"photosearcher" package in R: An accessible and reproducible method for harvesting large datasets from Flickr

Nathan Fox, Tom August, Francesca Mancini, Katherine E Parks, Felix Eigenbrod, James M Bullock, Louis Sutter, Laura J Graham

nf2g13@soton.ac.uk

https://doi.org/10.1016/j.softx.2020.100624

UNIVERSITY OF Southampton

UKRI Natural Environment Research Council

UK Centre for Ecology & Hydrology

@n_fox

Introduction

Flickr (*flickr.com*) contains a large database of photographs, with up to 25 million new uploads a day (Ding & Fan 2019)

The metadata of the photographs can contain both spatial and temporal information

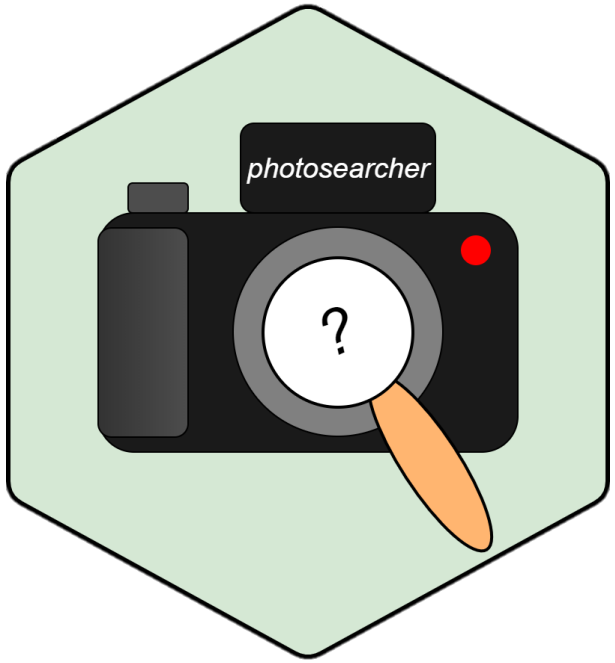
Metadata is accessible through the Flickr Application Programming Interface (API)

Overcomes many of the limitations of extensive large-scale social and ecological surveys

@n_fox

photosearcher R package

We developed the photosearcher R package (Fox *et al.* 2020; https://doi.org/10.1016/j.softx.2020.100624)



The package provides accessible and reproducible functions for searching for photograph metadata

Designed to overcome the limitations of the Flickr API

Current version available from rOpenSci at: https://github.com/ropensci/photosearcher

@n_fox

Overcoming Flickr API limitations

Limitation of API	***photosearcher* r package**
Accessibility is limited by the need for advanced coding skills	Provides a relatively simple method, accompanied by a user guide
Studies often do not share reproducible code	Method is easily shareable and reproducible
Maximum of 4,000 unique photographs returned per search	Automatically and dynamically splits searches to overcome the 4,000 results limit
You cannot search Flickr for images within a shapefile (Lee *et al*. 2019)	Provides functionality to search for any images that are within a shapefile

@n_fox

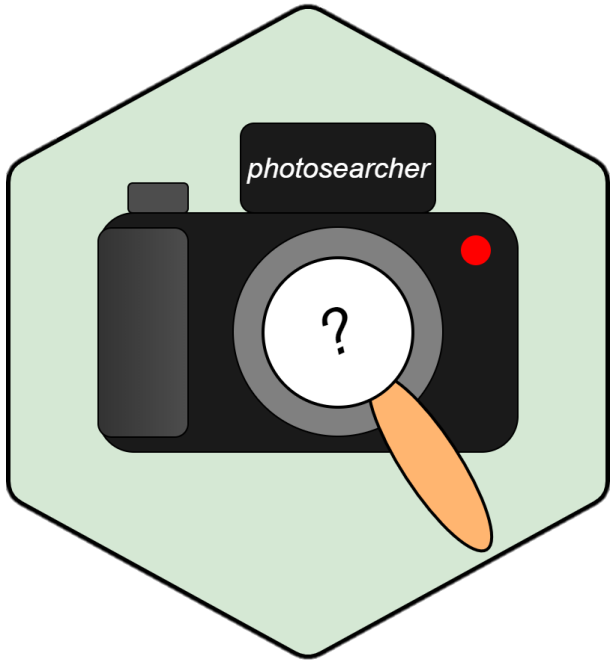
Uses of Flickr data from photosearcher

Assessing cultural ecosystem services:

- Recreational services (Graham & Eigenbrod 2019) *
- Wildlife watching (Mancini *et al*. 2019) *
- Assessing changes in cultural values (Thiagaraja *et al*. 2015)
- Visitation rates in protected areas (Kim *et al*. 2019)

Species distribution data:

- Identifying floral species (August *et al.* 2020) *
- Monitoring migratory and hibernation patterns (Fox *et al.* 2020)
- Tracking the spread of invasive species (Allain 2019)
- Assessing niche segregation (Peña-Aguilera *et al*. 2019)



* used early releases of photosearcher

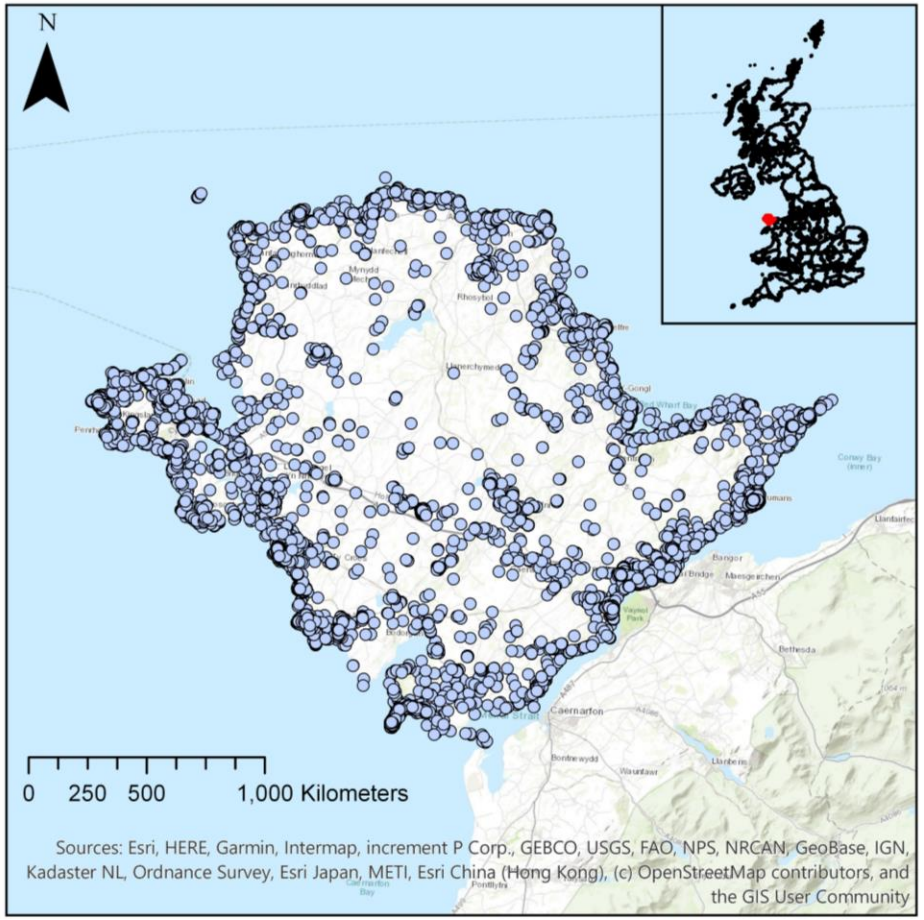
@n_fox

Example – Anglesey Geopark, Wales

Searches for spatial metadata are relatively straight forward and quickly provide large quantities of data:

```
photo_search(mindate_taken = “2000-01-01”,
        maxdate_taken = “2020-01-01”,
        sf_layer = sf::read_sf(“.\\area.shp”))
```

This search returned thousands of points in just a few minutes



@n_fox

Future Direction

- Add other API functions to the package If other available metadata is useful for environmental research
- Add searches for other social media sites: currently working on Reddit and Tumblr
- Please reach out to me if you would be interested in helping develop the package

Thank you for listening; I welcome any questions

@n_fox

References:

- Allain, S., 2019. Mining Flickr: a method for expanding the known distribution of invasive species. *Herpetological Bulletin*, **148**, pp.11-14.
- August, T.A., Pescott, O.L., Joly, A. and Bonnet, P., 2020. AI naturalists might hold the key to unlocking biodiversity data in social media imagery. *Patterns*, 1(7), p.100116.
- Ding, X., & Fan, H. (2019). Exploring the Distribution Patterns of Flickr Photos. *ISPRS Int. J. Geo-Inf,* **8**, 418.
- Fox, N., August, T., Mancini, F., Parks, K.E., Eigenbrod, F., Bullock, J.M., Sutter, L. and Graham, L.J., 2020. "photosearcher" package in R: An accessible and reproducible method for harvesting large datasets from Flickr. *SoftwareX*, 12, p.100624.
- Graham, L. J., & Eigenbrod, F. (2019). Scale dependency in drivers of outdoor recreation in England. *People and Nature*, **1**(3), 406–416.
- Kim, Y., Kim, C. ki, Lee, D. K., Lee, H. woo, & Andrada, R. I. T. (2019). Quantifying nature-based tourism in protected areas in developing countries by using social big data. *Tourism Management*, **72**, 249–256.
- Lee, H., Seo, B., Koellner, T., & Lautenbach, S. (2019). Mapping cultural ecosystem services 2.0 – potential and shortcomings from unlabeled crowd sourced images. *Ecological Indicators*, **96**, 505-515
- Mancini, F., Coghill, G. M., & Lusseau, D. (2019). Quantifying wildlife watchers' preferences to investigate the overlap between recreational and conservation value of natural areas. *Journal of Applied Ecology*, **56**(2), 387– 397.
- Peña-Aguilera, P., Burguillo-Madrid, L., Barve, V., Aragón, P. and Jiménez-Valverde, A., 2019. Niche segregation in Iberian Argiope species. *The Journal of Arachnology*, **47**(1), pp.37-44.
- Thiagarajah, J., Wong, S. K. M., Richards, D. R., & Friess, D. A. (2015). Historical and contemporary cultural ecosystem service values in the rapidly urbanizing city state of Singapore. *Ambio*, **44**(7), 666–677.