Citizen science – exploring tourists' observations on micro blogs as a tool to monitor environmental change

---

Author 1

Jinyan, Chen

Research Assistant

Griffith Institute for Tourism

Griffith University, Australia

Email address: jinyan.chen@griffith.edu.au

Author 2

Susanne, Becken

Professor

Griffith Institute for Tourism

Griffith University, Australia

Email address: s.becken@griffith.edu.au

Author 3

Bela, Stantic

Associate Professor

Institute of Integrated and Intelligent Systems

Griffith University, Australia

Email address: b.stantic@griffith.edu.au

Acknowledgement: This project (2.3.2) is funded through the Tropical Water Quality Hub as part of the National Environmental Science Programme (Australia).

ABSTRACT

The Great Barrier Reef, Queensland, Australia, is an important global asset of environmental significance. The GBR is at great risk from multiple stress factors and understanding trends of decline or recovery requires intensive and costly monitoring. Exploring cost-effective data sources that could complement biophysical monitoring is therefore an important opportunity that needs to be considered. Social media and reports of reef sightings provide one potential avenue to innovatively integrate data from different scales into a consolidated monitoring system. This paper presents research on how to harness micro-blog data in conjunction with other Open data to extract information on environmental and aesthetic conditions of the Great Barrier Reef, in real time. We present a conceptual framework to clarify the nature and contribution of different data sources, and discuss specific data requirements and analytical techniques. Whilst innovative in its approach, this research also advances computer science with regard to development of new algorithms to deal with this high volume of different data types in real time to enable extraction of valuable information and deep learning from Big Data. If proven successful, the approach presented here could provide a pathway for monitoring other at-risk environmental assets that are heavily used by tourists.

Keywords: Citizen Science, Big Data, Twitter, Great Barrier Reef,

Environmental Monitoring

INTRODUCTION

The Great Barrier Reef (GBR) World Heritage Area in Queensland, Australia, is one of the world's richest and most diverse natural ecosystems. Given its status as a natural asset of global significance, it is not surprising that the GBR features prominently in Australia's tourism advertising (Shafer and Inglis, 2000), drawing more than 2 million visitors on commercial tours to experience the Reef each year (Great Barrier Reef Marine Park Authority (GBRMPA), 2014).The protection of the GBR is challenging, and scientists have been concerned about the environmental integrity since its designation as a World Heritage Area in 1981 (Day, 2013). For instance, over the last three decades, the GBR has lost more than half its coral cover, mainly as a result of agricultural run-off, outbreaks of crown of thorns star fish, cyclones, and a warming of water temperatures due to climate change (De'ath, Fabricius, Sweatman and Puotinen, 2012). To balance the need to protect the reef with economic development in the region, the management of the GBRWHA is complex and multi-layered (Grech et al., 2013). Legislation and management are faced with the challenge to cater for multiple user groups and their specific needs, including a range of commercial and recreational activities, such as tourism.

In the face of multiple interacting and cumulative stress factor that compromise the health of the GBR, the GBRMPA (Addison et al., 2015) is now working towards an integrated monitoring program to help evaluate progress towards the Long Term Sustainability Plan. The goal is to better integrate the over 80 different monitoring programs and also enhance existing ones with new approaches, including through the use of citizen science. Citizen science has already been implemented in some areas, for example by involving school groups in monitoring of river ecology. Current monitoring programs therefore not only focus on the marine life (e.g. dugongs, turtles, seagrass, corals, acidity, temperature and nitrogen concentration), but also consider land use practices (e.g. in the different agricultural sectors) and socio-economic changes related to both residents and visitors. Monitoring of progress is essential against the background of deteriorating conditions and UNESCO's pending decision on listing the GBR as a World Heritage site in danger (Liburd and Becken, forthcoming).

The tourism industry has always seen themselves as a steward of the GBR (Becken, Moyle and McLennan, 2014) – knowing that considerable environmental degradation might reduce expenditures by visitor to the GBR by a minimum of 17% (Mustika et al., 2015). Recognising its dependence on a healthy Reef, tourism has engaged substantially in activities that help protect the GBR. For example, Reef operators are contributing to environmental monitoring through the Eye on the Reef program managed by GBRMPA. The Eye on the Reef program enables operators and visitors to contribute information about reef health, marine animals and incidents, which is then integrated into the Park's management platform. The program requires several levels of expertise. The mobile App allows anyone to share sightings, either by describing species or uploading photos. The Rapid Monitoring Survey requires more experience as an underwater monitoring slate is completed and submitted to the database afterwards through an online portal. The Tourism Operators Weekly Monitoring Survey requires ongoing commitment and monitoring of environmental indicators in the same location (i.e. where operators have a license to go diving). This comprehensive approach to citizen science is important as it not only gets people engaged, but it delivers much needed additional data that enhance the very costly scientific monitoring.

The use of social media or Twitter to monitor precarious situations is not new. One of the first fields of application of Twitter as a 'sensor' has been in disaster management. Analysis of 10 million tweets following hurricane Sandy in New York in 2012, for example, showed that tweets outperformed the national Federal Emergency Management Agency in reporting damage in real time (Bohannon, 2016). Capitalising on the quick spread of information via Twitter, the US Geological Service has now augmented its network of seismological sensors with real-time feed of tweets. Learning over time has shown that people being close to an earthquake tend to write brief tweets (because they are scared or uncomfortable), and the algorithm used now gives greater weight to shorter tweets over longer ones (Meyer, 2015). Based on this type of learning, agencies have begun to specifically encourage sending of relevant information via Twitter. For example by educating 'superspreaders' (users with over 1000 followers), the system of disaster preparedness and recovery can be greatly enhanced (Yeager et al. 2015).

The use of 'people as environmental sensors' has been successfully demonstrated in the context of climate change. Kirilenko, Molodtsova and Stepchenkova (2015) analysed the climate change discourse evident from over 2 million tweets in 157 cities in the United States. It was found that both deviations from 'normal temperatures' and climate change coverage in the mass media had a significant influence on the number of climate change-related tweets and their specific content. Elsewhere, and to replace costly surveys, researchers used photo imagery uploaded on a photo-sharing website (Flickr) to derive a measure of visitation to lakes. The resulting metric of 'photo-user-days' was then used to develop models of visitation in relation to water clarity and travel distance (Keeler et al., 2015). Twitter data have also been used in explicit tourism contexts, where tourists' perceptions, concerns about destinations or knowledge were analysed in real time to understand recent and future travel motivations (Claster, Pardo, Cooper and Tajeddini, 2013). To better understand service delivery, Philander and Zhong (2016) analysed all tweets that related to tourist resorts in Las Vegas and determined satisfaction via sentiment analysis. As such, the analysis of Big Data from visitors now allows studying the flow of information from visitors back to tourism providers at much greater scale and detail than previously possible.

An increasing application of Big Data analytics is sentiment analysis, which is a process of algorithmically identifying and categorizing opinions expressed in text to detect attitude toward the object, event, or person (Kasper and Vela, 2011). In other words, sentiment analysis helps to analyse feelings or emotions that underpin written or spoken language. Sentiment analysis has attracted significant attention in the literature and has been extensively investigated in recent years, particularly for the English language but also for other languages (e.g. German, see Kasper and Vela, 2011). Currently, existing approaches can be split into two main groups: methods based on the combination of lexical resources and *Natural Language Processing* techniques; and machine learning approaches. The social media, for example Twitter, due to the huge volume and velocity of data puts serious challenges to practical applications of sentiment analysis, and in such cases a comprehensive high quality lexicon is essential for fast and accurate sentiment analysis. A lexicon is a list of lexical features (e.g., words) which are generally labelled according to their semantic orientation as either positive or negative. Manually creating and validating such lists is one of the most time-consuming tasks. For this reason in the first instance, most of the applied research covering sentiment analysis relies on existing manually constructed lexicons. Sentiment analysis is looking beyond the number of *Likes* or *Comments* that certain product releases, services, or blog posts attract. Instead, it is a method that aims to understand *how* people are responding to a particular situation or event.

Typically, sentiment analysis of tweets has been coded into positive, negative or neutral (Philander and Zhong, 2016). Claster et al. (2013) described a sentiment analysis developed for tourism purposes. They used a classifier that was trained based on large volumes of text, which had also been evaluated by humans. Testing against human assessment is critical as this is what the algorithm seeks to replicate at large scale. The algorithm improves over time via deep (machine) learning. Accuracy of the computer-based classifier depends, and was 41.9% for a sample of Thai students based on tweets about Bangkok and Phuket. Other studies found higher levels of accuracy, for example for hotel reviews (Kasper, 2011) reported about 66% of accuracy. It is important to highlight that the challenge of ‘calculating a sentiment’ is easier when the data to be evaluated refer to specific and clearly identifiable targets. For example, data retrieved from tourism consumer review platforms (e.g. TripAdvisor) that explicitly discuss the service associated with identified elements of tourism (e.g. a hotel) are more likely to achieve high accuracy rates (see Kasper and Vela, 2011).

The changing condition of the GBR differs from disasters or mass media coverage of extreme events (e.g. climate change impacts, Kirilenko et al., 2015) in that time scales of change are slower and processes of resilience are more difficult to study (Becken, 2013). The use case also differs from application in the area of market intelligence and assessment of tourist satisfaction because information sought requires specific feedback on the Reef (e.g. sightings of particular species) rather than generic 'impressions' or emotions. It is therefore important to test whether micro-blogs can be useful to detecting such specific and slower changes. Thus, this project investigates the extent to which micro-blogs (e.g. Twitter) and other publicly available social media data could be harnessed to monitor and predict environmental changes. If successful in principle, GBRMPA could consider expanding its current monitoring systems to capture the additional information provided by 'human sensors' to enhance monitoring of conditions of the GBR.

METHODOLOGY

To answer the research question, several methodological steps are necessary. The first and basic step is to secure access to Twitter and download tweets (or a sample thereof) into a database for analysis. There are approximately 320 million monthly active users worldwide and about 2.8 million users in Australia (Social Media News, 2016). These users post about 500 million tweets per day (Twitter, 2016). Based on data for the month of April 2016, which we are currently analyzing in our database, we estimate that we can expect about 250,000 tweets per day originating from the GBR area. These can be sent by local users or by those visiting from other places within Australia or internationally. These numbers clearly indicate that sufficient tweets are available to be evaluated, acknowledging that only a small proportion of tweets will discuss the condition of the GBR. The analysis of tweets is twofold, and based on word counts (for keywords) and sentiment analysis. Both can be helpful in extracting information on environmental and aesthetic conditions of the Great Barrier Reef.

Micro-blog data are then integrated in several consecutive steps with other data, including Open data (e.g. weather data made publicly available by Australia's Bureau of Meteorology), and where possible data sources provided by various stakeholders and agencies. The methodological flow chart of our approach is visualised in Figure 1.

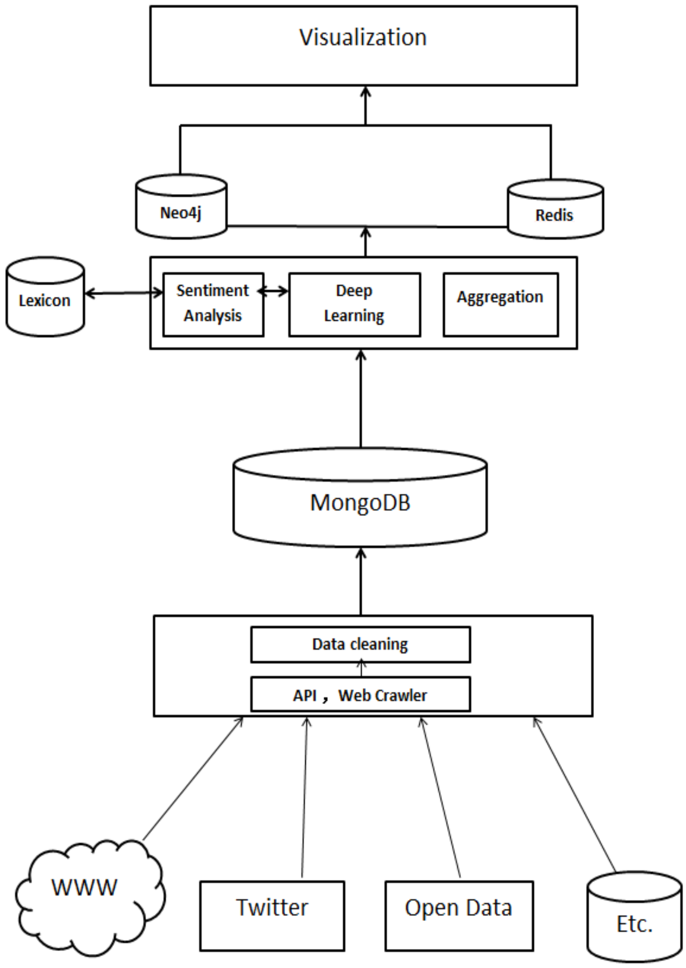

Figure 1 Methodological flow chart for collecting and analyzing data, and ultimately visualizing useful outputs for the end user.

In addition to storing tweets, further data are downloaded by crawling the web to access certain pages of interest, for example tourism service providers' pages (e.g. diving operators) which feature discussion blogs. Some data can be accessed 'per request' through different Application Programming Interfaces (API), such as in the case for accessing Eye on the Reef Program data collected and provided by the GBRMPA. Open data can also be accessed per request or periodically loaded onto our system for local access and processing (Faichney and Stantic, 2015). Data are locally stored in a MongoDB NoSQL document database, which is chosen as it is able to store and efficiently access a diversity of unstructured data. To simplify management, access and aggregation data are stored in individual collections organized by day. Furthermore, in order to ensure efficient access, data are processed, aggregated and stored in the Redis in-memory key value NoSQL database, which can serve web applications at very fast speed. High speed is particularly required for real time visualisation of data. In addition, we utilize Neo4j, which is a highly scalable native graph database that leverages data relationships as first-class entities. In Neo4j everything is stored in form of an edge, a node or an attribute. Each node and edge can have any number of attributes. Both the nodes and edges can be labelled, and labels can then be used to narrow the searches. This ensures efficient management of attributes and their relationships (i.e. which words are typically connected in text), which is particularly valuable in applications for sentiment analysis.

The analysis of Big Data from social media, such as Twitter, creates serious challenges to the practical applications of sentiment analysis. The key data characteristic that relate to volume, velocity, variety and veracity (Stantic and Pokorný, 2014) make traditional approaches to analysis not applicable, and it is not a trivial task to extract valuable information. In order to address these challenges there is a need to scale up both infrastructure and techniques. With regard to infrastructure we opted for horizontally scaling a so-called 'share nothing Hadoop style computing cluster', which refers to a distributed computing set up in which nodes (a computer) are connected but do not share memory or disk storage. In other words, the nodes are independent from each other, but can add up to almost infinite capacity – allowing users to store and process large volumes of data in real time. With regard to methods for sentiment analysis we accepted methods based on lexicons, however, we added features of machine (deep) learning to align positive or negative weights for particular words and phrases overt the time. The high volume and velocity of data reflects a large numbers of opinions that need to be integrated to make sense. In some cases, the information sought may not be reflected in the data and other sources need to be explored, such as Open weather data (see Figure 1 earlier).

Thus, in this research, the process of extracting sentiment from micro-blogs or other sources consists of several steps. First, it is important to identify the *Holder*; that is the person who is actually posting a microblog. This is important because we need to identify the location from where the tweet was posted (in our analysis it is mandatory that the tweet is sent from within the GBR region), and second to identify the user's profile. The profile will become important to establish decision rules on who could be classified as an 'expert' on Reef matters, who is a local and who is a visitor. This classification can then be used to implement a credibility weighting. The next step is to identify the *Target* (Qiu et al., 2011). The target tells us what the microblog is about, for example it could be about 'coral', or it could be about a visit to a restaurant in the GBR region, in which case it would be discarded. Finally there is need to identify the *Aspect*, which refers to the specific element of the target that the microblog is talking about, for example 'bleaching' of corals or 'colourful fish'.

All three steps are necessary to then derive by means of an algorithm the sentiment of the tweet, for example positive, neutral or negative, which is calculated by complex algorithms that uses the specified lexicon. Our system also relies on 'Deep Learning' methods as it learns over time from correct sentiment or failures by analysing the prediction with clearly identifiable sentiment and other measures. Through this process we continuously update the lexicon with correct weights that specify positive or negative contexts.

FINDINGS AND DISCUSSION

Assessment of sentiment tools

At the start of the project, a large-scale benchmark study of Twitter sentiment analysis methods was performed. More specifically, we examined the sentiment polarity classification performance of a number of existing methods. Out of all considered methods we paid particular attention to several recently proposed ones, including VADER (Gilbert, 2014) and a method on meta features (Canuto et al., 2016). The results revealed that method performances varied considerably across different domains of evaluations which obviously suggest presence of a domain interaction on method performance. This is because different words and also phrases in lexicon have different meaning and weight depending of domain. For example, the word 'lazy' could have different weight for positive and negative sentiment for tourism domain when compared to the business domain. Therefore, we decided to develop our own lexicon based method, which will be suitable to any domain as it will learn over the time and align weights in the lexicon according to the meaning relevant in a specific domain. As a starting point, and for preliminary analysis, we followed the literature and accepted publically available lexicons, which have been created by humans and tested previously in various contexts.

Preliminary analysis of GBR-related tweets

An initial structure to identify filter words and keywords was developed based on a) marine species, b) locations that indicate a visit to the GBR, c) tourist activities, d) risks to the Reef. These were extracted by reading initial sub-sample of tweets and by consulting a marine scientist. The structure, including examples of keywords and tweets is shown in Table 1.

Table 1 Overview of preliminary analysis of tweets to develop a structure

Category	Words	Sample tweet
Species	Wrasse, shark, turtle, clownfish/nemo/ Amphiprion ocellaris, coral	Hungry fish feat. glimpse of cute sea turtle ðŸ• ðŸ• ¢ #diveaustralia #yongala @ SS Yongala Dive Site https://t.co/PyBn5lgkSY
Location	Cairns, Townsville, Magnetic Island, Hamilton	#TBT Great Barrier Reef adventures. My favourite day. #cairns #greatbarrierreef

	Island, Whitehaven Beach	#visitqueenslandâ€¦ https://t.co/0WZy6SXmxu
Activities	Diving, snorkeling, boat trip, scuba, swimming	Star fish #scuba #greatbarrierreef #australia @ Great Barrier Reef (Australia) https://t.co/qkR1CQWEAZ
Risk	Bleaching, broken, algae, polluted, visibility, cyclone, damage	Tourist body downplays reef bleaching: The Great Barrier Reef is in the midst of its worst coral bleaching eve... https://t.co/vI5RkmBlyo

Using the Neo4j graph database, we have begun to store words from tweets as nodes and edges to represent relationship between the words. This type of analysis allows us to identify how many times specific words have been mentioned in the same microblog within a specific time frame (April 2016 data, see earlier). As can be seen in Figure 2, the word ‘coral’ is frequently related to several other words, namely ‘fish’, ‘nice’, and ‘bleaching’.

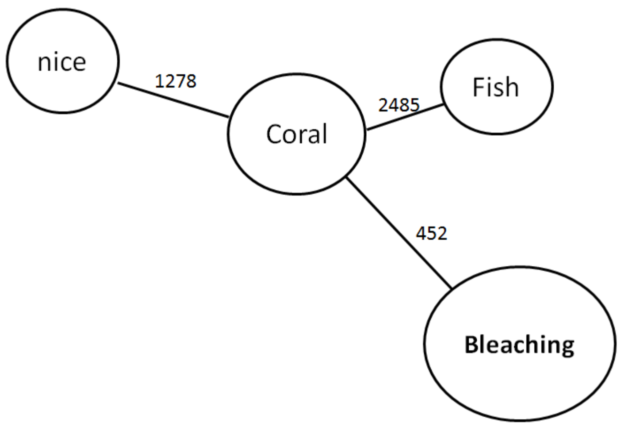

Figure 2. Representation of data in Neo4j graph database – example of the node ‘coral’.

The preliminary assessment of tweets has already brought to the light a number of challenges and insights:

1) Twitter data are ‘noisy’ and careful filtering is required. The sample is large enough that – with increasingly refined methods – it is possible to extract valid tweets;
2) Only about 15% of tweets are geocoded. However, it is possible, with relatively great accuracy to determine the location of non-geocoded tweets;
3) Visitors to the Reef large comment on the Reef at a generic level (e.g. “the coral was great”); they lack the experience and knowledge to articulate observations in a scientifically meaningful way.
4) Sentiment calculations will be achievable and provide useful information on a generic ‘satisfaction’ with the Reef and aspects of it.

As a result of some of the above challenges, we decided to consider additional data sources, including commercial Facebook pages from tourism operators, expert blogs from divers or other recreationists frequently visiting the reef, and data collected through other means of citizen science. Figure 2 provides an overview of how these different types of data can be used to develop an overall monitoring tool. Importantly, biophysical monitoring data are required to verify the observations made by visitors and shared through Internet-based channels. For example, data from remote sensing that provides an indication of dissolved sediment (resulting in turbidity) can be compared with visitors' observation on water quality and visibility, specified by location and time. Similarly, micro-blog data and other Internet-sources can be verified against more formalized citizen science approaches, such as the Eye on the Reef or various coral monitoring programs. In addition to verification, some biophysical will be used to calibrate tourist sentiment. For example, weather conditions are likely to impact on how visitors experience the Reef, and sentiment needs to be adjusted accordingly to extract the component that actually reflects Reef conditions and not the impacts of temperature, rain or wind, amongst others.

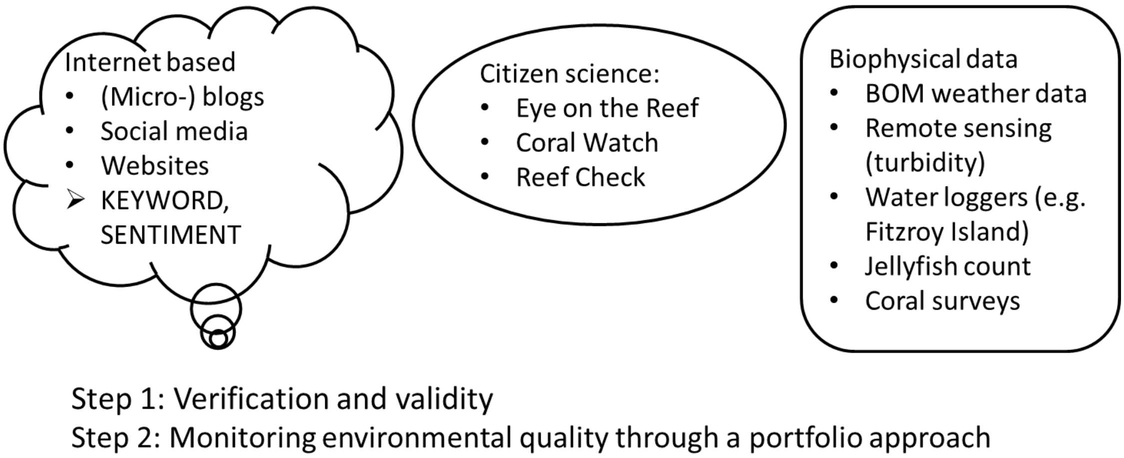

Figure 3 Advanced specification of different types of data sources needed for the development of an enhanced monitoring tool.

<u>Advances in computer science</u>
The research also advances computer science processes that are refined to maximise the efficiency and accuracy of consolidated data to enable real time extraction of valuable information and deep learning from Big Data. Specifically, we are working on advancing data structures that are required for efficient management of such huge volume and variety of data. Additionally, this research makes a contribution to developing new algorithms for sentiment calculation and also for automatic creation of lexicons, by utilizing deep

learning. To automatically create dedicated lexicons for specific domains we use a large set of tweets, which contains noisy opinion (positive and negative). In particular we are using tweets that contain both individual words and multi-word expressions (e.g. ‘bad’ indicating negative sentiment and ‘not bad’ indicating positive impressions) as features, which then become entries in the lexicon. The weights from the learned model are then used to define which lexicon items to keep. The scores from new lexicon items are then used to encode new tweets and they are further used to derive more advanced features. Sentiment labels are also automatically inferred from emoticons contained in tweets. The advantage of our approach is that it is easy to extend the lexicon, as emoticons can be considered as fairly good indicators for the general sentiment expressed in a tweet. Tweets with positive emoticons, like ’:)’, are assumed to be positive, while tweets with negative emotions, like ’:(’, are labelled as negative.

Next steps and further development

To date the research has shown that analysis of micro-blogs can provide useful insights into visitors perceptions of an environmental asset, but the level of detail on specific environmental conditions might be insufficient and needs to be extracted from other sources. However, the analysis of sentiment in particular has considerable ancillary benefits, for example for destination management and marketing.

Micro-blog and other Internet-based data can become part of a destination knowledge management system. Such an approach was discussed by Fuchs, Höpken and Lexhagen (2014) for a Swedish destination as part of a ‘learning destination’ project. Fuchs et al. (2014) argued that the competitiveness of a destination depends on new knowledge creation and knowledge application – both of which are increasingly supported by information and communication technologies (ICT). The early findings presented here provide insights into another jigsaw puzzle piece that can be added to evolving destination or city management systems that increasingly integrate various types of data to improve information flows and knowledge. As a next step we plan to perform a large-scale analysis that takes into consideration a greater diversity of data. This amongst others will allow us to compare our findings with biophysical data and accordingly modify our algorithms in order to compensate errors. Once our predictions are aligned with biophysical data we will be able with confidence to transfer findings to other locations and make predictions of environmental conditions.

CONCLUSION

The research presents a framework for harnessing information provided by human sensors, in this case visitors to the GBR who share observations via micro-blogs. In this paper, we discussed data requirements, computer infrastructure and analytical techniques. Initial finding indicate that microblogs can provide valuable information with regard to number of people in certain areas, their profile, and also their sentiment about certain targets and aspects. However, it is evident that more specific data in form of blogs, social media pages, and open data can offer additional valuable information. Our initial findings provide confidence in the applicability of harnessing tourists as environmental monitors, especially when ancillary benefits for tourist destinations are realised. Using an innovative Internet-based approach to monitoring environmental assets that are heavily used by tourists provides a positive way forward towards protecting the environment and understanding the visitor experience simultaneously. The GBR, with its iconic status, high visitation rates yet deteriorating environmental conditions, provides a perfect test case for this approach.

REFERENCES

Abbasi, A., Hassan, A. and Dhar, M. (2014) May. Benchmarking Twitter Sentiment Analysis Tools. In Proceedings of the International Conference on Language Resources and Evaluation (LREC) (Pp. 823-829), May 26-31, 2014, Reykjavik (Iceland). Available (22/05/16) http://www.lrec-conf.org/proceedings/lrec2014/pdf/483_Paper.pdf

Addison, P., Walshe, T., Sweatman, H., Jonker, M., Anthony, K., MacNeil, A., Thompson, A. and Logan, M. (2015) Towards an integrated monitoring program: Identifying indicators and existing monitoring programs to effectively evaluate the Long Term Sustainability Plan. Townsville, Queensland: Australian Institute of Marine Science.

Becken, S. (2013) Developing a framework for assessing resilience of tourism sub-systems to climatic factors. Annals of Tourism Research, 43, 506–528.

Becken, S., McLennan, C. and Moyle, B. (2014) World Heritage Area at Risk? Resident and Stakeholder Perceptions of the Great Barrier Reef in Gladstone, Australia. Griffith Institute for Tourism Research Report Series, Report No 2, Griffith University, Gold Coast, Australia.

Bohannon, J. (2016) Twitter can predict hurricane damage as well as emergency agencies. Science, Mar. 11, 2016. Available (15/05/16) http://www.sciencemag.org/news/2016/03/Twitter-can-predict-hurricane-damage-well-emergency-agencies

Canuto, S., Goncalves, M.A. and Benevenuto, F. (2016, Exploiting New Sentiment-Based Meta-level Features for Effective Sentiment Analysis' Proceedings of the Ninth ACM International Conference on Web Search and Data Mining WSDM '16, 53-62.

Claster, W., Pardo, P., Cooper, M. and Tajeddini, K. (2013) Tourism, travel and tweets: algorithmic text analysis methodologies in tourism. Middle Eastern Journal of Management, 1(1), 81-100.

Day, J. (2013) Operationalising the Outstanding Universal Value of the Great Barrier Reef World Heritage Area: addressing some challenges raised by the World Heritage Committee. (Pp. 118-127). In Figgis, P., Leverington, A., Mackay, R., Maclean, A. & Valentine, P. (eds). (2012). Keeping the Outstanding Exceptional: The Future of World Heritage in Australia. Australian Committee for IUCN: Sydney.

De'ath, G., Fabricius, K.E., Sweatman, H. and Puotinen, M. (2012) The 27–year decline of coral cover on the Great Barrier Reef and its causes. Proceedings of the National Academy of Sciences, 109(44), 17995-17999.

Faichney, J. and Stantic, B. (2015) A Novel Framework to Describe Technical Accessibility of Open Data, The First International Conference on Big Data, Small Data, Linked Data and Open Data – ALLDATA15, 1-12.

Fuchs, M., Höpken, W. and Lexhange, M. (2014). Big data analytics for knowledge generation in tourism destinations – A case from Sweden. Journal of Destination Marketing & Management, 3, 198-209.

Gilbert, C.H.E. (2014) VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. Eighth International Conference on Weblogs and Social Media (ICWSM-14). Available at (20/04/16) http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf

Great Barrier Reef Marine Park Authority [GBRMPA] (2014) Great Barrier Reef Outlook Report 2014. Townsville, Queensland: Great Barrier Reef Marine Park Authority.

Grech, A., Bos, M., Brodie, J., Coles, R., Dale, A., Gilbert, R., Hamann, M., Marsh, H., Neil, K., Pressey, R. L., Rasheed, M. A., Sheaves, M. & Smith, A. (2013). Guiding principles for the improved governance of port and shipping impacts in the Great Barrier Reef. Marine Pollution Bulletin, 75(1–2), 8-20.

Kasper, W. and Vela, M. (2011) Sentiment Analysis for Hotel Reviews, Proceedings of the Computational Linguistics-Applications Conference, 45–52. Available (12/04/16) file:///C:/Users/s2825673/Downloads/25.pdf

Keeler, B.L., Wood, S.A., Polasky, S., Kling, C., Filstrup, C.T. and Downing, J.A. (2015) Recreational demand for clean water: evidence from geotagged photographs by visitors to lakes. The Ecological Society of Marica, 13(2), 76-81.

Kirilenko, A.P. Molodtsova, T. and Stepchenkova, S.O. (2015) People as sensors: Mass media and local temperature influence climate change discussion on Twitter. Global Environmental Change, 30, 92-100.

Liburd, J. and Becken, S. (2016, forthcoming). Multi-scale Stewardship Alliances Protecting Nature, Tourism and UNESCO World Heritage. Submitted to Journal of Sustainable Tourism.

Meyer, R. (2015) How the USGS Detects Earthquakes Using Twitter. The Atlantic. Available (15/05/16) http://www.theatlantic.com/technology/archive/2015/10/how-the-usgs-detects-earthquakes-using-Twitter/409909/

Mustika, P., Stoeckl, N. and Farr, M. (2015) The potential implications of environmental deterioration on business and non-business visitor expenditures in a natural setting: a case study of Australia's Great Barrier Reef. Tourism Economics. DOI: http://dx.doi.org/10.5367/te.2014.0433

Philander, K. and Zhong, Y.Y. (2016) Twitter sentiment analysis: Capturing sentiment from integrated resort tweets. International Journal of Hospitality Management, 55, 16-24.

Qiu, G., Liu, B., Bu, J. and Chen, C. (2011) Opinion word expansion and target extraction through double propagation. Computational Linguistics, 37(1), 9-27.

Shafer, C.S. and Inglis, G.J. (2000) Influence of social, biophysical, and managerial conditions on tourism experiences within the Great Barrier Reef World Heritage Area. Environmental Management, 26(1), 73-87.

Social Media News (2016) Social Media Statistics Australia – January 2016. Available (01/05/16) http://www.socialmedianews.com.au/social-media-statistics-australia-january-2016/

Stantic, B. and Pokorný, J. (2014) Opportunities in Big Data Management and Processing. Frontiers in Artificial Intelligence and Appl., 270, 15–26.

Twitter (2016) Statistics. Available (12/05/16) http://www.internetlivestats.com/Twitter-statistics/

Yeager, V., Cooper, G.P., Burkle, F.M. and Subbarao, I. (2015) Twitter as a Potential Disaster Risk Reduction Tool. Part IV: Competency-based Education and Training Guidelines to Promote Community Resiliency. PLoS Currents. doi: 10.1371/currents.dis.ce3fad537bd666770a649a076ee71ba4