also how they are changing our interaction with fictional places through real places and vice-versa.

## References

- Abascal-Mena, R., & López-Ornelas, E. (2010). Geo information extraction and processing from travel narratives. *Proc. ELPUB*, 363-373.
- Andersen, P. (2007). What is Web 2.0?: ideas, technologies and implications for education (Vol. 1, No. 1). Bristol, UK: JISC.
- Batty, M., Hudson-Smith, A., Milton, R., & Crooks, A. (2010). Map mashups, Web 2.0 and the GIS revolution. *Annals of GIS*, 16(1), 1-13.
- Big data. (2015, July 19). In Wikipedia, The Free Encyclopedia. Retrieved 05:21, July 21, 2015, from [https://en.wikipedia.org/w/index.php?title=Big\\_data&oldid=672125766](https://en.wikipedia.org/w/index.php?title=Big_data&oldid=672125766)
- Behuniak, S. M. (2011). The living dead? The construction of people with Alzheimer's disease as zombies. *Ageing and Society*, 31(01), 70-92.
- Bolan, P., Boy, S., & Bell, J. (2011). "We've seen it in the movies, let's see if it's true" Authenticity and displacement in film-induced tourism. *Worldwide Hospitality and Tourism Themes*, 3(2), 102-116.
- Bolan, P., & Williams, L. (2008). The role of image in service promotion: focusing on the influence of film on consumer choice within tourism. *International Journal of Consumer Studies*, 32(4), 382-390.
- Brown, B., Chui, M., & Manyika, J. (2011). Are you ready for the era of 'big data'. *McKinsey Quarterly*, 4, 24-35.
- Busby, G., & Klug, J. (2001). Movie-induced tourism: The challenge of measurement and other issues. *Journal of vacation marketing*, 7(4), 316-332.
- Caquard, S., & Cartwright, W. (2014). Narrative cartography: From mapping stories to the narrative of maps and mapping. *The Cartographic Journal*, 51(2), 101-106.
- Cartwright W (2012) Neocartography: Opportunities, Issues and Prospects. *South African Journal of Geomatics* 1(1):14-31
- Caverlee, J., Cheng, Z., Sui, D. Z., & Kamath, K. Y. (2013). Towards Geo-Social Intelligence:

Mining, Analyzing, and Leveraging Geospatial Footprints in Social Media. *IEEE Data Eng. Bull.*, 36(3), 33-41.

Clements, M., Serdyukov, P., De Vries, A. P., & Reinders, M. J. (2010, July). Using flickr geotags to predict user travel behaviour. *In Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval* (pp. 851-852). ACM.

CLIFF: Entity Extraction and Geoparsing for News Articles. (n.d.). Retrieved January 5, 2015, from <http://cliff.mediameter.org/>

Coe, N. M. (2000). The view from out west: Embeddedness, inter-personal relations and the development of an indigenous film industry in vancouver. *Geoforum*, 31(4), 391-407.

Connell, J. (2005). Toddlers, tourism and Tobermory: Destination marketing issues and television-induced tourism. *Tourism Management*, 26(5), 763-776.

Crampton, J. W., Graham, M., Poorthuis, A., Shelton, T., Stephens, M., Wilson, M. W., & Zook, M. (2013). Beyond the geotag: situating 'big data' and leveraging the potential of the geoweb. *Cartography and geographic information science*, 40(2), 130-139.

Cunningham, H., Maynard, D., Bontcheva, K., & Tablan, V. (2002, July). A framework and graphical development environment for robust NLP tools and applications. In *ACL* (pp. 168-175).

Darabont, F., Marvin, N., Robbins, T., Freeman, M., Gunton, B., Sadler, W., Brown, C., ... Warner Home Video (Firm). (2004). *The Shawshank Redemption*. Burbank, CA: Warner Bros. Pictures.

Demetradi, J. (1996). The tele tourists: Are they a boon or just a nuisance?, *Hospitality*, October/November, 14-15.

Duggan, M., & Brenner, J. (2013). *The demographics of social media users, 2012* (Vol. 14). Washington, DC: Pew Research Center's Internet & American Life Project.

Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information science*, 38(2), 189-200.

Facebook Passes 1.23 Billion Monthly Active Users. (2014, January 29). Retrieved January 5,



- 2015, from  
<http://thenextweb.com/facebook/2014/01/29/facebook-passes-1-23-billion-monthly-active-users-945-million-mobile-users-757-million-daily-users/>
- Faby, H., & Koch, A. (2010, June). From maps to neo-cartography. *In Proceedings of the 3rd Interational Conference on Cartography & GIS* (p. 6).
- Fact Sheet - TripAdvisor. (n.d.). Retrieved March 4, 2015. From:  
[http://www.tripadvisor.ca/PressCenter-c4-Fact\\_Sheet.html](http://www.tripadvisor.ca/PressCenter-c4-Fact_Sheet.html)
- Fischer F (2012) VGI as Big Data. A New but Delicate Geographic Data-Source. *GeoInformatics* (April/May 2012): 46-47.
- Flickr API documentation. (n.d.). Retrieved March 19, 2014, from  
<https://www.flickr.com/services/api/>
- Flickr At 10: 1M Photos Shared Per Day, 170% Increase Since Making 1TB Free. (n.d.). Retrieved July 22, 2015, from techcrunch:  
<http://techcrunch.com/2014/02/10/flickr-at-10-1m-photos-shared-per-day-170-increase-since-making-1tb-free/>
- Former prison in Ohio draws captive audience, thanks to 'Shawshank Redemption' (n.d.). Retrieved April 5, 2015, from  
<http://www.post-gazette.com/ae/movies/2014/08/24/Former-Mansfield-reformatory-in-Ohio-draws-captive-audience-thanks-to-Shawshank-Redemption/stories/201408240010>
- Freire, N., Borbinha, J., Calado, P., & Martins, B. (2011, June). A metadata geoparsing system for place name recognition and resolution in metadata records. In Proceedings of the 11th annual international ACM/IEEE joint conference on Digital libraries (pp. 339-348). ACM.
- Friedman, T. L. (2006). The world is flat [updated and expanded]: A brief history of the twenty-first century. Macmillan.
- Frost, W. (2010). Life changing experiences: Film and tourists in the Australian outback. *Annals of Tourism Research*, 37(3), 707-726.
- Gazdar, G. (1996). Paradigm merger in natural language processing.
- Gelernter, J., & Mushegian, N. (2011). Geo-parsing messages from microtext. *Transactions in*

*GIS*, 15(6), 753-773.

GoGeo. (n.d.). Retrieved September 26, 2014, from <http://www.gogeo.ac.uk/gogeo/metadata/geodoc.htm>

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, 69(4), 211-221.

Haklay, M., Basiouka, S., Antoniou, V., & Ather, A. (2010). How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. *The Cartographic Journal*, 47(4), 315-322.

Haklay, M., Singleton, A., & Parker, C. (2008). Web mapping 2.0: The neogeography of the GeoWeb. *Geography Compass*, 2(6), 2011-2039.

Haklay, M., & Weber, P. (2008). Openstreetmap: User-generated street maps. *Pervasive Computing*, IEEE, 7(4), 12-18.

Herring, C. (1994). *An architecture of cyberspace: Spatialization of the internet*. US: Army Construction Engineering Research Laboratory.

Hoff, R.D. (2006) A virtual world's real Dollars, *Business Week*. Retrieved 18:19, July 19, 2014, from: [http://www.businessweek.com/technology/content/mar2006/tc20060328\\_688225.htm](http://www.businessweek.com/technology/content/mar2006/tc20060328_688225.htm)

Hollenstein, L., & Purves, R. (2015). Exploring place through user-generated content: Using Flickr tags to describe city cores. *Journal of Spatial Information Science*, (1), 21-48.

Hudson-Smith, A., Batty, M., Crooks, A., & Milton, R. (2009). Mapping for the masses accessing Web 2.0 through crowdsourcing. *Social science computer review*, 27(4), 524-538.

Hudson, S., & Ritchie, J. B. (2006). Promoting destinations via film tourism: An empirical identification of supporting marketing initiatives. *Journal of Travel Research*, 44(4), 387-396.

Iwashita, C. (2006). Media representation of the UK as a destination for Japanese tourists Popular culture and tourism. *Tourist studies*, 6(1), 59-77.

Jackson, J.B(1979). *Landscape as Theater: Landscape* 23(1): 3-7

- Jobst, M. (2009): Neo-cartographic interlacement as barrier for Cartographic Heritage. In: e-Perimtron, Vol. 4, No. 4, 2009, 212-220.
- Joliveau, T. (2009). Connecting real and imaginary places through geospatial technologies: Examples from set-jetting and art-oriented tourism. *The Cartographic Journal*, 46(1), 36-45.
- Jones, D., & Smith, K. (2005). Middle- earth Meets New Zealand: Authenticity and Location in the Making of The Lord of the Rings\*. *Journal of Management Studies*, 42(5), 923-945
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, 53(1), 59-68.
- Kennedy, C. B. (1994). The myth of heroism: man and desert in Lawrence of Arabia. *Place, Power, Situation, and Spectacle: A Geography of Film*. Lanham, MD: Rowman and Littlefield, 161-179.
- Kennedy, C., & Lukinbeal, C. (1997). Towards a holistic approach to geographic research on film. *Progress in Human Geography*, 21(1), 33-50.
- Kent, J. D., & Capello Jr, H. T. (2013). Spatial patterns and demographic indicators of effective social media content during theHorse Thief Canyon fire of 2012. *Cartography and Geographic Information Science*, 40(2), 78-89.
- King, S. (1982). Rita Hayworth and the Shawshank redemption. S. King *Different seasons*, 15-106.
- Krishnan, K. (2013). Data warehousing in the age of big data. Newnes.
- Lee, S., Scott, D., & Kim, H. (2008). Celebrity fan involvement and destination perceptions. *Annals of Tourism Research*, 35(3), 809-832.
- Leidner, J. L. (2008). *Toponym resolution in text: Annotation, evaluation and applications of spatial grounding of place names*. Universal-Publishers.
- Li, L., Goodchild, M. F., & Xu, B. (2013). Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartography and Geographic Information Science*, 40(2), 61-77.
- Lukinbeal, C. (2005). Cinematic landscapes. *Journal of Cultural Geography*, 23(1), 3-22.

- Lukinbeal, C. (2012). "On Location" Filming in San Diego County from 1985–2005: How a Cinematic Landscape Is Formed Through Incorporative Tasks and Represented Through Mapped Inscriptions. *Annals of the Association of American Geographers*, 102(1), 171-190. *Journal of Cultural Geography*, 23(1), 3-22.
- Luo, J., Joshi, D., Yu, J., & Gallagher, A. (2011). Geotagging in multimedia and computer vision—a survey. *Multimedia Tools and Applications*, 51(1), 187-211.
- MacEachren, A. M., Robinson, A. C., Jaiswal, A., Pezanowski, S., Savelyev, A., Blanford, J., & Mitra, P. (2011, July). Geo-twitter analytics: Applications in crisis management. In 25th International *Cartographic Conference* (pp. 3-8).
- Macionis, N. (2004, November). Understanding the film-induced tourist. In International tourism and media conference proceedings (Vol. 24, pp. 86-97). Melbourne: Tourism Research Unit, Monash University.
- Manovich, L. (2011). Trending: the promises and the challenges of big social data. *Debates in the digital humanities*, 460-475.
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity.
- Marjit, U., & Jana, S. (2009). *Mashups: An Emerging Content Aggregation Web2. 0 Paradigm*.
- November, V., Camacho-Hübner, E. & Latour, B., 2010. Entering a risky territory: space in the age of digital navigation. *Environment and Planning D: Society and Space*, 28(4), 581-599.
- O'Connor, N., & Bolan, P. (2008). Creating a sustainable brand for Northern Ireland through film-induced tourism. *Tourism Culture & Communication*, 8(3), 147-158.
- On Location: Mansfield, Ohio's 'Shawshank' Industry. (n.d.). Retrieved June 5, 2014, from <http://www.npr.org/2011/08/04/138986482/on-location-mansfield-ohios-shawshank-industry>
- O'Reilly, T. (2005). Web 2.0: compact definition. Message posted to [http://radar.oreilly.com/archives/2005/10/web\\_20\\_compact\\_definition.html](http://radar.oreilly.com/archives/2005/10/web_20_compact_definition.html).
- Riley, R., Baker, D., & Doren, C. S. V. (1998). Movie induced tourism. *Annals of tourism*

*research*, 25(4), 919-935.

Roche, S., Propeck-Zimmermann, E., & Mericskay, B. (2013). GeoWeb and crisis management: Issues and perspectives of volunteered geographic information. *GeoJournal*, 78(1), 21-40.

Social media (n.d.). In Oxford dictionary Online, Retrieved September 24, 2013, from <http://www.oxforddictionaries.com/definition/english/social-media>

The Shawshank Redemption. (2015, July 19). In Wikipedia, The Free Encyclopedia. Retrieved July 19, 2015, from [https://en.wikipedia.org/w/index.php?title=The\\_Shawshank\\_Redemption&oldid=672075380](https://en.wikipedia.org/w/index.php?title=The_Shawshank_Redemption&oldid=672075380)

Tsou, M. H., & Leitner, M. (2013). Visualization of social media: seeing a mirage or a message?. *Cartography and Geographic Information Science*, 40(2), 55-60.

Turner, A. (2006). Introduction to neogeography. " O'Reilly Media, Inc."

Twitter Developer Support. (n.d.). Retrieved March 22, 2014. <https://dev.twitter.com/overview/general/support>

Tzanelli, R. (2004). Constructing the 'cinematic tourist' The 'sign industry' of The Lord of the Rings. *Tourist Studies*, 4(1), 21-42.

Vu, H. Q., Li, G., Law, R., & Ye, B. H. (2015). Exploring the travel behaviors of inbound tourists to Hong Kong using geotagged photos. *Tourism Management*, 46, 222-232.

Web 2.0. (2015, July 15). In Wikipedia, The Free Encyclopedia. Retrieved 18:19, July 19, 2014, from [https://en.wikipedia.org/w/index.php?title=Web\\_2.0&oldid=671564440](https://en.wikipedia.org/w/index.php?title=Web_2.0&oldid=671564440)

Wenders, W. (2001). *On film*. Faber & Faber.

House, W. (2012). Consumer data privacy in a networked world: A framework for protecting privacy and promoting innovation in the global digital economy. *White House, Washington, DC*.

Xiang, Z., & Gretzel, U. (2010). Role of social media in online travel information search. *Tourism management*, 31(2), 179-188.

Xu, C., Wong, D. W., & Yang, C. (2013). Evaluating the “geographical awareness” of individuals: An exploratory analysis of Twitter data. *Cartography and Geographic Information Science*, 40(2), 103-115.

Zong, C. (2008). Statistical natural language processing. *Tsinghua University Publisher*, 342-343.

## Appendix I

2013-11-09 23:31:00 RT @MartinBelan: #Nature & Travel #Photo of the Day, Mansfield Reformatory - East #Cell Block - 11/9/13 <http://t.co/36me5SHTFZ> None

2013-11-09 23:10:33 #Nature & Travel #Photo of the Day, Mansfield Reformatory - East #Cell Block - 11/9/13 <http://t.co/36me5SHTFZ> None

2013-11-09 17:35:36 RT @MartinBelan: Nature & #Travel #Photo of the Day, Mansfield Reformatory - East #Cell Block - 11/9/13 <http://t.co/36me5SHTFZ> None

2013-11-09 17:03:33 Nature & #Travel #Photo of the Day, Mansfield Reformatory - East #Cell Block - 11/9/13 <http://t.co/36me5SHTFZ> None

2013-11-09 13:02:33 #Nature & #Travel #Photo of the Day, Mansfield Reformatory - East #Cell Block - 11/9/13 <http://t.co/36me5SHTFZ> None

2013-11-09 12:26:04 via @MartinBelan Nature & Travel Photo of the Day, Mansfield Reformatory – East Cell Block – 11/9... <http://t.co/Zd8Ep37uAi> #photography None

2013-11-09 09:39:14 Wow!! May 2, 3 George Noori of @coasttocoastam will be doing a live broadcast from Mansfield at the Ohio State Reformatory!! None

2013-11-08 03:41:48 @PatrickHTDoyle let me know when you guys ever make it to Ohio, Mansfield Reformatory is an awesome place to investigate. None

2013-11-08 00:46:24 Photographing the Mansfield Reformatory <http://t.co/H9hKgfZqE7> #photo None

2013-11-07 23:47:52 I liked a @YouTube video <http://t.co/x8tGpstybf> Ghost Adventures at Mansfield Reformatory Photos; April 14th 2011 None

2013-11-07 21:48:32 Talked to Cody Richardson about the overnight lockdown tour at the Mansfield reformatory and it sounds badass whose in!?!? None

2013-11-07 20:11:15 RT @Ktondo: I want to go to the Mansfield reformatory so bad None

2013-11-07 20:10:17 I want to go to the Mansfield reformatory so bad None

2013-11-06 03:27:05 RT @MartinBelan: Two Looks for an Old Prison Chair in the Mansfield Reformatory <http://t.co/kBiFZyiwB7> #photo None

2013-11-06 02:15:37 RT @MartinBelan: Two Looks for an Old Prison Chair in the Mansfield Reformatory <http://t.co/kBiFZyiwB7> #photo None

2013-11-06 00:36:12 Two Looks for an Old Prison Chair in the Mansfield Reformatory <http://t.co/kBiFZyiwB7> #photo None

2013-11-05 13:32:15 The\_Awesome\_Possum @Jills\_Awesome

2013-11-05 In 5 months... round 2. I'm ready. #MansfieldReformatory pic.twitter.com/EAdOcXB83t

2013-11-04 10:56:32 Betsy Sharp @BetsySharp I'm at Mansfield Reformatory (Mansfield, OH) <http://4sq.com/19uXUh5> [40.78412404404071,-82.50259280204773]



## Appendix II

```
{
  "metadata": {
    "result_type": "recent",
    "iso_language_code": "en"
  },
  "created_at": "Wed Nov 13 02:27:16 +0000 2013",
  "id": 400449757459800060,
  "id_str": "400449757459800064",
  "text": "Amazing Mansfield reformatory http://t.co/xnpdN1yZ6R",
  "source": "<a href='http://instagram.com' rel='nofollow'>Instagram</a>",
  "truncated": false,
  "in_reply_to_status_id": null,
  "in_reply_to_status_id_str": null,
  "in_reply_to_user_id": null,
  "in_reply_to_user_id_str": null,
  "in_reply_to_screen_name": null,
  "user": {
    "id": 518186788,
    "id_str": "518186788",
    "name": "Dom GhostHunter DGH",
    "screen_name": "doughosthunter",
    "location": " CANADA",
    "description": "here is my own fb page https://t.co/GIZozfBmbN and team fb page
https://t.co/7FisiU4J67 and youtube http://t.co/Uf7R3MDDZj",
    "url": null,
    "entities": {
      "description": {
        "urls": [
          {
            "url": "https://t.co/GIZozfBmbN",
            "expanded_url": "https://www.facebook.com/DomghosthunterParanormal?fref=ts",
            "display_url": "facebook.com/Domghosthunter...",
            "indices": [
              23,
              46
            ]
          }
        ]
      }
    }
  }
}
```

```

    },
    {
      "url": "https://t.co/7FisiU4J67",
      "expanded_url": "https://www.facebook.com/CanAmParanormal",
      "display_url": "facebook.com/CanAmParanormal",
      "indices": [
        64,
        87
      ]
    },
    {
      "url": "http://t.co/Uf7R3MDDZj",
      "expanded_url": "http://www.youtube.com/user/dominatorfn/videos",
      "display_url": "youtube.com/user/dominator...",
      "indices": [
        100,
        122
      ]
    }
  ]
}
},
"protected": false,
"followers_count": 10729,
"friends_count": 7434,
"listed_count": 153,
"created_at": "Thu Mar 08 04:11:58 +0000 2012",
"favourites_count": 1673,
"utc_offset": -18000,
"time_zone": "Eastern Time (US & Canada)",
"geo_enabled": false,
"verified": false,
"statuses_count": 39121,
"lang": "en",
"contributors_enabled": false,
"is_translator": false,
"profile_background_color": "1F1C1A",

```

```

        "profile_background_image_url":
"http://a0.twimg.com/profile_background_images/378800000025606502/8ee385615b7ceae0d4edc800b4
696ceb.jpeg",
        "profile_background_image_url_https":
"https://si0.twimg.com/profile_background_images/378800000025606502/8ee385615b7ceae0d4edc800b
4696ceb.jpeg",
        "profile_background_tile": true,
        "profile_image_url":
"http://pbs.twimg.com/profile_images/378800000735368935/17aa6fb6845070f18d481a848fa82b99_nor
mal.jpeg",
        "profile_image_url_https":
"https://pbs.twimg.com/profile_images/378800000735368935/17aa6fb6845070f18d481a848fa82b99_nor
mal.jpeg",
        "profile_banner_url": "https://pbs.twimg.com/profile_banners/518186788/1384360495",
        "profile_link_color": "EB2F2F",
        "profile_sidebar_border_color": "000000",
        "profile_sidebar_fill_color": "DDEEF6",
        "profile_text_color": "333333",
        "profile_use_background_image": true,
        "default_profile": false,
        "default_profile_image": false,
        "following": false,
        "follow_request_sent": false,
        "notifications": false
    },
    "geo": null,
    "coordinates": null,
    "place": null,
    "contributors": null,
    "retweet_count": 0,
    "favorite_count": 0,
    "favorited": false,
    "retweeted": false,
    "possibly_sensitive": false,
    "lang": "en"
}

```

## Appendix III

### Search by KEY WORDS: Mansfield reformatory

```
import time
import tweepy

mykey = 'Ix5uKcQgtOHvy9SIQwgt6Q'
mysecret = 'AiqWFehN0ikljulyeMS7TWAY2xYynW2UWaNrrJBY2M'
mytoken = '490027058-AWycQ0F6QLZh83QavHtd1he1xjaTyUA4R55HvH0c'
mysecret = 'm1nRlLmMcKDN1QW9BMuFYIXYtmliAHXkySAzfoYg'

auth = OAuthHandler(mykey, mysecret)
auth.set_access_token(mytoken, mysecret)

api = tweepy.API(auth)
for tweet in tweepy.Cursor(api.search,
                            q = "mansfield reformatory",
                            count=100,
                            result_type="recent",
                            include_entities=True,
                            lang="en"
                            ).items():
    print tweet.created_at, tweet.text, tweet.geo
```

### Search by geocode

```
from tweepy import Stream
```

```

from tweepy import OAuthHandler

from tweepy.streaming import StreamListener

import time

mykey = 'Ix5uKcQgtOHvy9SIQwgt6Q'
mysecret = 'AiqWFehN0ikljulyeMS7TWAY2xYynW2UWaNrrJBY2M'
mytoken = '490027058-AWycQ0F6QLZh83QavHtd1he1xjaTyUA4R55HvH0c'
mysecret = 'm1nRILmMcKDN1QW9BMuFYIXYtmliHXkySAzfoYg'

class listener(StreamListener):

    def on_data(self, data):

        try:

            print data

            saveFile = open("results.txt", "w")

            saveFile.write(data)

            saveFile.write('\n')

            saveFile.close()

            return True

        except BaseException, e:

            print 'failed ondata,',str(e)

            time.sleep(5)

    def on_error(self,status):

        print status

```

```
auth = OAuthHandler(mykey, mysecret)
auth.set_access_token(mytoken, mysecret)
twitterStream = Stream(auth, listener())
twitterStream.filter(locations =[40.78546,-82.5055,40.78269,-82.4987 ]
```