



South Atlantic Natural Capital Project: Mapping natural, historical and built heritage values in St Helena Island using geolocated social media





Dimitrios Bormpoudakis Robert Fish Ness Smith

March 2019

Review table

Name	Reviewed by	Date
Version 1	Ness Smith	14/03/19
Version 2	Tara Pelembe and Paul Brickle	17/03/19
Version 3	SHG	16/05/19
Version 4		

Acknowledgements:

We would like to thank the NCA project St Helena Advisory Group for their input and help with this report.

This research was funded by The UK Government via the Conflict, Security and Stability Fund.

Suggested citation:

Bormpoudakis, D., Fish, R., Smith, N. (2019). Mapping natural, historical and built heritage values in St Helena Island using geolocated social media. Final Report for the South Atlantic Overseas Territories Natural Capital Assessment

For more information, please contact the South Atlantic Environmental Research Institute (SAERI) at info@saeri.ac.fk or visit http://south-atlantic-research.org PO Box 609, Stanley Cottage Stanley FIQQ 1ZZ Falkland Islands Tel: +500 27374 www.south-atlantic-research.org

<u>SAERI is a registered</u> Charity in England and Wales (#1173105) and is also on the register of approved Charities in the Falkland Islands (C47). SAERI also has a wholly-owned trading subsidiary – SAERI (Falklands) Ltd – registered in the Falkland Islands.



MAPPING NATURAL, HISTORICAL AND BUILT HERITAGE VALUES IN ST HELENA ISLAND USING GEOLOCATED SOCIAL MEDIA



FINAL REPORT

NATURAL CAPITAL ASSESSMENT

SOUTH ATLANTIC OVERSEAS TERRITORIES

Dimitrios Bormpoudakis & Robert Fish, School of Anthropology and Conservation, University of Kent, Canterbury

Ness Smith, South Atlantic Environmental Research Institute, Stanley

Suggested Citation

Bormpoudakis, D., Fish, R., Smith, N. (2019) <u>Mapping Natural, Historical And Built</u> <u>Heritage Values In St Helena Island Using Geolocated Social Media</u> Final Report. Natural Capital Assessment. South Atlantic Overseas Territories (SAERI/University of Kent)

1. Mapping natural, historical and built heritage values using Social Media content (MSM)

Background

This report was commissioned by the South Atlantic Environmental Research Institute (SAERI). Its findings contribute evidence to a programme of natural capital assessment (NCA) being implemented by the UK Joint Nature Conservation Committee and conducted by SAERI in the UK South Atlantic Overseas Territories. Funded by the Foreign and Commonwealth Office (FCO) managed Conflict, Stability and Security Fund (CSSF), the work sits under its Environmental Resilience programme which includes objectives to integrate natural capital considerations into economic and social development planning.

Mapping natural, historical and built heritage values using social media content is a relatively recent methodology that capitalises upon the widespread use of web-based social media platforms. It typically involves the connection to the Application Programming Interface (API) of a social media provider (Facebook, Instagram, Twitter, Flickr, etc.) and downloading 'geolocated' public¹ posts, user data and metadata that are of interest for a specific application. For example, using the public Twitter API one can have access to recent posts ('tweets') by all public users from a specific location and to the tweets' and users' data and metadata such as: location the tweet was sent from (if the user has enabled location); users' stated home place and; time the tweet was sent. Thus, one can use this information to map Twitter activity. Using photograph-based social media such as Instagram and Flickr is done in a similar way (see below for a more detailed explanation).

There remains a question regarding the suitability of social media-based methodologies in places where internet connections (both home and mobile) are poor and the resident and tourist populations low, as is currently the case of St Helena. Social-media mapping has mostly found application in countries of the global north where mobile internet access is relatively good and uniform or in areas of the global south with significant amounts of tourists, e.g. Serengeti National Park in Tanzania. The practicalities and usefulness of social media mapping in places like St Helena, which are so isolated as to have limited amounts of tourists and poor mobile and home internet connections, have not been evaluated to date. Furthermore, considering that each social-media platform has its own demographic profile of users as well as its own way of handling geographic information, it is worth considering the advantages and disadvantages of using particular platforms for mapping natural, historical and built heritage values.

2. Methodology

Flickr allows access to public photographs and photograph metadata through its API. The API enables searching the Flickr archive by many means (title, user, tags, etc.). Crucially for our purposes, it is possible to search through photograph location metadata. We downloaded all photographs that were geolocated (that means we only searched for photographs that have geographical coordinates in their metadata; photographs that do not have this kind of information were excluded from the search) in St Helena between, taken between1st of December 2005 and 31st of July 2018.

Due to recent changes in the way one can access the Instagram archive, Instagram only allows access to the most recent 100 posts for geographical searches. To overcome this limitation, we used a web-

¹ In the vast majority of cases, public APIs do not allow access to private posts or access to accounts and posts that are visible only after invite by/ application to their respective owners.

service provided by Netlytic (<u>https://netlytic.org/</u>) to schedule an hourly download of Instagram photographs and metadata that allowed us to "scrape" all Instagram posts from St Helena. We scheduled an hourly data collection from 25 of August to 4th of December 2018.

We downloaded all geolocated photographs from Instagram and Flickr and retained:

- The image itself;
- A series of metadata associated with the photographs: title, tags, latitude, longitude, user, and date the photograph was taken.

Using this information, we were able to provide summary statistics and map all photographs to precise locations on St Helena. Furthermore, because we downloaded the photograph, we could select photographs that depict the natural environment, the built environment or historical heritage. Namely, from the total set of photographs downloaded from Flickr and Instagram, we viewed every single photograph and excluded photographs that did not depict elements of natural, historical or built heritage, e.g. face close ups, advertisements, 'selfies', house interiors etc. Furthermore, to assess the mapping accuracy of the two datasets, we focused on photograph hot spots and using a satellite image of the area we calculated the percentage of photographs that depict the area they are mapped on. For example, for both datasets the harbour area is a photograph hotspot; we viewed all photographs from that area to assess whether they were indeed taken there or if that is where the poster obtained internet connection and uploaded the photograph/post. Finally, we deleted duplicate photographs.

We chose three ways to display the photographs on a map of St Helena. First, we present the photographs as raw location data, i.e. as points on the map based on where the photograph was taken. Second, to provide a sense of the magnitude of the difference between parts of the Island, and to provide a continuous layer for further modelling, we employed a commonly used technique, kernel density estimation. However, all these visualisation techniques suffer from user bias, i.e. few active users taking photographs of the same areas in the same day. To overcome this limitation, we adapted a metric from Wood et al. (2013), total visitation user days (henceforth UD). UD, as defined by Wood et al. (2013), is *"the total number of days, across all users, that each person took at least one photograph within each site"*. This approach allows for standardising mapped social media output, particularly for cases when there are a few active users taking photographs in the same area. We followed the approach of Wood et al. (2013) because initial analysis of the photograph and photograph metadata revealed that only a few users took the majority of the photographs. Instead of focusing on particular sites as Wood et al. (2013) did, we adapted their approach and applied it to a grid superimposed on the St Helena island map. Arguably, the latter approach for visualising the photographs is the most unbiased.

For all visualisation methods, we distinguish between photographic datasets that include natural, historical and built heritage, and datasets that include only photographs of the natural environment. Note that sometimes a strict distinction between these two categories is not feasible. For example, photographs that include landscape elements, built heritage and the natural environment would be classified in both categories (see e.g. following links to photographs on Instagram: https://goo.gl/x4qkFu; https://goo.gl/xmyka; https://goo.gl/pXmGhH).

3. Results

General characteristics of the downloaded datasets

We downloaded a total of 1020 photographs from Flickr, taken between 1st of December 2005 to 31st of July 2018, and 964 photographs from Instagram, taken between 25th of August 2018 to 4th of December 2018². After filtering the data for photographs that depict natural, historical or built heritage, the Flickr dataset was reduced to 902 photographs (with their associated metadata as described above). Applying the same filtering procedure, the Instagram dataset was reduced to only 272 photographs, less than a third compared to the Flickr dataset, reflecting the different uses and audiences of the two social media platforms.

The Instagram dataset provided further complications for analysis. Unlike Flickr, for which the only geolocation for each photograph is the one contained in the data uploaded to the website as photograph metadata, in Instagram there are a variety of ways of 'geo-tagging' a post (notice the difference between post and photograph for Instagram and Flickr respectively). The user has the option to use a location based on his/her IP, or tag one of the already available 'places' (e.g. '#StHelena') or use the photograph metadata. As a result, in the dataset there are cases where many photographs are seemingly taken from exactly the same location, but are depicting completely different locations. In just one case, there are 129 photographs (almost half of the 272 photographs of the Instagram dataset) taken from the same spot, depicting wildly different locations (from Jacob's Ladder to Lot's Wife Ponds) (Fig. 1a). This is probably because this location is the default St Helena location according to Instagram, hence all posts geolocated in '#St Helena' were allocated to that particular spot. Also, in many cases, photograph mapping did not correspond to where the photograph was taken. In the analysis of photograph hot-spots we conducted, only seven out of the 272 locations where the photographs were mapped actually correspond to where they were taken (e.g. Fig. 1b). After filtering the Instagram dataset, only these seven locations remained. As a result, further analysis of the Instagram dataset was not undertaken, and all analyses below refer to the Flickr-derived dataset only.

The Flickr dataset, following an identical filtering procedure was reduced to 685 records (and 459 for



Figure . a: Location where 129 Instagram photographs from St Helena were mapped. b: Example of a Instagram-derived photograph location for which its spatial location does not match the content of the photographs. In the insets we can see the photographs that were mapped to that location. Circled point refers to the location where the photographs were mapped.

the dataset that excludes purely historical or built heritage photographs). Due to differences in how the two social media websites use geolocation (see above), and the years of available photographs the

² Flickr allows access to its archive. On the contrary, Instagram only allows for 'live' data download.

Flickr dataset was much less reduced than the Instagram one, and didn't suffer from the same amount of problems. General characteristics of the Flickr dataset are (see Fig. 2 below):

- 31 users in total have uploaded photographs on Flickr; circa 1/3 of the photographs uploaded by two users from the UK
- Similarly, some particular years dominate the dataset, with 2011, 2015, and 2018 sharing between them more than half of the photographs uploaded on Flickr.
- The number of Flickr users uploading geo-tagged photos on Flickr per year is not very big. Some years (2005, 2006, 2007, 2008, 2009, 2010, 2012 and 2014) have 0 to 2 users uploading photographs on Flickr.

Thus, drawing robust inferences from this dataset could be problematic, as it is biased towards certain users and years. However, it can, and as we will see below does, provide a fairly good outline of the areas that are of particular interest in St. Helena.



Figure 2. Top: Number of photographs taken in SH and uploaded to Flickr by user (2005 – 2018); Middle: Number of photographs uploaded to Flickr from SH per year; Bottom: Number of different users posting photographs from St Helena on Flickr per year. All tabulations refers to the final edited dataset of 685 records.

Mapped Flickr output

a. Mapped Flickr photographs (1/2005 - 9/2018)

The Flickr dataset of mapped photographs (Figure 3) revealed that the majority of photographs are taken from the area around Jamestown, including the nearby sea and coast. Well known sites of historical and built heritage are also prominent, such as Napoleon's Tomb, Longwood House, or Plantation House. Interestingly, we can see that not all photographs from these predominantly historical heritage sites depict elements of the historical or built environment. Photographs that depict the natural environment are also present, particularly around Napoleon's Tomb, or in the grounds of Longwood House. As expected, for almost all cases photographs far from Jamestown, other settlements and sites of historical heritage depict the natural environment. These differences will be more prominent when we take a look at the metric of photographic user days below.



Figure 3. Mapped photographs downloaded from Flickr. Boundaries of Peaks National Park are in green.

b. Kernel density estimation

To generalise the distribution and better visualise the differences in photographers' interest in specific locations, we used kernel density estimation (Fig. 4). The continuous surface we produced indicates how many photographs per square kilometre we would expect to find in a particular location. This approach highlights similar areas to simply mapping the photographs (Figure 3), namely the area around Jamestown and particular sites of built or historical heritage are the most prominent. Comparing the two images we can see that they both show the same pattern, although for the map that excludes photographs of the built or historical environment, natural areas appear more prominent (e.g. Peaks National Park, Lot's Wife Ponds or South West Point).



Figure 4. Flickr photographs with a kernel density layer as background. Top: calculations include all Flickr photographs. Bottom: calculations only include photographs that depict the natural environment. Cell resolution = 50 m; search distance: 1000 m.

Total photographic user days

However, both ways of visualising and mapping photographs from Flickr are biased towards photographers who take many photos of the same location, or by photographers who take many photographs in general (see general description of the Flickr dataset above). To overcome this limitation we present the metric of photographic user days, which for every grid cell in the maps below, counts the number of photographers taking at least a picture in a single day (Figures 5 and 6). Thus, for example, if a photographer takes several pictures of a particular tree in the same day, it will count as one.

In the maps below (Figures 5 and 6) we see the same familiar pattern although the maps, since photographic user days (PUD) rather than individual photographs are being counted, are more homogeneous. In contrast to kernel density maps, areas of natural beauty like the Peaks National Park appear to have roughly the same magnitude of PUD as Plantation House for example, and other historical landmarks. Jamestown still dominates the magnitude of PUD, however natural areas are equally represented. The representation we have opted for here also makes it clear that while some areas may be well known for their built or historical environment, they do not lack interesting elements of the natural environment. For example Longwood House and the area around High Knoll Fort remain the most photographed even when excluding photographs that depict only built or historical heritage.



Figure 5. Total user days calculated for St Helena using Flickr photographs that depict the natural, historical or built heritage. Boundaries of Peaks National Park are in green. From Wood et al. (2013): *"for photographs, user-days are defined as the total number of days, across all users, that each person took at least one photograph within each site (PUD)."* Wood et al. (2013) use this Flickr-derived statistic for whole National Parks (which could be the size of St Helena or more), but we used it here for 500 m by 500 m grid cells.



Figure 6. Total user days calculated for St Helena using Flickr photographs that depict the natural environment, excluding photographs that include only elements of built or historical heritage.

4. Conclusion

We used social media-derived geolocated data to map natural environment, historical and built heritage values in the island of St Helena. St Helena being a remote oceanic island with very limited home or mobile internet connection, part of the task was to gauge whether this type of internet-based methodology can be applied to similar situations. Flickr and Instagram were used as paradigmatic social media which are built around sharing photographs of specific locations.

We found that photographic interest in St Helena is gathered mainly around the capital, Jamestown, and its landmarks (e.g. Jacob's Ladder; various traditional houses; the port). Secondarily, most photographers are attracted by historical and built heritage areas. Notably though, the natural environment appears to play a role in the appreciation of the built and historical environment, as many photographs taken in places of historical interest (High Knoll Fort, Longwood House, Napoleon's Tomb or even in Jamestown) are actually photographs of nature. The most telling (and interesting) example were photographs of Jonathan the tortoise from Plantation House. Appreciation of the 'pure' natural environment also appeared to be significant, especially after correcting for photographer bias using Wood et al. (2013) photographic user days. Areas such as Lot's Wife Ponds, Sandy Bay and Peaks National Park did have a marked presence in the dataset of Flickr photographs.

The methodology of using geolocated social media data to map interest in the natural, built or historical environment appears to be feasible, although with some qualifications. Instagram, because of its different ways of capturing posts' location, and because it does not allow access to its archives, was deemed unsuitable for St Helena since only seven posts could be retained for a period of six months. Flickr on the other hand, proved much more useful, although the dataset was dominated by a few photographers and particular years. Nevertheless, Flickr managed to capture both natural, historical and built heritage areas of interest in St Helena, and we believe could be a valuable tool for mapping interest in the natural, historical or built environment in other areas of the South Atlantic with no or little internet connections – and beyond. With the new fibre optic cable bringing faster and cheaper internet to St Helena in 2020, this method could prove even more effective for long term monitoring.