Moataz Medhat ElQadi, Alan Dorin, Adrian Dyer, Martin Burd, Zoë Bukovac, Mapping species distributions with social media geo-tagged images: Case studies of bees and flowering plants in Australia. Ecological Informatics (2016), Version of record: doi: 10.1016/j.ecoinf.2017.02.006

Mapping Species Distributions with Social Media Geo-Tagged Images: Case studies of bees and flowering plants in Australia

Moataz Medhat ElQadi[1]*, Alan Dorin[1], Adrian Dyer[2], Martin Burd[1], Zoë Bukovac[1]

[1] Monash University, Clayton, Australia, 3800

[2] RMIT University, Melbourne Australia, 3000

Abstract

Data sources on species distribution and range are typically expensive and time consuming to build, and traditional survey techniques often have spatial, temporal, or scale-related gaps. Social network sites, on the other hand, can provide massive amounts of cost effective data that may potentially yield information of direct benefit to supplement and understand ecological phenomena. Previous research explored using social network site content to enhance information collected by experts or professional surveys in domains including species distribution and land cover. However, the data quality and general suitability of social network sites data for answering questions related to species distribution and range is highly variable and this aspect of its value to science remains underexplored.

In this research we investigate some causes of social network site data unreliability and explore how to mitigate it. We filter data points based on our estimates of reliability and relevance. We then use the filtered data to infer species ranges and distributions in concert with Global Biodiversity Information Facility (GBIF) data. Our proposed methodology was applied to four Australian case studies including two insect pollinators, and two flowering plants. The case studies were chosen from Australia because of its unique geographical features, large landmass, sparse population, and the many tourists and residents who travel across it taking photos and sharing them through social media. We show that, despite some barriers, there are instances where the social network site data clearly complements the existing source, making our technique a valuable means of making repeatable, efficient additions to traditional species distribution data.

Keywords: bio-diversity data, social network sites, geo-tagged images, species distribution mapping

1 Introduction

Climate change and habitat alteration threatens our planet's biodiversity, food security and, of particular relevance to this article, insect pollinator populations (Kjøhl, Nielsen, & Stenseth, 2011). It is estimated that one third of world food production requires pollination by animals, especially insects

* Corresponding authors {moataz.mahmoud / alan.dorin} @ monash.edu

(Klein et al., 2007), and that the global economic value of insect pollination in agriculture is approximately €153 billion annually (Gallai, Salles, Settele, & Vaissière, 2009). Information on the geographic ranges and abundances of insects is important to understand, predict, and manage pollinator services (Biesmeijer et al., 2006; Moritz, Kraus, Kryger, & Crewe, 2007). Databases of species occurrence, such as the Global Biodiversity Information Facility (GBIF, www.gbif.org), can play an important role in research on the effects of climate change and habitat alteration on pollinator availability. GBIF is a data source based on biodiversity records of participating institutions and governments, but there are often inconvenient gaps in its data (Robert P. Anderson, 2016). For instance, some literature (Beck, Ballesteros-Mejia, Nagel, & Kitching, 2013) specifically addresses the inventory completeness of a tropical insect, hypothesising that it is impacted by human factors including "road and tourism infrastructure, habitat encroachment, population density, conflict and colonial history". Filling such gaps is an important challenge for both biologists and information scientists.

GBIF species occurrence data is sometimes without supporting photographs or video, even though such image-based media can be very useful. For instance, images may enable ecologists to determine attributes of the specimens that are not reported in the textual data. Additionally, the potential to better inform or supplement biodiversity research with social media images may have widespread research value. Such a framework may also lead to improved or targeted use of public science for contributing to important research questions.

In contrast to GBIF, Flickr (www.Flickr.com) is a photo sharing social network site that is not usually considered a formal source of information on species distribution for ecological research. Nevertheless, in the last three years Flickr had an average of two million public photos added every day worldwide (Michel, 2016). While few of these images are likely to be relevant to bio-diversity research, the pool is so large that the ability to extract even a very small fraction of useful images may help extend our current knowledge of species ranges, and particularly of ranges undergoing rapid change. Other researchers (Barve, 2014; Beck et al., 2013) have previously demonstrated the possibilities that Social network sites like Flickr offer for filling gaps in our knowledge about species ranges, especially given the precise temporal and spatial information encapsulated in digital image file metadata. Our own study builds upon this growing body of knowledge by enhancing aspects of the method. Also, we apply the idea in a different geographical region, with its own unique character, to discover the extent to which it may apply beyond its initial application.

Social network sites (SNS) have been used as a source of data in a wide variety of research contexts. For example they have been applied to detect outbreaks of epidemics (Culotta, 2010), for measuring public opinion and sentiment (O'Connor, Balasubramanyan, Routledge, & Smith, 2010), and to predict outcomes of elections (Tumasjan, Sprenger, Sandner, & Welpe, 2010). In the context of conservation ecology, previous research has used Flickr to enhance the data quality of the Coordination of information on the environment (Corine) land cover maps (Estima & Painho, 2013; Jacinto & Marco, 2014), and in monitoring invasive species (Daume, 2016). Kirkhope et al. (2010) used images uploaded in Flickr by volunteers in a bee identification project, and Barve (2014) explored the usefulness of Flickr images as occurrence records for the Monarch butterfly and the Snowy owl.

While these efforts have been promising, questions surrounding taxonomic accuracy and scope and other data quality issues remain. For instance, "Questions still remain concerning how many of the SNS-derived records will pass data quality and fitness-for-use tests", says Barve (2014). Stafford et al. (2010) found that even images solicited in a citizen science project did not conform to instructions intended to ensure geographic accuracy in about 40% of cases. Nonetheless, the potential to mine SNSs for useful biodiversity information has been sufficiently well demonstrated that it now makes sense to elaborate and refine the search techniques and seek improved validation of potential occurrence data. Thus, in the current study, we seek to improve validation of data sourced though social media using the framework suggested by Barve (2014), and discover how this might operate in a very large, sparsely populated country like Australia that is, nevertheless, technologically well resourced.

A second important issue for the use of data obtained from social network sites is that images are typically added by non-experts. So, it is probably helpful to search using species' common names (e.g. Honeybee), and scientific names (e.g. *Apis mellifera*).

In this paper we use a novel, accessible approach to test the relevance of geo-tagged image content to search keywords (section 2). The filtered images, together with ALA species distribution data, contribute to the construction of land map overlays. Both sets of data are compared to investigate how the filtered Flickr images complement the available information on species distribution obtained from ALA (section 3). We outline our methodology below, and then present our results for test cases, along with a discussion of the strengths and weaknesses of this new research tool (section 4).

2 Methodology

In overview, we searched for geo-tagged images in Flickr using species common and scientific names. These images were subsequently fed into Google's reverse image-search to find tags that best describe the content of these images. These tags were next used to exclude images deemed irrelevant to the studied species. The scientific name of the species is used to search the Atlas of Living Australia (ala.org.au), henceforth referred to as ALA, a data source that is a participant node of GBIF. The filtered images from Flickr, and the occurrences from ALA, are overlayed on a geographical map to allow us to draw comparisons and conclusions. An outline of the methodology is depicted in figure 1 and each step is detailed below.

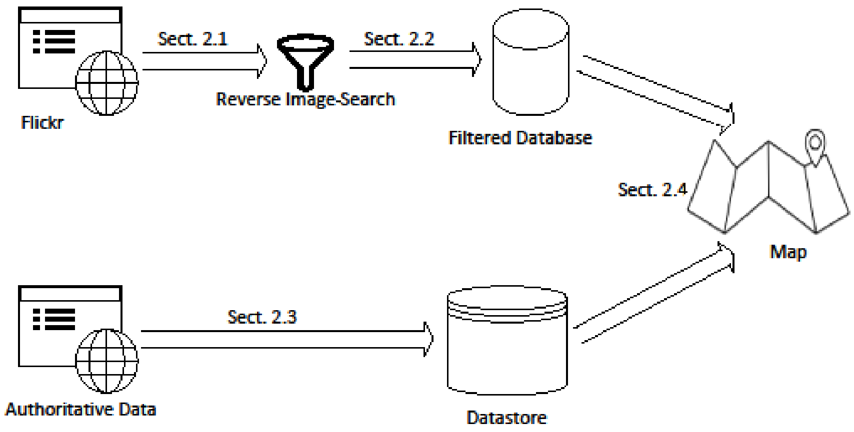


Figure 1. Overview of the process: Data from Flickr are filtered using reverse image-search. This filtered data is overlayed with the ALA data for further analysis.

2.1 Flickr image retrieval

Flickr was selected as the social network site with which to test our study for several reasons. It is currently very popular with photographers, images uploaded can include detailed spatial location information (i.e. they are *geo-tagged*), and Flickr also has an Application Programming Interface, (API), that consists of a set of callable methods allowing search by keywords and location in the same query. This was a useful property, allowing for validation type searching and cross-referencing to other data sources.

There exist many API kits that implement wrappers of the Flickr HTTP API in various programming languages. We used the Python-Flickr-api (Mignon, 2016) which we edited slightly to allow storing image URLs instead of having to store complete image files. The image URLs were saved to a database along with the specific search keywords used, and image location coordinates.

2.2 Image content check

Images returned from search queries in Flickr may not always be related to the species sought. For this reason, image content requires a separate stage of validation. We used Google's reverse image-search, a tool for finding images that are visually similar to an input reference image. As a result of this search, both a text-label estimate of the image subject, and a set of visually similar images are returned as potential matches. Using Google search's result, we can thus differentiate between images of Honeybees, and, for example, images of jars of honey. The relevance of each member of the set of images previously returned from Flickr, can now be assessed by referring to its associated text tag.

Our filtering is therefore a coarse but useful way to remove images that may be related to the search keyword in ways irrelevant to the ecological study goals.

2.3 Obtaining reference data

Data on species distribution are usually collected by experts or citizen scientists, then subsequently vetted by experts, before being standardised and published. Due to their quality and the recognition by researchers, we benchmark our data against available data for Australia obtained from ALA, a node of GBIF.

2.4 Geographic map overlay

In order to visualise the obtained data, ArcGIS software (ESRI, 2015) was used to create geographic maps. The locations associated with the filtered geo-tagged images (section 2.2) and ALA data (section 2.3) were plotted on the map. The ALA reference points were buffered to create polygons on the map extending a distance of 100 Km around species' occurrence points. The 100 km buffer is a coarse proxy for the variable maximum coordinate uncertainty in the obtained ALA reference data. It has been chosen for clarity, keeping in mind the scale of the map on which we have plotted the data. Actual uncertainty range values in the dataset were: 100 km Blue-banded bee, 125 km Sturt's Desert Pea, 125 km Pink Heath, 10 km Honeybee. The Flickr image locations were then overlayed as shown in figures 6, 7, 8, and 9. Results are discussed in section 3.

2.5 Case study selection

The methodology described in section 2 was applied to four case studies: two insect pollinators and two flowering plants. The pollinators were the Honeybee (*Apis mellifera*: Apidae) and the Blue-banded bee (*Amegilla cingulata*: Apidae), an Australian native pollinator. The flowering plants were Sturt's Desert Pea (*Swainsona formosa*: Fabaceae) and Pink Heath (*Epacris impressa*: Ericaceae), both native Australian species.

The two insects were chosen because both are reasonably distinctive but differ in abundance, distribution, and behaviour. The Honeybee is an important pollinator of agricultural and horticultural crops and natural ecosystems around the world, and has almost iconic public recognition from school aged children to seniors in the community. The Honeybee is an introduced, but abundant and economically important pollinator on mainland Australia, the domain of our current study. Additionally, the Honeybee is relatively slow moving, making it an easy and popular subject for amateur photographers. The Blue-banded bee by contrast, although relatively common in many parts of Australia, forms solitary nests rather than large colonies and so is not nearly as abundant as the Honeybee. Perhaps consequently, it is not well recognised by the general public. Also, the Blue-banded bee is not currently well known as a pollinator of global significance, although there is interest in how it might be employed as a pollinator in some circumstances (Hogendoorn, Coventry, & Keller, 2007; Switzer, Hogendoorn, Ravi, & Combes, 2016). Hence, it makes an interesting contrast against which to compare the utility of Flickr user images for the purposes of collecting species range data.

The floral case studies were chosen from the Australian states' emblems. Each of the seven states and two territories into which Australia is divided has a floral emblem that is loosely associated with the region. The South Australian emblem, the Sturt's Desert Pea, has a bright red flower with very unusual structure. The Victorian emblem, the Pink Heath, by contrast, is certainly bright and striking *en masse*, but structurally it is not as iconic as the South Australian emblem. Both are short perennials of open woodlands, but Sturt's Desert Pea occurs more frequently in the continental interior while Pink Heath is found in cooler habitats in the southeast. These flowers, perhaps surprisingly to an outsider, are not uniformly well known to the public, despite their official status as state emblems. Because the two species differ in the degree of visual spectacle they offer, Pink Heath is a less popular subject for photography than the Sturt's Desert Pea, making the combination of the two floral emblems interesting cases for this study.

3 Results

In section 3.1, we present the results of applying our image content filter (section 2.2) to our case studies (section 2.5). The findings obtained from laying the data over geographic maps (section 2.4) are presented in section 3.2.

3.1 Image content validation

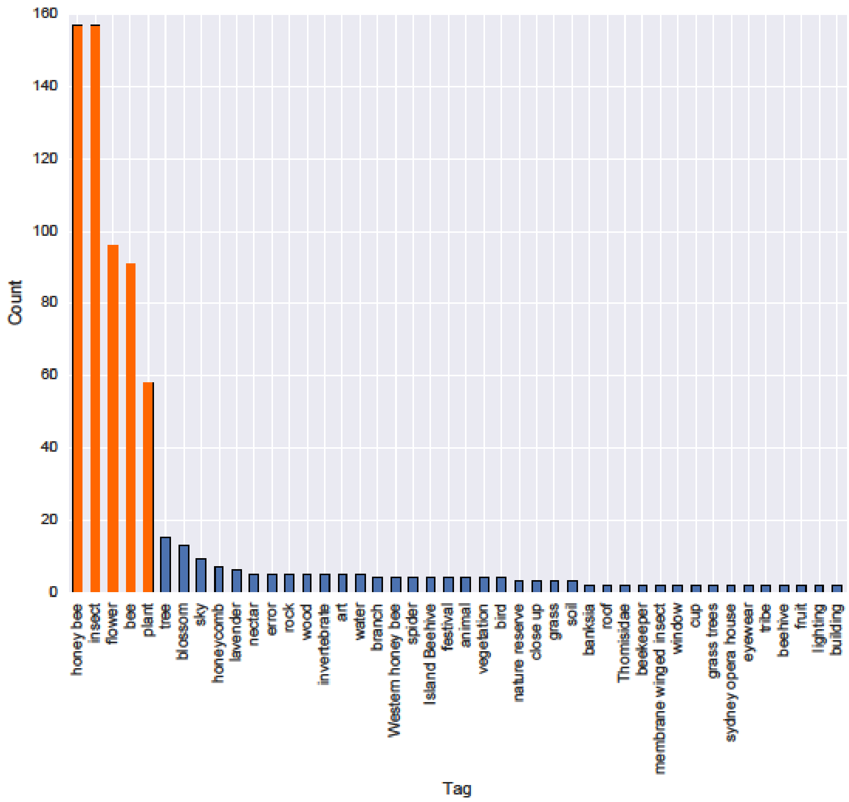
Flickr was searched for our case studies using the scientific and common names shown in Table 1. The table also indicates the number of images returned by each search, and the percentage of these misclassified as false positives and false negatives using the image text tags returned by Google's reverse image-search. The image text tags' frequencies are plotted in detail in Figures 2, 3, 4, and 5 for each case study.

In the case of the Honeybee, the five most frequent image text tags accounted for 69% of all images returned. By visual inspection of the images themselves, these tags were determined to match the intent of the search. In the case of the Blue-banded bee, the six most frequent tags were intuitively related to the search term, except for the tag "performance" which related instead to a performance of Debbie Harry (Blondie) and her rock band. The remaining five relevant tags accounted for 68% of all images.

Table 1 highlights the extent to which the non-expert community of photo-sharers chooses to label their uploads using common name labels, rather than scientific names, as well as the variation in this practice (from 2% for Sturt's Desert Pea to 28% for the Pink Heath). Hence, at this step in the process of data collection from the SNS, it seems worthwhile to conduct a search using both types of name.

Table 1. Flickr search results for Australian case studies[1]*. Search terms were species scientific and common names. Percentages are provided for the fraction of images returned using scientific names over the total number of returned images. "False positive %" refers to the percentage of images included in the data set after filtering using Google reverse image-search that were then determined by visual inspection not to be photographs of the target species. "False negative %" refers to the percentage of images excluded from the data set after filtering using Google reverse image-search that were then determined by visual inspection to be photographs of the target species.*

Scientific name (No. of images)	Common name (No. of images)	Images returned by scientific name %	No. of images included by tag (False positive %)	No. of images excluded by tag (False negative %)
Apis Mellifera (50)	honey bee (759)	6%	559 (12%)	250 (13%)
Amegilla cingulate (23)	Blue banded bee (342)	6%	248 (13%)	117 (27%)
Swainsona Formosa (7)	Sturt pea (419)	2%	310 (3%)	116 (34%)
Epacris impressa (103)	Pink Heath (267)	28%	223 (43%)	147 (13%)



[1] Search dates: 1/Jun/2016 Honeybee, 16/Jun/2016 Blue-banded bee and Sturt's Desert Pea, 2/Oct/2016 Pink Heath.

Figure 2. Flickr was searched for images of the Honeybee (Apis mellifera). The results were passed to Google reverse image-search. The frequencies of tags returned by Google from this image set are shown. The five most frequent tags (in orange) are relevant, they account for 69% of the 809 images returned. Note that a long tail of tags with frequency 1 is not shown in the graph.

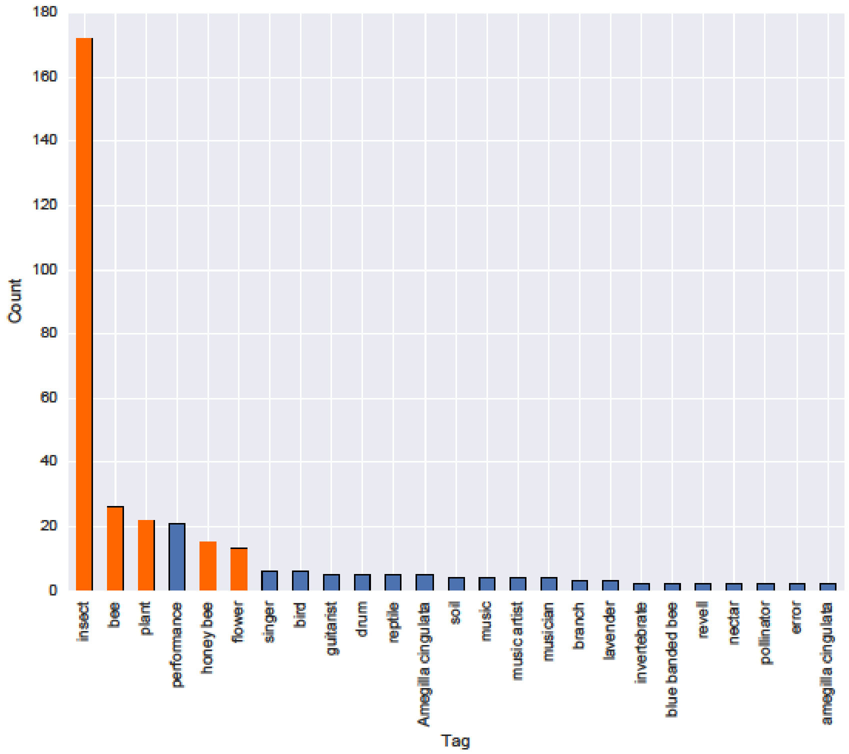


Figure 3. Flickr was searched for images of the Blue-banded bee (Amegilla cingulata). The results were passed to Google reverse image-search. The frequencies of tags returned by Google from this image set are shown. Five out of the six most frequent tags are relevant (in orange), they account for 68% of the 365 images returned while the irrelevant "Music Performance" tags successfully excluded account for ~13% of all images.

A search for Sturt's Desert Pea returned 426 images. Google reverse image-search classified a 73% majority of these into two categories, "plant" and "flower", and we visually confirmed that each related to our search intent. Applying the same process to Pink Heath yielded 370 images, with a 60% majority divided between the four tags "plant", "flower", "Epacris impressa", and "Epacris", in decreasing order of frequency.

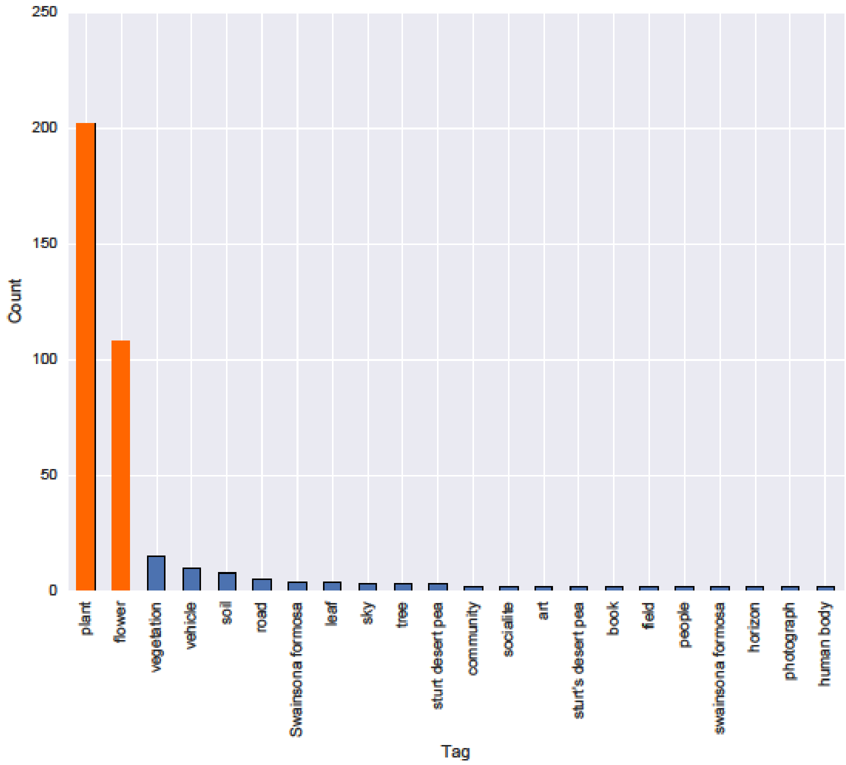


Figure 4. Flickr was searched for images of Sturt's Desert Pea (Swainsona formosa). The results were passed to Google reverse image-search. The frequencies of tags returned by Google from this image set are shown. The two most frequent tags (in orange) are relevant, they account for ~73% of the 426 images returned.

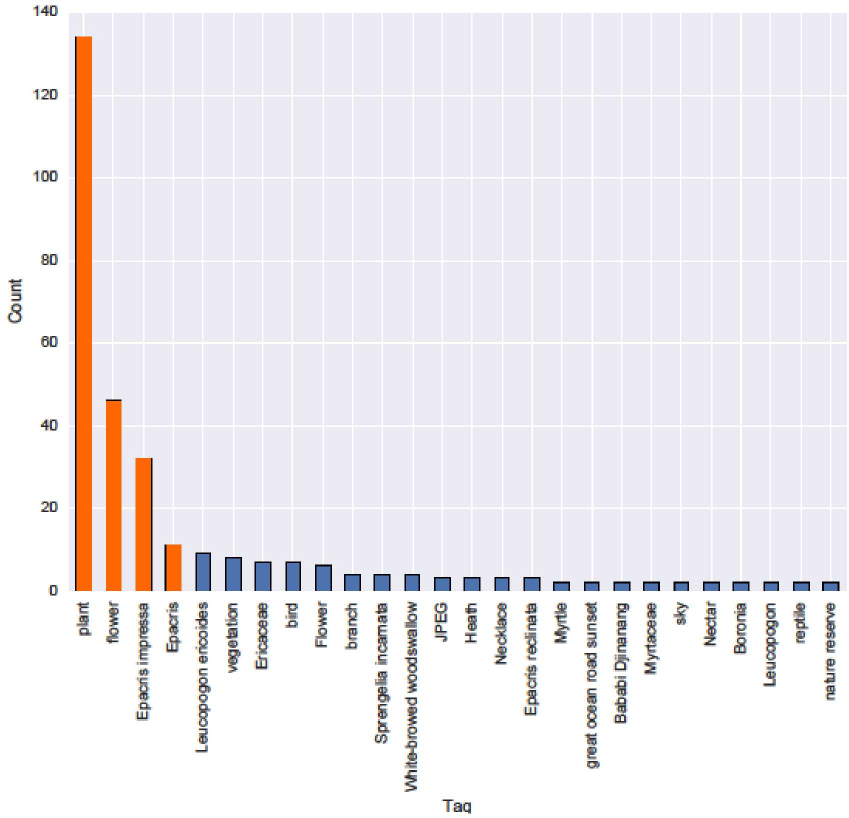


Figure 5. Flickr was searched for images of Pink Heath (Epacris impressa). The results were passed to Google reverse image-search. The frequencies of tags returned by Google from this image set are shown. The four most frequent tags (in orange) are relevant, they account for 60% of the 370 images returned.

3.2 Geographic Results

Data obtained from ALA and from social network site (Flickr) were overlaid on a map of Australia for each of the species in our case studies. Data points obtained from Flickr appear to expand the ALA-based ranges. For instance, the Honeybee map (Figure 6) shows that the Flickr data generally confirm the ALA data, but extend these data on the East coast (close to Australian urban centres) and in the centre of the continent (near Alice Springs, a popular tourist destination, remote from major cities).

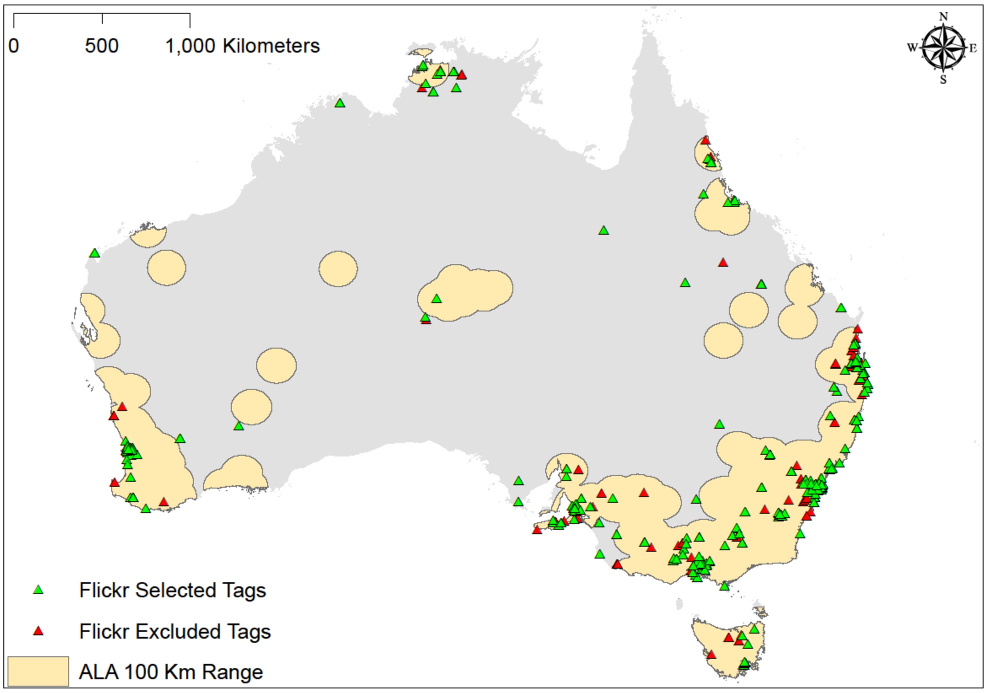


Figure 6. Honeybee (Apis mellifera) images obtained from Flickr (triangle symbols) mostly fall within a 100 km region around ALA reference points (region in yellow, ALA points not shown).

In the Blue-banded bee map (Figure 7), the Flickr points generally fall in the range known from ALA close to the coast, while suggesting new habitat in the centre, north and south of the continent. We have confirmed by visual inspection that the images in these locations warrant inclusion as insect sightings. Interestingly, the occurrence marked with a square in Figure 7, was of a Neon-cuckoo bee (*Thyreus nitidulus*) which is parasitic on the Blue-banded bee (Cardale, 1968). Hence, its appearance may point to the presence of the Blue-banded bee.

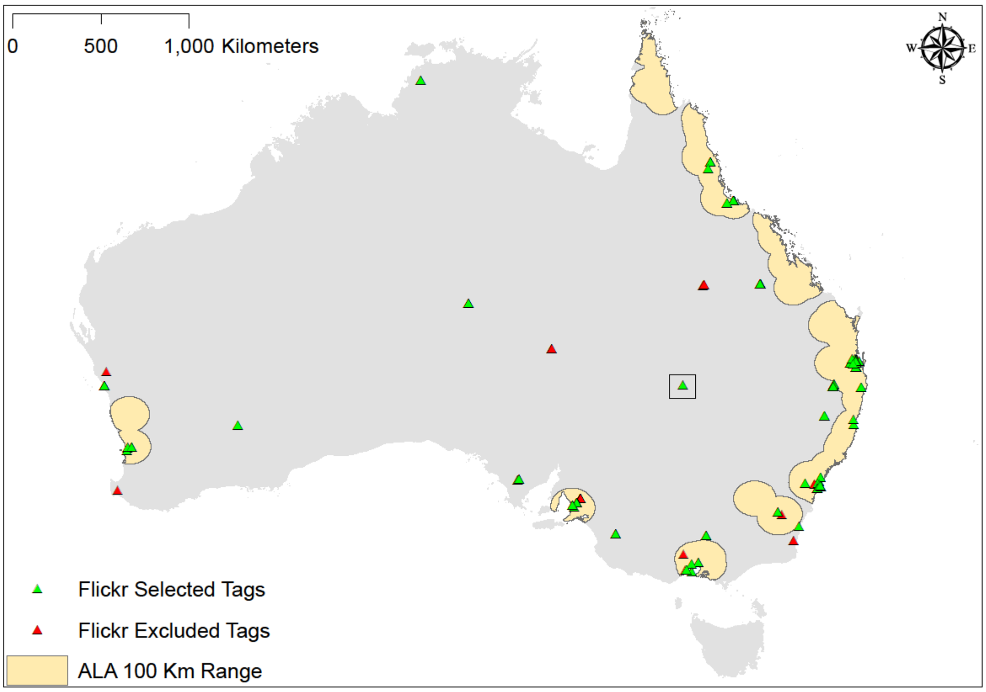


Figure 7. Blue-banded bee (Amegilla cingulata), images obtained from Flickr (triangle symbols) mostly fall within a 100 km region around ALA reference points (region in yellow, ALA points not shown). However, The Flickr filtered images showed occurrences outside the coastal regions marked in ALA (Yellow). The Flickr point inside the square is created from an image of a Neon-cuckoo bee (Thyreus nitidulus) which is parasitic on the Blue-banded bee.

Flickr data also extend the Sturt's Desert Pea range in the centre of the continent (Figure 8). However, although the data suggest the existence of the plant in Victoria, the images in Victoria were examined individually and subsequently found to be located in gardens and museums. Similarly, one specimen in the Northern Territory was found to be a potted plant at a roadside "Holiday Park". In the case of Pink Heath, Flickr data generally coincides with the ALA range in Victoria and Tasmania, but a few images were registered in locations far from the ALA ranges. These images were confirmed to be false matches that happened to have the words 'pink' and 'heath' in their Flickr tags without being specimens of Pink Heath.

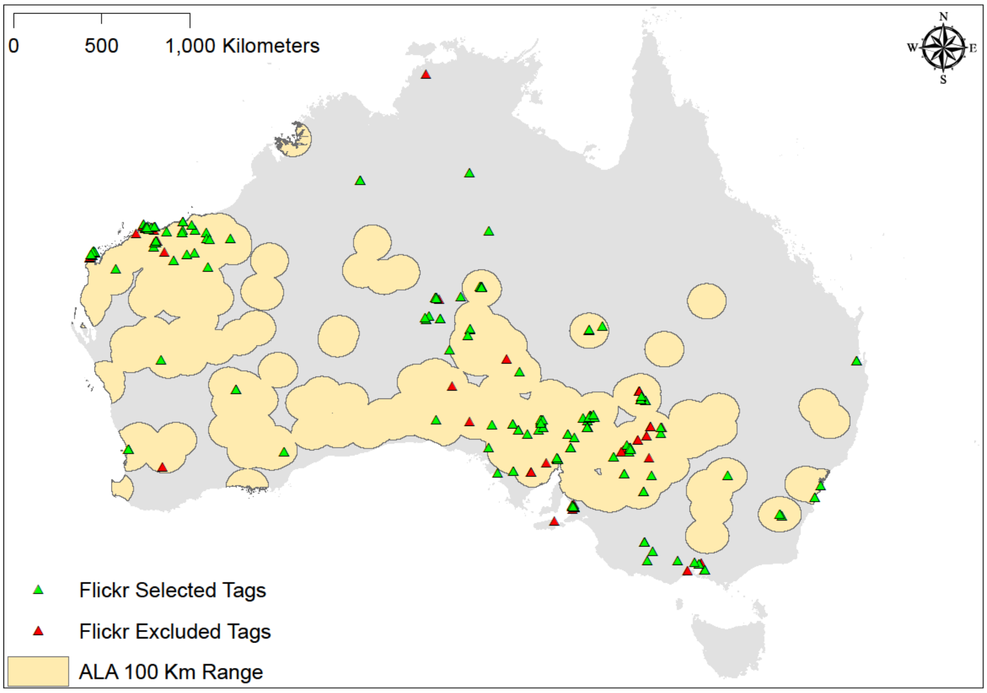


Figure 8. Sturt's Desert Pea (Swainsona formosa) images obtained from Flickr (triangle symbols) mostly fall within a 100 km region around ALA reference points (region in yellow, ALA points not shown). However, although ALA occurrences are missing in some areas in the centre of the continent, Flickr Geo-tagged images cover those regions.

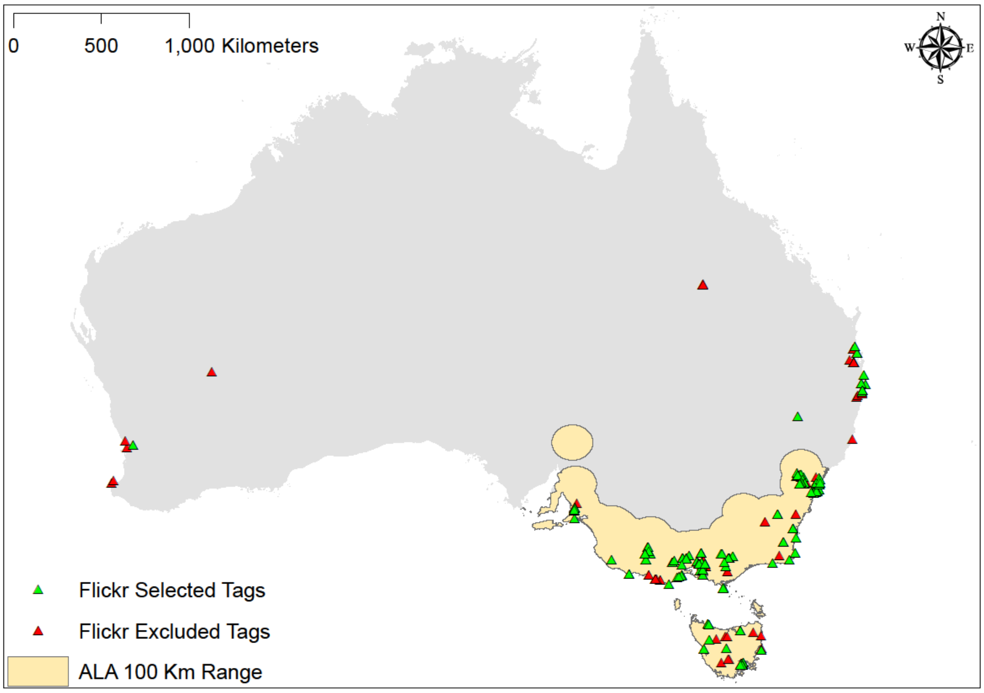


Figure 9 Pink Heath (Epacris impressa) images obtained from Flickr (triangle symbols) mostly fall within a 100 km region around ALA reference points (region in yellow, ALA points not shown). Flickr points far from ALA regions were visually confirmed to be false positives due to similar text keywords.

4 Discussion

Our results suggest that the SNS data can be helpful in determining the existence of species in certain regions. In particular, geo-tagged images can be used to detect or verify the occurrence of a species outside its previously documented or inferred range. Tools for obtaining and filtering such information have many potential uses in a world in which species ranges are shifting at unusually high rates due to climate change (Chen et al., 2011). Timely detection of expanding ranges, for example, can assist in biocontrol and other management efforts directed at invasive species (Fagan et al., 2002).

Our results also show that selecting images identified by the most frequent tags is generally a useful approach, with an error rate (false positive identification) that varies from as little as 3% in the case of Sturt's Desert Pea (*Swainsona formosa*) to 43% for Pink Heath (*Epacris impressa*) (Table 1). Selecting images with the most frequent tags is certain to overlook other tags that are less-frequent yet relevant. This is evident in the case of the Blue-banded bee where 27% of the images excluded based on the tag filtering were found by visual inspection to, in fact, be relevant to our search (Table 1). This is not surprising when we notice that "Ameguilla cingulata" and "blue banded bee" are among the less-frequent tags reported in Figure 3. Similarly, in the case of Sturt's Desert Pea (*Swainsona formosa*), Figure 4 shows that "Swainsona formosa" and "Sturt's desert pea" are among the infrequent, hence excluded, tags. These tags' exclusion by our simple approach accounts for the high number of false negatives reported in Table 1. Our goal is to devise a method that facilitates excluding irrelevant images, based on their tag frequency. If unnecessary exclusion of relevant images is of a concern in an application, for instance in cases where data are rare, we would advise including images tagged with the search keywords (being scientific and common names) in the "relevant" image set, even if these tags were infrequent.

Taxonomic accuracy of SNS images remains a concern (e.g., Stafford et al., 2010), but our experience indicates that visual validation by a specialist can greatly reduce the error rate. The pre-filtering techniques we have proposed assist to save time in this aspect of the process. The combination of machine filtering and human verification may be especially viable for targeted uses of SNS data, such as monitoring efforts directed at a single or limited number of species. In common with others (Barve 2014; Daume 2016), therefore, we see considerable potential in social media sources of biodiversity information and a need to develop tools to realise that potential.

The inherent limitation of SNS-derived biodiversity information must be kept in mind, however. The "social" in "social network sites" denotes the relation of these sites to human social behaviour which is complex, variable and not necessarily aligned with the goals of any ecological study. It is doubtful, for example, whether population density of species should be inferred directly from this data; the density of images in a social network site may be dependent on user activity levels, rather than on the abundance of the photographed subject matter. Figure 10 shows, for instance, how the capital cities are hotspots of high image density.

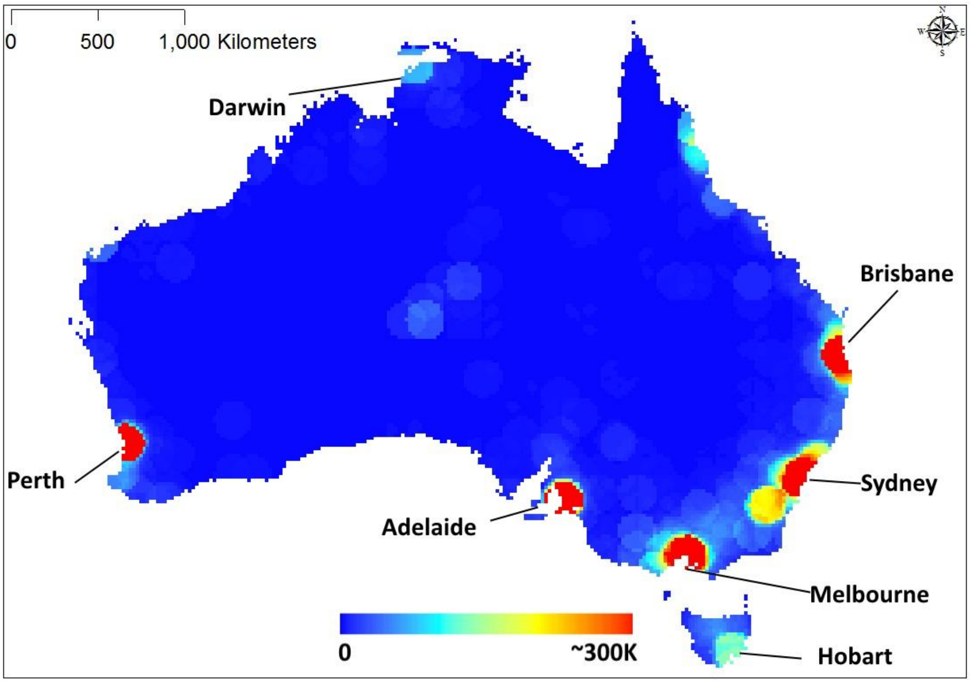


Figure 10. Heat map of all geo-tagged Flickr images in Australia. Capital cities show a much higher density, Oct. 2016

Similarly, inference of the dynamics of species presence over seasons is not recommended using data obtained from SNS (Figure 11), since users' activity inherently changes over time (e.g. in unfavourable weather people may prefer staying indoors rather than taking nature photos).

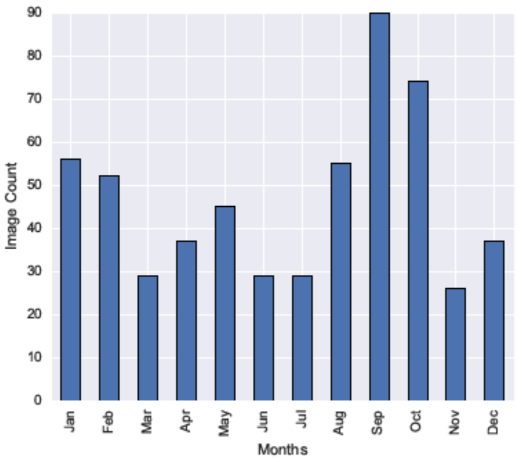


Figure 11 Seasonality of Honeybee images (search terms: "honey bee" and "Apis mellifera") appearing on Flickr tracked monthly. It remains to be seen whether this pattern has been generated by bee activity levels, photographer activity levels, or a combination of both.

Another aspect of human behaviour involved in social networks is people's willingness to share what they deem "interesting". Plants and insects that are common, or are not photogenic, or are difficult to photograph (e.g. insects that are very fast in flight and do not visit flowers for long, such as the Blue-banded bee) may be overlooked as photographic subjects. In other words, the availability of data on species is not necessarily related to their importance to ecologists.

Despite these limitations, SNS images can provide an inexpensive and abundant source of geographic occurrence data. Even limited data on species occurrences may, in conjunction with models of range expansion (Fagan et al., 2002) or range-abundance relationships (Gaston et al., 2000), greatly assist basic research and applications in ecology, conservation and environmental management.

We found Google reverse-image search to be a useful tool, but not at the level of distinguishing between similar insect species unless a separate validation stage was implemented. For instance, it provided the tag “honey bee” in response to an input image of a Blue-banded bee (Figure 3). We confirmed visually that the images on Flickr were in fact Blue-banded bees.

5 Conclusion

Social network sites provide a potential wealth of geo-tagged images that many researchers could use to complement existing ecological information. However, since social network sites are often used by non-specialists, they may suffer from two major problems including that common names are frequently used to describe species which might by synonymous with other objects or events; and uploaded images can be misclassified by non-specialists.

Our new findings suggest that checking the image content using Google reverse image search is useful in filtering out images broadly unrelated to the species sought. However, the method is not sufficiently fine-grained to distinguish between species. Expert human validation is still needed for reliable classifications. Despite this, the effort to manually confirm species classifications among the potentially large amounts of data we have available is relatively minor given our approach of identifying relevant tags. Future image classification systems may improve classification reliability further, and at any rate, the use of human expertise for validation of images is more cost efficient than sending experts to the field, especially in cases when distances between study sites are great.

We have shown that the filtered geo-tagged images from Flickr can in fact complement existing data in many cases, by providing valuable data in locations where the existing data is thin. This research method can also reveal the presence of species in areas not previously considered, allowing for improved planning for more traditionally focussed ecological methodologies.

6 Acknowledgements

This research was supported under Australian Research Council’s Discovery Projects funding scheme (project number DP160100161). Thank you to Mani Shrestha for discussions regarding plant species distribution.

7 References

Barve, V. (2014). Discovering and developing primary biodiversity data from social networking sites: A novel approach. *Ecological Informatics, 24*, 194-199.

Beck, J., Ballesteros-Mejia, L., Nagel, P., & Kitching, I. J. (2013). Online solutions and the 'Wallacean shortfall': what does GBIF contribute to our knowledge of species' ranges? *Diversity and Distributions, 19*(8), 1043-1050.

Biesmeijer, J. C., Roberts, S., Reemer, M., Ohlemüller, R., Edwards, M., Peeters, T., . . . Thomas, C. (2006). Parallel declines in pollinators and insect-pollinated plants in Britain and the Netherlands. *Science, 313*(5785), 351-354.

Cardale, J. (1968). Parasites and other organisms associated with nests of Amegilla Friese (Hymenoptera: Anthophorinae). *Austral Entomology, 7*(1), 29-34.

Culotta, A. (2010). *Towards detecting influenza epidemics by analyzing Twitter messages.* Paper presented at the Proceedings of the first workshop on social media analytics.

Daume, S. (2016). Mining Twitter to monitor invasive alien species—An analytical framework and sample information topologies. *Ecological Informatics, 31*, 70-82.

ESRI. (2015). ArcGIS for Desktop (Version 10.4). Retrieved from http://www.esri.com/software/arcgis/arcgis-for-desktop

Estima, J., & Painho, M. (2013). Flickr geotagged and publicly available photos: Preliminary study of its adequacy for helping quality control of corine land cover *Computational Science and Its Applications–ICCSA 2013* (pp. 205-220): Springer.

Gallai, N., Salles, J.-M., Settele, J., & Vaissière, B. E. (2009). Economic valuation of the vulnerability of world agriculture confronted with pollinator decline. *Ecological Economics, 68*(3), 810-821.

Hogendoorn, K., Coventry, S., & Keller, M. A. (2007). Foraging behaviour of a blue banded bee, Amegilla chlorocyanea in greenhouses: implications for use as tomato pollinators. *Apidologie, 38*(1), 86-92. doi: 10.1051/apido:2006060

Jacinto, E., & Marco, P. (2014). Photo Based Volunteered Geographic Information Initiatives: A Comparative Study of their Suitability for Helping Quality Control of Corine Land Cover. *International Journal of Agricultural and Environmental Information Systems (IJAEIS), 5*(3), 73-89. doi: 10.4018/ijaeis.2014070105

Kirkhope, C. L., Williams, R. L., Catlin-Groves, C. L., Rees, S. G., Montesanti, C., Jowers, J., . . . Goodenough, A. E. (2010). *Social networking for biodiversity: the BeeID project.* Paper presented at the Information Society (i-Society), 2010 International Conference On.

Kjøhl, M., Nielsen, A., & Stenseth, N. C. (2011). *Potential effects of climate change on crop pollination*: Food and Agriculture Organization of the United Nations (FAO).

Klein, A.-M., Vaissiere, B. E., Cane, J. H., Steffan-Dewenter, I., Cunningham, S. A., Kremen, C., & Tscharntke, T. (2007). Importance of pollinators in changing landscapes for world crops. *Proceedings of the Royal Society of London B: Biological Sciences, 274*(1608), 303-313.

Michel, F. (2016). How many public photos are uploaded to Flickr every day, month, year? Retrieved 31/8/2016, from https://www.flickr.com/photos/franckmichel/6855169886/in/photostream/

Mignon, A. (2016). python-flickr-api [Python package]: github.com. Retrieved from https://github.com/alexis-mignon/python-flickr-api

Moritz, R. F., Kraus, F. B., Kryger, P., & Crewe, R. M. (2007). The size of wild honeybee populations (Apis mellifera) and its implications for the conservation of honeybees. *Journal of Insect Conservation, 11*(4), 391-397.

O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM, 11*(122-129), 1-2.

Robert P. Anderson, M. A., Antoine Guisan, Jorge M. Lobo, Enrique Martínez-Meyer, A. Townsend Peterson, Jorge Soberón. (2016). Report of the Task Group on GBIF Data Fitness for Use in Distribution Modelling. http://www.gbif.org/resource/82612

Stafford, R., Hart, A. G., Collins, L., Kirkhope, C. L., Williams, R. L., Rees, S. G., . . . Goodenough, A. E. (2010). Eu-social science: the role of internet social networks in the collection of bee biodiversity data. *PloS one, 5*(12), e14381.

Switzer, C. M., Hogendoorn, K., Ravi, S., & Combes, S. A. (2016). Shakers and head bangers: differences in sonication behavior between Australian Amegilla murrayensis (blue-banded bees) and North American Bombus impatiens (bumblebees). *Arthropod-Plant Interactions, 10*(1), 1-8. doi: 10.1007/s11829-015-9407-7

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment. *ICWSM*, *10*, 178-185.