Ecological Informatics 52 (2019) 35–45

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf

A hybrid is born: Integrating collective sensing, citizen science and professional monitoring of the environment

Susanne Becken[a,d,*], Rod M. Connolly[b], Jinyan Chen[a], Bela Stantic[c]

[a] Griffith Institute for Tourism, Griffith Business School, Griffith University, Gold Coast 4222, QLD, Australia
[b] Australian Rivers Institute, Coast and Estuaries, and School of Environment and Science, Griffith University, Gold Coast 4222, QLD, Australia
[c] Institute for Integrated and Intelligent Systems, Griffith Sciences, Griffith University, Gold Coast 4222, QLD, Australia
[d] School of Hospitality and Tourism Management, University of Surrey, Guildford, United Kingdom

ARTICLE INFO

Keywords:
Citizen science
Collective sensing
Monitoring
Hybrid system
Coral
Great barrier reef

ABSTRACT

Including members of the public in the development of effective environmental monitoring systems is gaining traction. This research assesses the potential for a hybrid monitoring system for the case of coral at the Great Barrier Reef. Based on a review of citizen-derived data sources, the paper first develops a framework and then populates it with five datasets. These are then compared based on data volumes, type of data, spatial coverage, and bleaching patterns. The results reveal the inherent difficulties – both in terms of quantity and quality – for collective sensing data (Twitter in this case) and more structured human sensors approaches (Eye on the Reef Sightings). However, more targeted approaches, such as CoralWatch and tourism-operator based data collection, emerged as important contributors to information generation on the state of coral. Citizen-based data that either deliver a high data density per location, a wide geographic coverage, or regular observations over time are particularly valuable. Recommendations are made for developing a hybrid monitoring system that integrates citizen-derived with professionally collected data.

1. Introduction

The monitoring of environmental change has become increasingly important for environmental managers at a time of accumulating and accelerating stress on ecosystems. However, scientific monitoring is costly, especially if vast areas are to be covered over long periods. To facilitate cost-effective data collection at greater scale, members of the public are encouraged to contribute observations to centralised databases, often managed by scientists or non-governmental organisations. The potential benefits of data collection through citizen science are broadly accepted (Lodia and Tardin, 2018; McKinley et al., 2017; Tiagoa et al., 2017); however, further evidence is needed to assess how well these types of data complement, or integrate with, more professional monitoring systems.

Citizen science programs refer "to the inclusion of members of the public in some aspect of scientific research" (Eitzel et al., 2017, p. 1) can have different levels of order and structuring (Welvaert and Caley, 2016). Structured programs provide observations or perceptions specific to particular locations and times, using a purpose-built monitoring and reporting tool. The majority of studies to date focused on structured citizen science programs (e.g. Theobald et al., 2015). In contrast, information supplied unwittingly, such as that related to social media communication, generates unstructured data. Social media information is often only indirectly relevant to the particular monitoring interest. However, with the appropriate filtering mechanisms these unstructured data can potentially generate useful insights with a high geo-temporal resolution (Becken et al., 2017). Indeed, Daume and Galaz (2016) refer to twitter conversations as "embryonic citizen science communities" (p. e0151387).

Citizen science platforms and social media sharing have benefitted from substantial progress in information technology and Internet availability, and the transmission of digital data 'from the field' has become possible at large scale. This has led to an increasing interest in less structured 'crowd sourced' data that are transmitted at a high-volume and velocity. Twitter data has attracted considerable research activity (for a review see Steiger et al., 2015), in particular with a focus on event detection (predominantly acute crisis and disaster, e.g. Vivacqua and Borges, 2012). A small number of researchers have explored the usefulness of Twitter posts for environmental monitoring (e.g. Becken et al., 2017; Daume, 2016; Daume and Galaz, 2016).

The opportunity to collect data from members of the public is particularly pertinent for places that face rapid environmental change, and

* Corresponding author.
E-mail address: s.becken@griffith.edu.au (S. Becken).

https://doi.org/10.1016/j.ecoinf.2019.05.001
Received 12 June 2018; Received in revised form 1 May 2019; Accepted 2 May 2019
Available online 03 May 2019
1574-9541/

that – at the same time – are visited by large numbers of people that have the potential to deliver a high volume of data (ElQadi et al., 2017). The Great Barrier Reef (GBR) in Australia is an example of a natural asset that faces environmental decline, yet is a major tourist destination visited by over 2 million people each year (Becken et al., 2017). Recognising the urgent need for a comprehensive and integrated monitoring program, the role of citizen science has been acknowledged and has explicitly been advocated (Addison et al., 2015). Hence, the Great Barrier Reef Marine Park Authority (GBRMPA) is assessing the role of citizen science programs as part of the wider goal of ensuring adequate and cost-effective coverage to help evaluate progress towards long-term sustainability targets as outlined in the Reef 2050 Plan (Department of Environment, 2015).

Our research therefore aims to increase the understanding of how traditional monitoring programs can work alongside, or be integrated with, a range of citizen science programs to enhance the overall monitoring capacity in a cost-effective way. We draw on multiple data sets related to the GBR as a case study, contrasting and comparing the spatial coverage and environmental information they contain We have specifically focussed on coral monitoring, but future research could seek to compare other aspects of GBR monitoring, for example of particular fish species, cetaceans, dugongs, water quality, beach erosion or environmental incidents.

This research has three objectives: 1) to review the broad range of citizen-based data collection for environmental monitoring, 2) to understand the availability and nature of multiple types of data sources that provide information on the state of coral at the GBR; and 3) to compare the spatial coverage and the bleaching patterns of the GBR based on the different data sources. The paper concludes by making recommendations for developing a hybrid monitoring system that integrates citizen-derived with professionally collected data.

1.1. Categories of environmental monitoring

Environmental monitoring typically requires the expertise of trained scientists who understand the details of the ecosystem they are observing. Expert knowledge ensures that changes are recognised and interpreted correctly. Some monitoring requires precise measurement (e.g. water temperature, salinity, turbidity) using specific technology and instruments, whereas other monitoring relies on human (expert) assessment of environmental conditions (e.g. coral cover, species abundance).

Hiring professional staff is expensive, and transporting them to monitoring sites, often in remote areas, is costly and at times logistically difficult. It has therefore become increasingly attractive to involve members of the public in monitoring activities. Building on a tradition of utilising lay observations of the environment (Kearns et al., 2003), an increasing number of programs have formalised public involvement. Collecting data involves three elements: the observation, time, and geographic location (Resch, 2013). Several authors refer to these elements as 'volunteered geographic content' (Connors et al., 2012). To date, such volunteered data has mainly informed research projects related to biodiversity and conservation biology (e.g. McKinley et al., 2017), and has been less prominent in the context of monitoring environmental change (Connors et al., 2012).

The literature uses multiple terms to describe approaches that involve members of the public, including 'citizen science', 'citizen surveillance', 'human sensing', and 'crowdsourcing'. Often these terms are used inconsistently (or as subsets of each other) and there is limited consensus in relation to how different approaches could be classified within one conceptual framework (relevant work is presented in See et al., 2016 and Liu et al., 2015). This paper provides an attempt towards a classification based on a coherent set of variables, and specifically for the purpose of environmental monitoring. We note, however, that the field is "in a state of flux" (Eitzel et al., 2017, p.1) and involves researchers from many disciplines, who bring with them different definitions and research motivations (e.g. Boulos et al., 2011). Moreover, this paper acknowledges – but will not resolve – the ongoing discussion on whether citizen science (broadly or with its sub-forms) is a tool, a specific research method, or a new way of developing or implementing research collaborations (Eitzel et al., 2017).

Despite conceptual and definitional challenges, we suggest a framework that serves as an underlying logic for this present paper. Our classification begins by differentiating structured from unstructured approaches (Welvaert and Caley, 2016). More specifically, Welvaert and Caley argue that reporting can be intentional or unintentional (i.e. involuntary), and detection of observations can be controlled (e.g. through a rigorous framework) or opportunistic. Social media, for example, generates unintentionally supplied data that are collected with an opportunistic detection method. The different types of approaches discussed in the literature cover the whole spectrum of these two dimensions. In addition, existing approaches differ in terms of their resource requirements, the nature of data, potential for bias, the types of people involved in data generation, training needs and expertise, and data processing (Table 1).

1.2. Collective sensing

Collective Sensing – also referred to as crowdsourcing[1] by Welvaert and Caley (2016), passive crowdsourcing by See et al. (2016) or e-participation by Gharesifard et al. (2017) – draws on aggregated data that stem from an open source network, typically accessed via mobile technology. Liu et al. (2015) introduce the term social sensing, referring to the use of big data with a spatial reference and socio-economic context (alluding to the established remote sensing). However, different to this study Liu et al. do not differentiate between unintentional (i.e. collective) and intentional sensing (i.e. human sensing, see below). Collective sensing data could originate, for example, from mobile phone usage or transport smart card patterns (Liu et al., 2015), but the most commonly used sources evident in the scientific literature comes from social media. In particular, Twitter and Flickr data have been used widely because they can be accessed without a password and they provide actual content beyond geographic coordinates. Twitter is a microblogging platform where users can post short messages, photos or short video clips. In 2017, there were about 330 million monthly active users of Twitter (Statistica, 2018). Twitter releases the equivalent of 1% of the total tweets free of cost to researchers, who can choose a random sampling approach for data collection using specified filtering mechanisms. Flickr is an online photo management and sharing application that is commonly used by travellers and photographers.

For both Twitter and Flickr, users share information for a range of purposes that are unrelated to subsequent research questions. Thus, the analysis of data contained in posts or photographs occurs without the knowledge or intent of the provider. Subject expertise is coincidental; instead data provide insights into what 'matters to people' at a particular location and time. Since these types of social media posts are often geo-time stamped, it is possible to extract geographic information in combination with specific content of interest (e.g. Barve, 2014).

Earlier research noted that people are more likely to share positive content (Alaei et al., 2017; Mitchell et al., 2013) or experiences that are outside the norm. Unusual weather events, for example, feature more often in social media than expected weather (Hyvarinen and Saltikoff, 2010). Thus, whilst not being useful as a monitoring vehicle for every day conditions, social media posts may add detail to scientific records

[1] Note that Boulos et al. (2011) refer to crowdsourcing as any type of data provided by the general public, both intentionally (e.g. by using smart phone apps) or unintentionally (e.g. Twitter). In other words, they combine the two categories of collective and human sensing, as shown in Table 1. Instead they specifically highlight the approach of 'participatory sensing' where mobile phones are used as part of a (urban) network, for example to detect traffic.

Table 1
Categories of environmental monitoring and key differences (Sources: after Resch (2013) and Welvaert and Caley (2016)).

	Collective sensing	Human sensors	Citizen science	Professional monitoring
Intentional reporting	Unintentional	Intentional	Intentional	Professionally
Detection mechanism	Unstructured	Somewhat structured	Structured	Highly structured
Expertise/knowledge	None required	None to some	Some to high	Considerable
Cost and infrastructure	Low, but requirement for IT infrastructure	Moderate, investment into data collection mechanism	High, as training and supervision required	Very high
Data volumes	Potentially very large	Large	Low to medium	Low
Time-space resolution	Depends on several factors	Depends on several factors	Low to medium	Typically very low
Data quality	Poor	Poor to medium	Medium	High

for extreme events. The implication of such information is that observations posted on social media are not representative of all types of conditions, and the fact that a particular state is not commented on does not mean it did not occur.

Given that social media user profiles also contain a significant amount of metadata it is possible to provide some context around posts (e.g. past behaviour, favourite topics, location of residence), and this may help understand possible sources of bias and levels of expertise. In this sense, social media could be superior to other more structured approaches that often collect very little information on the providers of data (See et al., 2016). Some projects are not primarily interested in the content provided but only focus on the related metadata. Mapping Ocean Wealth, for example, is a project by The Nature Conservancy (2017) that uses the locations of photographs posted on Flickr to derive estimates of visitation to coral reefs, and associated economic impact.

Due to the open source nature of many collective sensing data, no investment is required in producing them; however, there is a cost in developing a computing system that can access, download, store and process the large volume of data. Data cleaning and filtering systems are required to eliminate redundant or irrelevant data. Despite such processes, the information extracted from collective sensing often lacks accuracy and reliability, but as stated by Hyvarinen and Saltikoff (2010), the volumes are potentially large and the benefit of real time information compensates for some of the data shortcomings. Thus, despite considerable uncertainty it may still be possible to draw important conclusions from the data.

1.3. Human sensors

Increasingly, organisations obtain data from people who are willing to voluntarily provide information for a particular purpose. Eitzel et al. (2017) define a human sensor as an “individual who is part of a network by sending data and observations that are often taken and transmitted via modern communication tools, like smartphones, to a central database” (p. 1). Often, specialised web apps help collect the relevant data, but observation sheets and clip board-type approaches are also being used. In most applications, there is little or no expertise required in providing the information as systems rely on common sense and general types of human observations. Due to a predefined format for data collection, the quality is higher than collective sensing. Similar to collective sensing, an upfront investment into the digital infrastructure and computing system is required. Ongoing expertise is needed to maintain the system and analyse incoming data.

The ‘Did You Feel It?’ (DYFI) website developed by the U.S. Geological Survey (USGS) to understand the effect of earthquakes is a primary example of how human sensors can complement traditional measurements (Crooks et al., 2013). Thus, by sharing (often real-time) observations or measurements, people can contribute to targeted data collection. Another example is a dedicated Facebook-based citizen science group that reports sightings of cetaceans off the coast of Brazil. A comparison with scientifically collected data showed acceptable levels of correlation (Lodia and Tardin, 2018).

1.4. Citizen science

This approach involves an active involvement of citizens in the collection of scientific data (Resch, 2013), and as such it is quite different from both collective and human sensing. Eitzel et al. (2017) emphasise that this distinction is particularly relevant for funding agencies who need to understand that the ‘more engaged participatory approaches’ of citizen science are superior in their data quality outcomes. Indeed, citizen science is a targeted approach that requires significantly higher levels of involvement and expertise. McKinley et al. (2017) argue that citizen science does not differ substantially from professional science. They define it “…as the practice of engaging the public in a scientific project—a project that produces reliable data and information usable by scientists, decision makers, or the public and that is open to the same system of peer review that applies to conventional science” (p. 16). The contribution by citizens can be substantial. Theobald et al. (2015) reviewed 388 citizen science projects related to conservation and biodiversity. They found that collectively these involved 1.3 million volunteers, representing a value of US$2.5 billion in-kind every year.

Citizen science projects often involve some training, and this can take several shapes. Most often, citizens collect data as part of an existing and well-supported scientific project, for example in relation to birds, fish or butterflies, or water quality monitoring (Connors et al., 2012). One of the oldest examples of data collection involving citizens is Audubon's Christmas Bird Count, whereby volunteers have counted birds in the Western Hemisphere since 1900 (Connors et al., 2012). Another well-known example is eBird, which provides an online program for bird watchers to report and access information about bird abundance and distribution (Cornell Laboratory of Ornithology, 2018). Hyder et al. (2015) provide a review of marine citizen science projects and their relevance to specific policy areas of marine management.

A new form of citizen science project has evolved with increasing computing power. It involves the interpretation of large amounts of data delivered by personal computers or devices of citizens. For example, participants view and classify photographs of animals and their behaviours, taken by automated cameras. Snapshot Serengeti is one of the most prominent examples where 30,000 volunteers helped to identify rapidly images from 200 cameras installed in the park (Swanson, 2015).

1.5. Integrated approaches

A small number of projects have compared the reliability and accuracy of citizen-derived data with professional monitoring. Romana et al. (2017), for example, examined observations of street trees and found that generally the consistency between experts and citizens is high (e.g. 90% for site type, land use, dieback, and genus identification), but lower for a number of specific parameters (e.g. wood condition). They recommended to either drop overly complex variables or use additional techniques, such as photo collection. Similarly, Tiagoa et al. (2017) compared the data collected by citizens on reptile and amphibian species in Portugal with a scientific dataset. They concluded that the citizen science project delivered relatively accurate predictions of species distributions. In a marine context, the benefits of using citizen-supplied sightings to expand scientific databases was recognised by Lodia and Tardin (2018), who obtained records on dolphins and whales from citizens for areas that were not covered by their survey. Citizens also identified new species not previously present in the database.

Limited work has been published that compares several forms of citizen-based and scientific data. One exception is the OakMapper.org (Connors et al., 2012), which constitutes an online platform at the intersection of citizen science and crowdsourced data (i.e. collective sensing). OakMapper.org collects and visualises information on a disease called ‘sudden oak death’. The platform integrates and visualises official data monitoring with an iPhone application, and Flickr and Twitter data. Connors et al. conclude that using environmental monitoring data from a broad spectrum of sources and intentionality can be suitable for “other cases of highly visible environmental problems” (p. 1285). For citizen-supplied data to enhance scientific data, it is important to build or strengthen the connections between the different contributors to scientific enquiry (Theobald et al., 2015). This current paper interrogates and compares several types of data sources for the case of coral bleaching at the GBR in Australia to explore how a broad spectrum of citizen-based data can be integrated and calibrated with relevant scientific data.

2. Method

2.1. Context

Despite the GBR's status as a marine protected area and a UNESCO World Heritage site, and the considerable investment by the Australian Government into Reef management through the GBRMPA, the deterioration in Reef health has been ongoing. Over the last 30 years, the GBR has lost more than half its coral cover; a consequence of the deterioration in water quality, warming of water temperatures due to climate change, ocean acidification and the destructive impact of cyclones (Department of Environment, 2015; Hughes et al., 2017). In 2016 and 2017, the GBR was affected by the worst sequence of coral bleaching events on record. The most severely affected area in the 2016 event was the northern part of the GBR, with recent information indicating that 29% of shallow water corals died (GBRMPA, 2017a). Due to continuing warm water temperatures, bleaching continued in 2017, with the main impact in the central parts of the GBR.

There are several sets of data that contain information on the state of coral. For this study, we are drawing on Twitter, Eye on the Reef Sightings (Sightings, hereafter), CoralWatch, Tourism Weekly and the Reef Health and Impact Survey (Reef Health, hereafter). All data sets were imported into MongoDB NoSQL database, however, some analysis required to load some structured data into MYSQL relational database and harness the power of SQL language. The data are located on a cluster computer with a Hadoop Distributed File System. An overview of the measures and scales used in the comparison is provided in Table 2. Their collection and analysis are described below.

2.2. Twitter data

We used a public Twitter API with restrictions to capture geo-tagged tweets posted from a defined geographic area. Geo-tagged tweets are a subsample of tweets associated with explicit geographic coordinates measured by either an exact coordinate or an approximate polygon. To determine an approximate region of the GBR a rectangular bounding box was defined (Southwest coordinates: 141.459961, -25.582085 and Northeast coordinates: 153.544922, -10.69867). Whilst the bounding box does not perfectly overlay with the ‘Great Barrier Reef region’, it provides a sufficient filtering mechanisms to obtain relevant tweets (Becken et al., 2017).

Table 2
Overview of data sources and measures used to assess the state of coral. Note that for Tourism Weekly and Reef Health averaging of multiple observations gave a scaled response between 1 and 5.

Data source	Category	Measure used	Scale
Twitter	Collective sensing	Sentiment (polarity of text in tweet)	Calculated sentiment score transformed into a scale from 1 to 5. 1 = Most negative 3 = Neutral 5 = Most positive
Sightings	Human sensors/ Citizen science	Extent of bleaching	1 = Totally bleached white 2 = Bleached only on upper surface 3 = Pale light or yellow 4 = Fluorescing 5 = No bleaching
CoralWatch	Citizen science	Coral colour type and shade on a chart	1 = Bleached 2 = Lighter 3 = Medium-light 4 = Darker 5 = Darkest colour
Tourism Weekly	Citizen science	Bleaching observed	1 = Yes bleaching 5 = No bleaching
Reef Health	Professional monitoring	Bleaching observed	1 = Yes bleaching 5 = No bleaching

Further steps were necessary to obtain a suitable sample of tweets. The project only started in March, which is when data collection began, and Twitter posts could not be accessed for earlier periods. For the period of data collection from 18/03/2016 to 31/12/2016, a total of 275,324 tweets were retrieved. These were then filtered using keywords related to selected locations, marine activities, marine life, aspects of water quality, and coral condition (Becken et al., 2017). This resulted in the extraction of 12,400 tweets, which were further reduced to those tweets that mentioned the word ‘coral’. A manual examination revealed that some posts referred to a Queensland politician (i.e. Ms. Coralee O'Rourke) and “coral trout” as a seafood meal, resulting in the elimination of 23 tweets and a final number of 434 coral-related tweets.

Since Twitter data do not provide structured information on coral health, it was necessary to develop a proxy. Tweets mentioning coral were processed using Natural Language Processing (NLP) and sentiment analysis technologies to assess whether the post reflected a positive or negative perception. The sentiment algorithm used was modified Valence Aware Dictionary for Sentiment Reasoning (VADER) (Alaei et al., 2017), which is a rule-based model that combines a general lexicon and a series of intensifiers, punctuation transformation, emoticons, and many other heuristics to compute sentiment polarity of a review or text (Hutto and Gilbert, 2014). The output of sentiment analysis is a score that ranges between −1 (very negative) and +1 (very positive), with 0 indicating a neutral sentiment. In order to improve accuracy and performance we modified the original VADER algorithm and started a dedicated lexicon suitable for environmental changes.

In summary, for each coral-related tweet, three variables were recorded, namely time, location and sentiment score. To enable better comparison with the other data sources, the sentiment score was normalised to a scale from 1 to 5, with 3 indicating a neutral point and 5 being most positive (eq. $Y = 2*X + 3$ to normalize Twitter Sentiment from −1 to +1 to a scale from 1 to 5) (Table 2).

2.3. Sightings data

The Eye on the Reef program operated by the GBRMPA (2017b) includes a data collection mechanism for general users of the Reef, referred to as ‘Sightings’. Using a mobile app or online system, visitors report observations of particular species or environmental hazards. As with other programs, the information provided describes the particular subject of interest, the time and location. GBRMPA's Sightings data were attractive as they constitute an intermediate form of human sensors and citizen science (Table 1). The Sightings platform provides a relatively structured approach for people to report what they see underwater and what they feel was noteworthy. GBRMPA staff provided the data to the research team in an Excel spreadsheet, which has been imported into MySQl relational database.

The information contained in the Sightings database contains a wide range of species and incident variables. The data were filtered to obtain only those observations that reported some aspect of coral (including bleaching). The time and location of observation, along with an assessment of the severity of the incident were extracted. Severity was measured as follows: Totally bleached white (score of 1), bleached only on upper surface (2), pale light or yellow (3), fluorescing (4), no bleaching (5) (Table 2).

2.4. CoralWatch data

CoralWatch (2017) is a citizen science project based at the University of Queensland. It has been developed to engage non-scientists in Australia and elsewhere to not only appreciate coral reef management, but also to contribute by adding data into a tailored system. CoralWatch uses a Coral Health Chart to provide assistance to citizens to make decisions about the state and type of coral. The chart standardises coral colour as a guide to bleaching, and provides a simple way for people to quantify coral health and contribute to the database (Marshall et al., 2012).

Observations were obtained from CoralWatch in the format of an Excel spreadsheet. The following variables are of interest: i) Date and Time of observation, ii) Reef name, and geographic location (Longitude, Latitude), iii) Coral Colour (colour type and shade measured as 1 to 6, with 6 being darkest) (Table 2). To generate a scale in line with the other data sources, the darkest scores (6 and 5) were collapsed into a single category.

CoralWatch scores were then processed to analyse frequencies of observation by time and location, and extent of bleaching. In many cases, the surveyor provided multiple data points for the same geographic coordinates, indicating that they considered different types of coral (or aspects of one and the same coral) within one diving location. In those cases, an average score was calculated to generate one reliable bleaching score per location. Furthermore, it became apparent that diving locations were surveyed repeatedly throughout the year and by different people. Again, to avoid bias by overweighting certain locations, the average score was calculated for 2016.

2.5. Tourism Weekly data

GBRMPA's Eye on the Reef platform also provided the Tourism Weekly dataset, which involves marine tourism operators. Tourism Weekly monitoring requires training and commitment and clearly represents a citizen science approach. It requires more resources than the Sightings program (e.g. in terms of training) and is likely to produce better quality data since many of the divers who work for tourism companies have considerable expertise or even a professional background.

Tourism Weekly data are collected on a regular basis and for the same locations, which is defined by where tourism businesses have a license to operate. The survey collects detailed information on types of corals, but for this research the most simple indicator was used, namely bleaching “Yes” (quantified as 1) or bleaching “No” (quantified as 5) (Table 2). Each locations had multiple observations over the course of the year, and by averaging these we obtained a scaled response along the 1–5 scale.

2.6. Reef health data

The Reef Health data represent the most professional data set in this project. Reef Health involves a quantitative method that assesses reef health in a series of circular survey areas 10 m in diameter. Trials have shown that trained surveyors achieve similar results to each other, making this a robust method. The data are mainly collected by staff from the Queensland Parks and Wildlife Service, the Great Barrier Reef Marine Park Authority, universities and other government agencies (GBRMPA, 2018).

GBRMPA have provided an excel spreadsheet with the Reef Health data. Whilst a range of coral-related indicators were collected (e.g. recently dead coral), for consistency we used the same variable as in the Weekly Monitoring survey, namely whether coral was bleached (Yes = 1, as above) or not (a score of 5) (Table 2). The average score was calculated for each unique location in 2016 for which more than one observation was recorded, so that a scale response along the 1–5 axis was obtained.

2.7. Data limitations

There are several limitations to the data and their analysis, in addition to previously reported shortcomings of the quality of citizen-supplied data (e.g. ElQadi et al., 2017). One limitation relates closely to the very objective of this research of data integration. All data sources measure something slightly different, with the biggest assumption being made for Twitter data where sentiment is used to approximate coral quality. Even for the other more structured data sources, the original research purpose, variables and scales differ. An attempt has been made to normalize variables, but it is acknowledged that the scaled response might not be strictly linear. However, for the case of this study it was deemed acceptable as it allowed for a better comparison among data sets. In addition to issues around generating comparable scales, the data sources are characterised by different sample sizes and distribution across time and space, making direct comparisons more challenging – but potentially supporting the argument to using them in a complementary way.

In addition, all data sets display a high degree of geographic dispersal, and coordinates (longitude and latitude) had to be aggregated into grid cells of approximately 10 km across. We took this approach because our interest was in synthesis of data at the scale of reefs rather than in variability within individual reefs. Observations were also aggregated for the whole of 2016 (with the exception of Twitter for which only 9 months of data were available). This results in an over-simplification of change throughout the year and presents a major limitation to this analysis. However, attempts to compare data on finer time scales (e.g. monthly) failed because data volumes were insufficient.

In terms of limitations specific to the data source, the biggest challenge was encountered with Twitter. Using a predefined list of keywords to extract Reef-relevant tweets is pragmatic but likely to miss useful information. Relying on geo-coded tweets also limits the number of usable posts. Advanced recognition of content would enable automated recognition of relevant tweets and their location even for those tweets where location-enabled was switched off. In the meantime, using the coordinates of where the tweet was posted as an indication of what particular area of the Reef the information refers to is a first approximation, but somewhat simplistic. The uncertainty of location is aggravated by the fact that many locations visited by diver or snorkelers have no Internet reception and observations are likely to be shared once people return to the shore.

3. Results

In line with the research objectives, and building on the framework presented earlier, the result section will first present insights into the

Table 3
Comparison of data sources used in this research.

Data source	Volume	With geographic information	Unique locations covered	Data density (observations per location)	Number of days on which data were collected
Twitter	434	335	41	8	158
Sightings	259	259	50	5	93
CoralWatch	6093	6093	39	157	81
Tourism Weekly	665	665	19	35	296
Reef Health	1840	1840	85	22	121

nature of the data and key differences, followed by a visualisation of their geographic coverage and content.

3.1. Characteristics of the different data sets

The data sources differ substantially in their volume and coverage (Table 3). For Twitter, only 335 posts included geographic metadata, covering 41 unique locations. On average, the density of data was eight tweets per location, which outperforms the Sighting data. The quality of the collective sensing data, however, is inferior. A manual assessment shows that tweeters often discuss the state of the coral reef in general, rather than referring to a particular experience or observation in situ. Thus, reference to 'coral' in a tweet does not always lead to useful information about Reef health.

The size of the Sightings data set is relatively small ($N = 259$), but broadly comparable to the Twitter data in terms of locations covered and density (Table 3). The more sophisticated citizen science databases, Coral Watch and Tourism Weekly, comprise more observations than the collective sensing and human sensor data sets, but cover fewer locations. CoralWatch is by far the largest dataset of all and has the highest data density (157 data per location). The data density for Tourism Weekly observations is also relatively high, reflecting the fact that a selected number of operators regularly collect data in 19 specific locations. The frequency with which Tourism Weekly data are collected (i.e. on 296 days in 2016) is far greater than for any of the other sources.

The professional Reef Health surveys supplied 1840 data points for a total of 85 locations in 2016. The data density is still relatively high at 22 observations on average per location, but lower than that of the citizen science projects. Data were collected for about one third of the year.

3.2. Geography of observations and coral bleaching

A map-based visualisation of observations obtained from each data source provides further insight into similarities and differences between the data sets. For both unstructured sources, Twitter and Sightings, observations tended to cluster around population and tourist centres, including Cairns, Townsville and Airlie Beach, the gateway to the Whitsunday Islands (Fig. 1). The average sentiment of all tweets is positive, and this can be seen in the high average scores for the locations shown in Fig. 1. Examples of useful tweets that revealed insight into environmental conditions include:

- Posted from Fitzroy Island: *Hundreds of blue damsels tucked into a coral. I see you! @ GreatBarrier Reef* (original sentiment score: 0.66)
- Posted from Cairns: *Watching the stages of #coralbleaching go the wrong way. Stressed to fluorescent, bleached to algae*. #coralnotcoal (original sentiment score: −0.81)

The Sightings data indicate significant bleaching in the area between Port Douglas and Mackay (Fig. 1). Because of the mode of detection, which is serendipitous and depends on where visitors go and what they wish to report, the absence of bleaching incidents does not mean that bleaching did not occur.

CoralWatch broadly covers all parts of the GBR, with the largest number of observations found in the Southern Great Barrier Reef (Fig. 2). Furthermore, CoralWatch observations extend to the Far North to areas that are rarely visited by tourists or other members of the public, presumably being made by professional or semi-professional divers. The CoralWatch records reveal severe bleaching in the Northern GBR and no bleaching in the South. The Tourism Weekly data is naturally confined to key tourism dive sites, namely around Cairns and Port Douglas, and the Whitsunday Islands around Airlie Beach. Whilst there is a trend of relatively more bleaching in the Cairns and Whitsunday regions than in the south, the findings from this dataset are more mixed than those for CoralWatch.

The Reef Health survey has broad geographic coverage, a reflection of the aim of this survey to monitor coral conditions for the whole GBR (Fig. 3). For the Cairns and Whitsunday areas this means that the professional data overlap with the citizen science based datasets, but apart from these areas, the observations cover many areas that are not included in any of the other data sets.

3.3. Correlations

To explore the comparability of data sets, observations from the CoralWatch and Tourism Weekly sources are juxtaposed with those from the Reef Health survey, respectively (Fig. 4). We tested the relationships using linear regression for all locations for which values for each source existed were identified (18 locations for the CoralWatch-Reef Health comparison, and 16 for the Tourism Weekly-Reef Health correlation. Moran's I statistic was used to check for spatial autocorrelation (SAC)), using the R package *ape* (Paradis and Schliep, 2018). SAC was weak but significant, and to account for SAC error we ran the regressions using the lm_robust function from the *estimatr* package in R (Blair et al., 2019). Whilst neither correlations was significant ($p = .115$ and $p = .250$ for CoralWatch and Tourism Weekly, respectively), Fig. 4 shows a similar pattern of scores for the CoralWatch data compared with the Reef Health, and a much less clear relationship between Tourism Weekly and Reef Health data.

CoralWatch scores indicate less bleaching than Reef Health data for the same locations. Perhaps this can be explained by the different timing of data collection (Fig. 5). CoralWatch data are often collected by school groups that might select timing for a range of reason, and not in response to a specific incident. The Reef Health data collection is more targeted in response to events. Data collection was more frequent in early January, and intensified again during April and October 2016. The Tourism Weekly data show even coverage all year round, which probably explains the higher average score as periods of bleaching are covered equally as non-bleaching times.

4. Discussion

This research contributes to a rapidly growing body of literature on using citizen science for environmental monitoring. In particular, the aim was to understand the nature, geographic coverage and content of multiple data sources to assess opportunities of data integration into a hybrid model. A categorisation of different types of monitoring data sources presented in the first part of the paper informed the selection of datasets to ensure that each of the categories, ranging from collective sensing to human sensors and citizen science approaches, and scientific

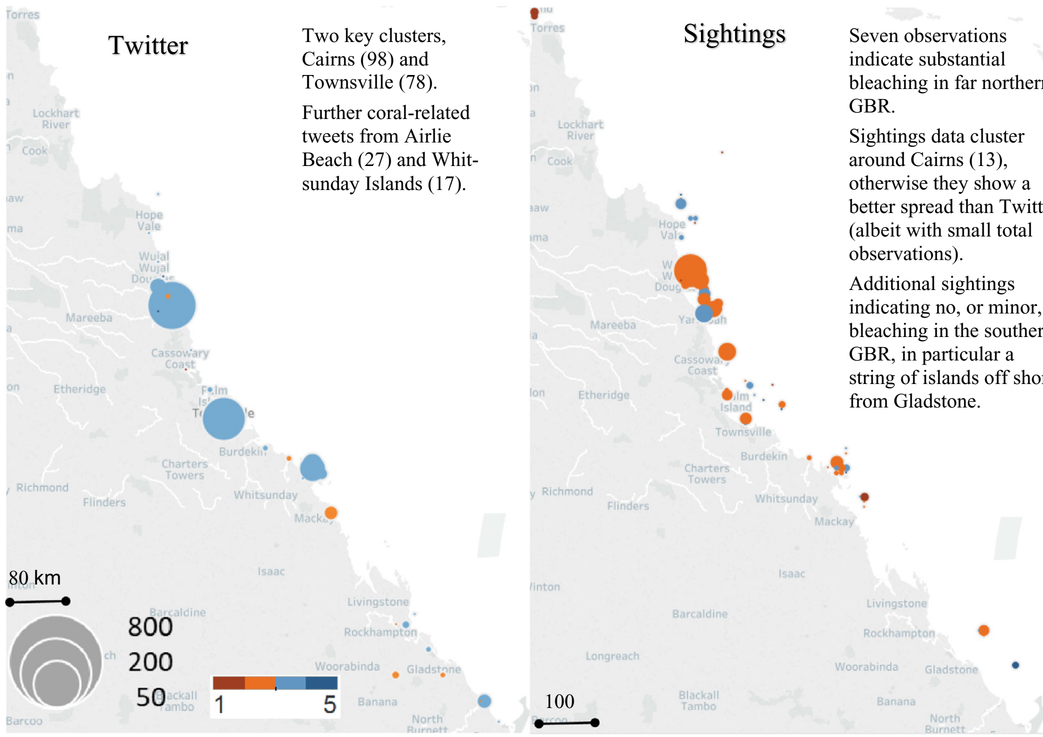

Fig. 1. Observation counts and bleaching values for unstructured sources, Twitter and Sightings. Circles size represents number of observations, with circles size scaled to be standardised with Figs. 2 and 3. Colour indicates intensity of “bleaching” with 1 being bleached and 5 representing unbleached.

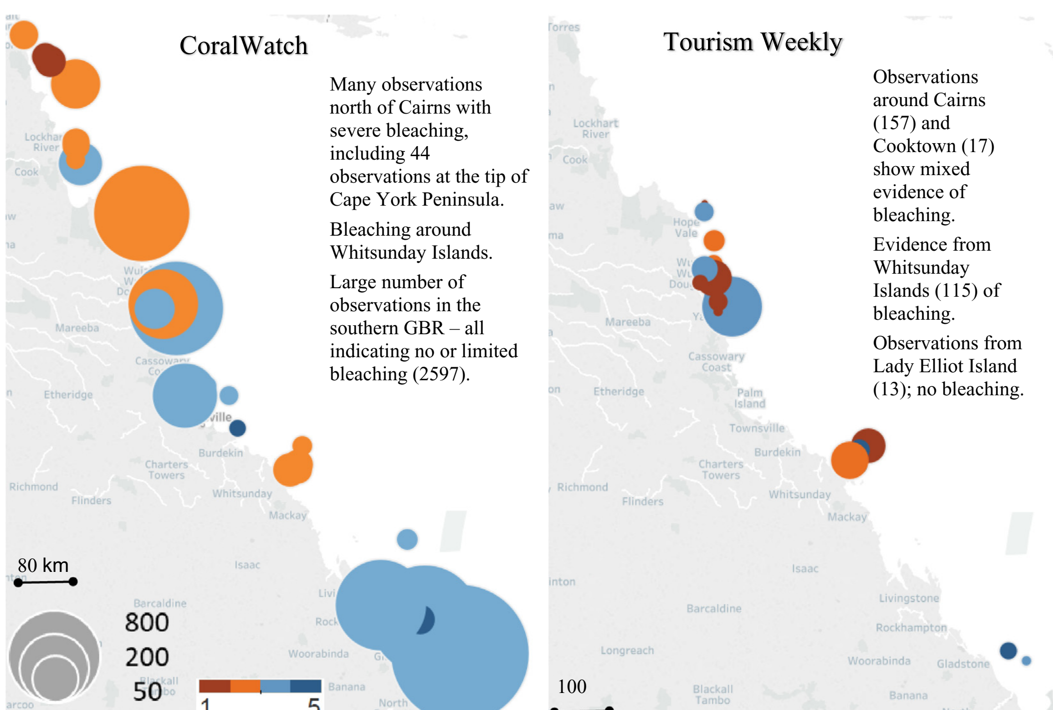

Fig. 2. Observation counts and bleaching values for structured citizen science sources CoralWatch and Tourism Weekly.

data were included. As noted by See et al. (2016) and Eitzel et al. (2017), the terminology is changing (and subject to fashion) and the classification presented in this paper is likely to evolve and perhaps contradict other schemes presented elsewhere. However, its basic dimensions are in line with the literature and provide a useful logic for the purpose of environmental monitoring.

The unique contribution of this research is the comparative assessment of these very different types of data, based on one common variable of interest, namely coral bleaching in the GBR. The 2016 coral bleaching event occurred on a North-South gradient, due to the water temperatures being warmer in the lower latitudes of the northern GBR (Hughes et al., 2017). The visualisation of data presented in Figs. 1 to 4

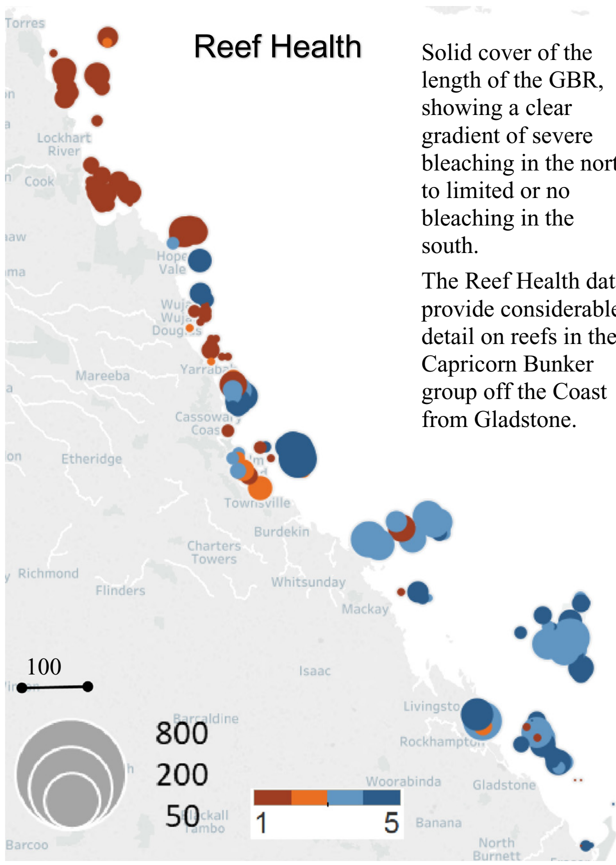

Fig. 3. Observation counts and bleaching values for Reef Health data set.

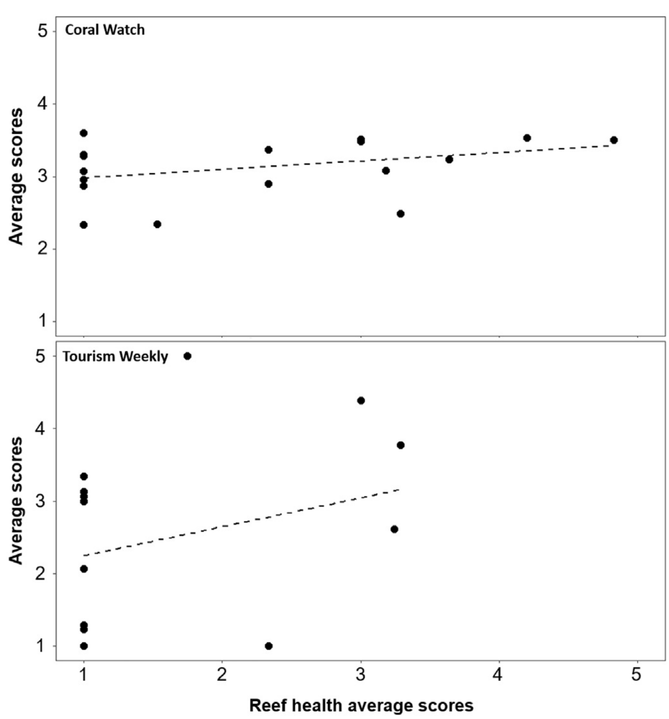

Fig. 4. Scatterplots for CoralWatch and Reef Health observations, and for Tourism Weekly and Reef Health observations at common locations within a 10 km grid.

indicated some broad alignment between data sources, with the exception of Twitter. Specific insights and implications from this research are discussed in the following.

4.1. Data quantity

The first part of the comparative assessment involved the quantity and density; that is the number of observations per unique location. An unexpected finding was the low volume for Twitter data, considering that one of the main perceived benefits of collective sensing data is the considerable amount of data. However, as this research showed, the number of relevant tweets reduces very quickly, once several steps of cleaning and filtering were implemented. It appears that extreme events (Steiger et al., 2015), attract larger volumes of tweets, but a topic as specific as 'coral' may be less suited for social media sharing. However, given that the open source tweets used in this research were only a small sample of all tweets from the region, it is possible to boost the dataset by purchasing a larger sample, or by considering multiple sources of social media (including open access Facebook pages, Sino Weibo and Flickr). Another way to increase social media activity is to encourage visitors to provide information. A combination with the Sightings platform is conceivable where social media promote the more structured human sensor program run by the GBRMPA. Such a combined approach would move this type of data source from being one of involuntary information provision to one of intentional submission, making it more akin to 'human sensing'.

An existing human sensing platform is the Eye on the Reef Sightings, which prompts users to provide observations following a pre-designed template. This mechanism ensures that only those visitors contribute to the program, who 'detected' something of interest. The number of GBR visitors to engage in this activity is small, but in comparison with Twitter the total number of 259 sightings is of higher value. Similarly, providers of observations to CoralWatch make a deliberate choice to engage in the collection of data. Indeed, the use of a designated colour-coded sheet requires a minimum level of preparation. CoralWatch has a high followership, ranging from occasional and amateur divers/snorkelers to frequent and professional 'experts'. Indeed, an earlier analysis of CoralWatch submissions revealed that the majority of users are school groups and dive centres, with only 13% of data provided by tourists (Marshall et al., 2012). This share could potentially be boosted with dedicated campaigns, for example through the 'Citizens of the Great Barrier Reef' (2018) movement. CoralWatch provides evidence that well-designed programs can attract significant support, and have the potential to add substantially to environmental monitoring (Lodia and Tardin, 2018; Tiagoa et al., 2017), including in the marine context (Hyder et al., 2015).

The main benefit of the Tourism Weekly program is the intensity with which each location is monitored. With an average of 35 observations per location in 2016, the tourism operators engaged in marine monitoring provide a greater density than GBRMPA's key tool of coral incident detection, the Reef Health survey. The regular collection of data in the same collection makes the Tourism Weekly a valuable data source.

4.2. Geography

The analysis of spatial coverage for the five different data sets revealed clear overlaps, as well as differences. The collective sensing and human sensor type datasets generate most information from frequently visited areas (ElQadi et al., 2017). Considering that the provision of information of coral is not the main purpose of visiting the Reef, but a coincidental by-product of other activities, this is not surprising. The clustering of observations around tourist spots is further reinforced by the Tourism Weekly program, which by nature focuses on those parts of the Reef that are used for diving and snorkelling.

At the same time, those areas that are not frequented by tourists are

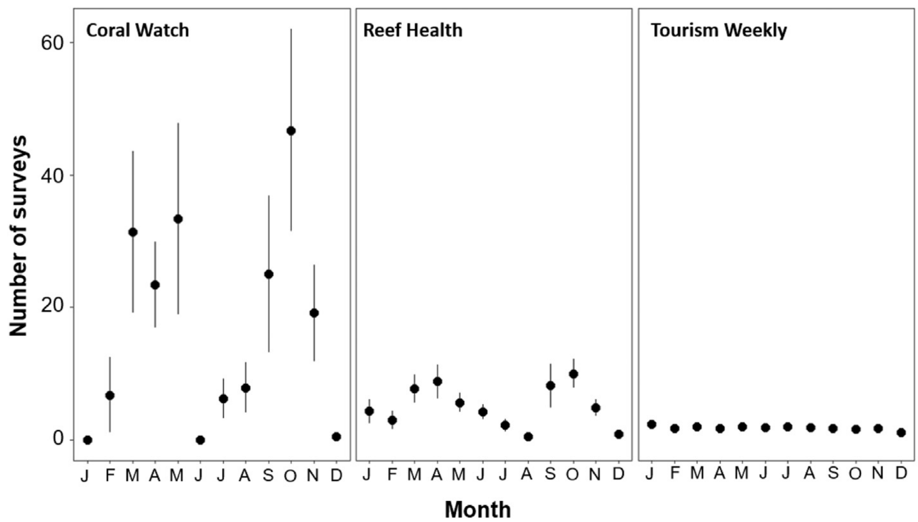

Fig. 5. Volume and temporal dispersion of data collection for the Coral Watch, Tourism Weekly and Reef Health data.

likely only monitored as a result of targeted programs or due to the activities of highly dedicated individuals (e.g. scientists, professional divers, rangers). The extent of the CoralWatch dataset is noteworthy in this regards. Whilst a citizen science program, the database includes a considerable amount of observations from remote locations (including by the 16% of scientists or 28% of dive centre staff, Marshall et al., 2012), although clustering around tourist sites is still evident. The geographic extent of data makes this dataset particularly valuable, although the dispersion of data is not quite the same level of the Reef Health survey, which is the most purposeful and centrally managed of all datasets.

Understanding the spatial differences is critical for exploring redundancies and complementarities, and also the limitations of each individual data set. The Tourism Weekly, for example, could provide an opportunity to reduce the number of Reef Health surveys in the same area, considering that both data collections follow similar standards and levels of training. The CoralWatch could be used to complement the Reef Health data in areas where there are very limited other observations, for example in the Southern Great Barrier Reef.

4.3. Data quality

Whilst geographic coverage is an important attribute, ultimately, the quality of data is critical. Perhaps not surprisingly, the Twitter posts are subject to a positivity bias (Mitchell et al., 2013), lacked specificity and did not emerge as a robust data source for coral-related incidents (confirming the broader trend that most citizen science approaches involve active participation, See et al., 2016). However, the tweets still present some useful insights into how people communicate about coral, where and when they feature in social media posts and what the key topics are. As such, they are more likely of value to socio-economic monitoring programs (e.g. visitor experience), with the additional possibility of identifying 'red flags' of particular events (Becken et al., 2017). Future use of machine learning and topic detection will help detect those tweets that discuss the state of the Reef but do not explicitly mention any of the keywords, for example 'coral' used in this present study. Other improvements to the filtering mechanism could include identifying synonyms and misspelt keywords to ensure no valuable data are lost and increase the sample size (Barve, 2014). Further improvements to sentiment detection are necessary to increase the usefulness of this proxy and to derive target-specific sentiment (Alaei et al., 2017).

One challenge with both the social media and the Sightings sources is that these only provide data on 'presence', and not 'absence'. In other words, the fact that a particular location is not mentioned or does not have a sighting does not mean that there was no incident. The site may simply not have been visited, or those having been there did not engage in providing information, for example because the experience did have limited emotional impact (Mitchell et al., 2013). There are some inherent biases associated with this issue and correcting from them is a major research challenge (Welvaert and Caley, 2016). Presence bias represents a substantial limitation of using social media and citizen supplied data sets for ecological monitoring. Some work has been done on more structured citizen science projects (e.g. koala detection along transport corridors, Paul et al., 2014), but very little research has assessed bias control for collective sensing or crowdsourced data. If the social media and Sightings data are merely complementing other data sets, this may not be a pressing issue at this point.

4.4. Hybrid approach

More recently, organisations have begun to explore hybrid approaches whereby citizen science data are integrated with traditional measurements. Researchers have demonstrated the benefits of such dual approaches, for example in the context of earthquakes (Crooks et al., 2013), air pollution (Riga & Karatzas, 2014), and diseases of popular terrestrial species (Connors et al., 2012). The parameters of data density, quality and temporal distribution, as proposed in this research, can guide this process, and a structured benchmarking of the features of each data source, as proposed and operationalised by Gharesifard et al. (2017) is recommended should an organisation such as GBRMPA seek to invest in a hybrid system. Future research could develop a tool or guide for organisations to assess and compare the suitability of different data sources.

Data integration in the marine context can be challenging, because the logistics of involving the public in data collection on or under the water is more complex than that compared with terrestrial projects (Stuart-Smith et al., 2017). Visitors to the GBR only spend little time at the Reef, and collection of data (e.g. through photographs, user generated online content, or specific apps) may require specific equipment or Internet availability (Hyder et al., 2015). Possibly for this reason, existing attempts to combine citizen science with professional monitoring are either terrestrial or rely on those programs that involve considerable training of volunteers. Two examples of marine programs are the Reef Life Survey (Stuart-Smith et al., 2017) and Reef Check Australia (Done et al., 2017). Both delivered valuable results on specific aspects of environmental monitoring. It may now be time to extend these approaches and integrate a broader array of citizen based data for a more "effective long-term management of the Marine Protected Areas" (Lodia and Tardin, 2018, p. 52).

The results presented here show considerable potential to develop a hybrid system that either displays multi-source information simultaneously, or pools data with the intention to exploit redundancies. Combining data records in areas of high data density could result in cost savings, or allow the GBRMPA to reallocate resources to those areas that are not visited by citizen scientists. The next step will be to use multiple years of data and develop a full spatial model to help optimise integration of different data sources by explicitly drawing on the varying spatial and temporal data collection patterns.

5. Conclusion

This research addresses the increasing need to develop cost-effective and comprehensive systems to monitor environmental change in natural environments. It is proposed that those areas that benefit from frequent visitation by members of the public or tourists could invest into a broad range of citizen based monitoring schemes, ranging from collective sensing, to more structured human sensor approaches or fully trained citizen science projects. The comparison of five data sets containing information on the state of coral at the Great Barrier Reef reveals useful insights into compatibility and complementarity of data. Whilst Twitter data, at this point, appears to mainly provide information on how people 'experience' the Reef, rather than its condition, the Eye on the Reef Sightings platform goes some way in delivering useful additional data on coral bleaching incidents. Similar to the Tourism Weekly survey, data are concentrated in those areas that are popular with tourists. The CoralWatch citizen science program was found to deliver higher-quality data with a wide geographic spread, thanks to its diverse base of users that include tourists, school groups, scientists and other professionals. The paper concludes, however, that there is considerable potential for existing or newly developed citizen based programs to support, or in some locations replace, the professional and costly Reef Health survey, which in turn could then focus on covering those locations for which no data are provided by the public.

Acknowledgement

This project was funded through a National Environment Science Program (NESP) fund, within the Tropical Water Quality Hub (Project No: 2.3.2). We would also like to thank GBRMPA for providing us with their data, and we also greatly appreciate CoralWatch's collaboration in sharing their data.

References

Addison, P., Walshe, T., Sweatman, H., Jonker, M., Anthony, K., MacNeil, A., Thompson, A., Logan, M., 2015. Towards an Integrated Monitoring Program: Identifying Indicators and Existing Monitoring Programs to Effectively Evaluate the Long Term Sustainability Plan. Australian Institute of Marine Science, Townsville, Queensland.
Alaei, A.R., Becken, S., Stantic, B., 2017. Sentiment analysis in tourism: beginning of a new research paradigm. J. Travel Res. 1–17. https://doi.org/10.1177/0047287517747753.
Barve, V., 2014. Discovering and developing primary biodiversity data from social networking sites: a novel approach. Ecol. Inform. 24, 194–199.
Becken, S., Stantic, B., Chen, J., Alaei, A.R., Connolly, R., 2017. Monitoring the environment and human sentiment on the Great Barrier Reef: assessing the potential of collective sensing. J. Environ. Manag. 203, 87–97.
Blair, G., Cooper, J., Coppock, A., Humphreys, M., Sonnet, L., 2019. Estimatr: Fast Estimators for Design-Based Inference. R Package Version 0.16. https://CRAN.R-project.org/package=estimatr.
Boulos, M.N., Resch, B., Crowley, D.N., Breslin, J.G., Sohn, G., Burtner, R., Pike, W.A., Jezierski, E., Chuang, Y., 2011. Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. Int. J. Health Geogr. 10 (67ff).
Citizens of the Great Barrier Reef, 2018. Homepage. Available (05/06/18). https://citizensgbr.org/.
Connors, J.P., Lei, S., Kelly, M., 2012. Citizen science in the age of neogeography: utilizing volunteered geographic information for environmental monitoring. Ann. Assoc. Am. Geogr. 102 (6), 1267–1289.
CoralWatch, 2017. Homepage. Available (20/02/18). www.coralwatch.org.
Cornell Laboratory of Ornithology, 2018. EBird. Homepage. Available (20/02/18). http://www.birds.cornell.edu/citscitoolkit/projects/clo/eBird/.
Crooks, A., Croitoru, A., Stefanidis, A., Radzikowski, J., 2013. #earthquake: twitter as a distributed sensor system. Trans. GIS 17 (1), 124–147.
Daume, S., 2016. Mining twitter to monitor invasive alien species — an analytical framework and sample information topologies. Ecol. Inform. 31, 70–82.
Daume, S., Galaz, V., 2016. "Anyone know what species this is?" – twitter conversations as embryonic citizen science communities. PLoS One 11 (3), e0151387. https://doi.org/10.1371/journal.pone.0151387.
Department of Environment, 2015. Reef 2050 Long-Term Sustainability Plan. Commonwealth of Australia, Canberra.
Done, T., Roelfsema, C., Harvey, A., Schuller, L., Hill, J., Schläppy, M., Lea, A., Bauer-Civiello, A., Loder, J., 2017. Reliability and utility of citizen science reef monitoring data collected by Reef Check Australia, 2002–2015. Mar. Pollut. Bull. 117, 148–155.
Eitzel, M.V., Cappadonna, J.L., Santos-Lang, C., Duerr, R.E., Virapongse, A., West, S.E., Kyba, C.C.M., Bowser, A., Cooper, C.B., Sforzi, A., Metcalfe, A.N., Harris, E.S., Thiel, M., Haklay, M., Ponciano, L., Roche, J., Ceccaroni, L., Shilling, F.M., Dörler, D., Heigl, F., Kiessling, T., Davis, B.Y., Jiang, Q., 2017. Citizen science terminology matters: exploring key terms. Citizen Sci. 2 (1), 1. https://doi.org/10.5334/cstp.96.
ElQadi, M.M., Dorina, A., Dyer, A., Burda, M., Bukovac, Z., Shrestha, M., 2017. Mapping species distributions with social media geo-tagged images: case studies of bees and flowering plants in Australia. Ecol. Inform. 39, 23–31.
GBRMPA, 2017a. Significant Coral Decline and Habitat Loss on the Great Barrier Reef (29/05/2017). Available (09/07/17). http://www.gbrmpa.gov.au/media-room/latest-news/coral-bleaching/2017/significant-coral-decline-and-habitat-loss-on-the-great-barrier-reef.
GBRMPA, 2017b. Eye on the Reef. Available (08/09/17). http://www.gbrmpa.gov.au/managing-the-reef/how-the-reefs-managed/eye-on-the-reef.
GBRMPA, 2018. Reef Health and Impact Survey. Available (08/02/18). http://www.gbrmpa.gov.au/managing-the-reef/how-the-reefs-managed/eye-on-the-reef/reef-health-and-impact-survey.
Gharesifard, M., When, U., van der Zaag, P., 2017. Towards benchmarking citizen observatories: features and functioning of online amateur weather networks. J. Environ. Manag. 193, 381–393.
Hughes, T.P., Kerry, J.T., Alvarez-Noriega, M., Alvarez-Romero, J.G., Anderson, K.D., Baird, A.H., Babcock, R.C., Beger, M., Bellwood, D.R., Berkelmans, R., Bridge, T.C., Butler, I.R., Byrne, M., Cantin, N.E., Comeau, S., Connolly, S.R., Cumming, G.S., Dalton, S.J., Diaz-Pulido, G., Eakin, C.M., Figueira, W.F., Gilmour, J.P., Harrison, H.B., Heron, S.F., Hoey, A.S., Hobbs, J.A., Hoogenboom, M.O., Kennedy, E.V., Kuo, C.Y., Lough, J.M., Lowe, R.J., Liu, G., McCulloch, M.T., Malcolm, H.A., McWilliam, M.J., Pandolfi, J.M., Pears, R.J., Pratchett, M.S., Schoepf, V., Simpson, T., Skirving, W.J., Sommer, B., Torda, G., Wachenfeld, D.R., Willis, B.L., Wilson, S.K., 2017. Global warming and recurrent mass bleaching of corals. Nature 543, 373–377.
Hutto, C., Gilbert, E., 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the 8th International AAAI Conference on Weblogs and Social Media. AAAI, Palo Alto, CA, pp. 216–225.
Hyder, K., Townhill, B., Anderson, L.G., Delany, J., Pinnegar, J.K., 2015. Can citizen science contribute to the evidence-base that underpins marine policy? Mar. Policy 59, 112–120.
Hyvarinen, O., Saltikoff, E., 2010. Social media as a source of meteorological observations. Mon. Weather Rev. 138, 3175–3184.
Kearns, F.R., Kelly, M., Tuxen, K.A., 2003. Everything happens somewhere: UsingWebGIS as a tool for sustainable natural resource management. Front. Ecol. Environ. 1, 541–548.
Liu, Y., Liu, X., Gao, S., Gong, L., Kang, C., Zhi, Y., Chi, G., Shi, L., 2015. Social sensing: a new approach to understanding our socioeconomic environments. Ann. Assoc. Am. Geogr. 105 (3), 512–530. https://doi.org/10.1080/00045608.2015.1018773.
Lodia, L., Tardin, R., 2018. Citizen science contributes to the understanding of the occurrence and distribution of cetaceans in South-Eastern Brazil – a case study. Ocean Coast. Manag. 158, 45–55.
Marshall, N., Kleine, D.A., Dean, A.J., 2012. CoralWatch: education, monitoring, and sustainability through citizen science. Front. Ecol. Environ. 10 (6), 332–334.
McKinley, D., Miller-Rushing, A.J., Ballard, H.L., Bonney, R., Brown, H., Cook-Patton, S., Evans, D., French, R.A., Parrish, J.K., Phillips, T.B., Ryan, S.F., Shanley, L.A., Shirk, J., Stepenuck, K., Weltzin, J.F., Wiggins, A., Boyle, O., Briggs, R.D., Chapin, S., Hewitt, D.A., Preuss, P.W., Soukup, M., 2017. Citizen science can improve conservation science, natural resource management, and environmental protection. Biol. Conserv. 208, 15–28.
Mitchell, L., Frank, M.R., Harris, K.D., Dodds, P.S., Danforth, C.M., 2013. The geography of happiness: connecting twitter sentiment and expression, demographics, and objective characteristics of place. PLoS One 8 (5), e64417.
Paradis, E., Schliep, K., 2018. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35 (3), 526–528.
Paul, K., Quinn, M.S., Huijser, M.P., Graham, J., Broberg, L., 2014. An evaluation of a citizen science data collection program for recording wildlife observations along a highway. J. Environ. Manag. 139 (15), 180–187.
Resch, B., 2013. People as sensors and collective sensing-contextual observations complementing geo-sensor network measurements. In: Krisp, J.M. (Ed.), Progress in Location-Based Services, Lecture Notes in Geoinformation and Cartography. Springer-Verlag, Berlin Heidelberg, pp. 391–406.
Riga, M., Karatzas, K., 2014. Investigating the relationship between social media content and real-time observations for urban air quality and public health. In: Proc 4th Int Conf Web Intell Min Semant, Thessaloniki, Greece.
Romana, L.A., Scharenbroch, B.C., Östberg, J.P.A., Mueller, L.S., Henning, J.G., Koeser,

A.K., Sanders, J.R., Betz, D.R., Jordan, R.C., 2017. Data quality in citizen science urban tree inventories. Urban For. Urban Green. 22, 124–135.
See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, J., Fritz, S., Kerle, N., Jiang, B., Laakso, M., Liu, H., Milcinski, G., Niksic, M., Painho, M., Pődör, A., Olteanu Raimond, A., Rutzinger, M., 2016. Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. ISPRS Int. J. Geo Inf. 5. https://doi.org/10.3390/ijgi5050055.
Statistica, 2018. Twitter Users. Available (26/02/18). https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/.
Steiger, E., de Albuquerque, J.P., Zipf, A., 2015. An advanced systematic literature review on spatiotemporal analyses of twitter data. Trans. GIS 19 (6), 809–834.
Stuart-Smith, R.D., Edgar, G.J., Barrett, N.S., Bates, A.E., Baker, S., Bax, N.J., Becerro, M.A., Berkhout, J., Blanchard, J.L., Brock, D.J., Clark, G.F., Cooper, A.T., Davis, T.R., Day, P.B., Duffy, E., Holmes, T.H., Howe, S.A., Jordan, A., Kininmonth, S., Knott, N.A., Lefcheck, J.S., Ling, S.D., Parr, A., Strain, E., Sweatman, H., Thomson, R., 2017. Assessing national biodiversity trends for rocky and coral reefs through the integration of citizen science and scientific monitoring programs. Bioscience 67 (2), 134–146.
Swanson, A., 2015. Crowdsourcing the Serengeti: How citizen scientists classified millions of photos from home. In: The Conversation, June 9. Available (26/02/18). https://theconversation.com/crowdsourcing-the-serengeti-how-citizen-scientists-classified-millions-of-photos-from-home-42930.
The Nature Conservancy, 2017. Mapping Ocean Wealth. Available (12/05/17). http://oceanwealth.org/.
Theobald, E.J., Ettinger, A.K., Burgess, H.K., DeBey, L.B., Schmidt, N.R., Froehlich, H.E., Wagner, C., HilleRisLambers, J., Tewksbury, J., Harsch, M.A., Parrish, J.K., 2015. Global change and local solutions: tapping the unrealized potential of citizen science for biodiversity research. Biol. Conserv. 181, 236–244.
Tiagoa, P., Pereirab, H.M., Capinha, C., 2017. Using citizen science data to estimate climatic niches and species distributions. Basic Appl. Ecol. 20, 75–85.
Vivacqua, A.S., Borges, M.R.S., 2012. Taking advantage of collective knowledge in emergency response systems. J. Netw. Comput. Appl. 35 (1), 189–198.
Welvaert, M., Caley, P., 2016. Citizen Surveillance for Environmental Monitoring: Combining the Efforts of Citizen Science and Crowdsourcing in a Quantitative Data Framework. vol. 5 SpringerPlus (1890-ff).