

A text and image analysis workflow using citizen science data to extract relevant social media records: Combining red kite observations from Flickr, eBird and iNaturalist

Maximilian C. Hartmann^{a,*}, Moritz Schott^b, Alishiba Dsouza^c, Yannick Metz^d, Michele Volpi^e, Ross S. Purves^a

^a University of Zurich, Department of Geography, Switzerland

^b Heidelberg University, Department of Geography, Germany

^c University of Bonn, Data Science & Intelligent Systems, Germany

^d University of Konstanz, Department of Computer and Information Science, Germany

^e ETH Zurich, Swiss Data Science Center, Switzerland

ARTICLE INFO

Keywords:

User-generated content
Volunteered geographic information
Data integration
Image content analysis
Convolutional neural networks

ABSTRACT

There is an urgent need to develop new methods to monitor the state of the environment. One potential approach is to use new data sources, such as User-Generated Content, to augment existing approaches. However, to date, studies typically focus on a single data source and modality. We take a new approach, using citizen science records recording sightings of red kites (*Milvus milvus*) to train and validate a Convolutional Neural Network (CNN) capable of identifying images containing red kites. This CNN is integrated in a sequential workflow which also uses an off-the-shelf bird classifier and text metadata to retrieve observations of red kites in the Chilterns, England. Our workflow reduces an initial set of more than 600,000 images to just 3065 candidate images. Manual inspection of these images shows that our approach has a precision of 0.658. A workflow using only text identifies 14% less images than that including image content analysis, and by combining image and text classifiers we achieve almost perfect precision of 0.992. Images retrieved from social media records complement those recorded by citizen scientists spatially and temporally, and our workflow is sufficiently generic that it can easily be transferred to other species.

1. Introduction

Environmental change is one of the key challenges of the 21st century. Understanding this change and developing effective policies to mitigate it, leads to a need for data which can be used to develop indicators capturing, for example, drivers, pressures, states, impacts and responses of the environment (Masó et al., 2020). Despite the massive growth in data production witnessed in recent decades, the United Nations Environment Programme recently reported that for 69% of all environmental indicators insufficient data were available to allow continuous monitoring (UNEP, 2019). This lack of data has led to increasing interest in new data sources, and their possible use in the production of official statistics, as demonstrated through the creation of the UN Global Working Group on Big Data for Official Statistics which explicitly recognises social media data as a potential source.

In parallel to this policy driven need for new data sources as inputs to indicators, researchers in Geographic Information Science have sought to describe the properties of new data forms exploiting, as Goodchild (2007) memorably proposed, the idea of ‘citizen sensors’. This has led to the use of a wide range of terminology characterising digital data ranging from the general e.g. User-Generated Content (UGC), through the more explicitly spatial Volunteered Geographic Information (VGI) (Goodchild, 2007) and distinguishing between different modes of production (active and passive crowdsourcing (Haklay, 2013)). Most recently, the potential of digital platforms has led to a rapid growth in projects categorised as citizen science, where the broader public actively contributes to scientific knowledge, through data collection, analysis and hypothesis generation (Trojan et al., 2019).

One area currently witnessing particularly rapid growth is the application of new data sources to create a wide range of ecological

* Corresponding author at: 8057 Zurich, Winterthurerstrasse 190, Y25 J 82, Switzerland.

E-mail address: maximilianchristoph.hartmann@geo.uzh.ch (M.C. Hartmann).

indicators. These relate to themes including biodiversity (e.g. the presence or absence of species (Jeawak et al., 2018)), the added value of protected areas (e.g. by exploring patterns of visitation (Hamstead et al., 2018)) and more generally cultural ecosystem services (e.g. ways in which humans benefit through experiencing nature (Lee et al., 2019; Lee et al., 2022)). Many current efforts to develop ecological indicators using new data forms take the form of case studies, and demonstrate how indicators can be developed using bespoke methods at specific locations and using a single data source (Ghermandi and Sinclair, 2019). There is a pressing need for the development of more generic and scalable workflows allowing creation of indicators of the state of the environment, with a focus on integrating all available data sources related to a study area. Such approaches should consider how different forms of UGC can usefully be combined. One possibility is to use data created by expert contributors in the form of citizen science to train and evaluate models run on much larger, heterogeneous and passively generated social media datasets. Creating indicators using diverse data sources thus also requires that attention be paid to data integration and considering how data quality varies with respect to different platforms.

Particularly promising with respect to environmental indicators are social media data which include both images and some form of related metadata, for example in the form of locations, timestamps and textual descriptions. These data have already been used to explore the environment, for example by identifying the presence of particular species (Toivonen et al., 2019) or human interactions with the environment (Figuerola-Alfaro and Tang, 2017). Methods used in such endeavours may use purely manual annotation, extract information using keywords and simple heuristics or apply machine learning to extract records containing useful content with respect to, for example, individual species or concepts (Edwards et al., 2022). However, most approaches presuppose that the information extracted is not available through other modalities, and few studies have explored the comparative benefits of analysing images and text (You et al., 2016). Where different data sources have been used, they are typically compared rather than combined (Tenkanen et al., 2017), leaving a gap with respect to policy relevant indicators - where the aim is to document, as completely as possible a relevant concept, irrespective of whether one or multiple data sources are used.

In this paper we propose, implement and evaluate a workflow taking advantage of citizen science data documenting and recording sightings of birds, and more specifically red kites (*Milvus milvus*). Analysing social media data until recently has often used simple keyword-based methods to perform an initial filtering or search step, meaning that content tagged in other ways was not found. However, improvements in content-based classification now mean that it is also possible to use off-the-shelf, pre-trained algorithms to reliably identify predefined classes such as presence of buildings, people or birds in image data with reasonable accuracy. Our workflow uses these improvements, together with training data generated from citizen scientists, to identify further relevant sightings in a social media dataset using both textual metadata and image content. Since both datasets include metadata capturing unique users, locations, times and descriptions of sightings, we can describe not only where red kites were seen, but also when, by whom and potentially in what context. In developing our workflow, we aim to assess the performance of individual steps so that we can evaluate the added value of both text and image analysis. Having evaluated the workflow, we explore the properties of the complete integrated dataset, in particular the extent to which different data sources complement one another in describing variation in our target species.

2. Background

Our starting point is twofold. Firstly, the opportunities afforded by new data forms in environmental monitoring have been discussed extensively, not only in research but also by policymakers (Zaccheddu, 2019). Secondly, the need for new data sources has accelerated as

increasing attention is paid to measuring progress towards the Sustainable Development Goals (SDGs). One area where Geographic Information Science can make a concrete contribution is in developing robust and reusable approaches to spatio-temporally explicit indicators recording change of state across a range of scales. MacFeely, 2019 set out opportunities and challenges with respect to the use of 'big data' for compiling SDG indicators. Key opportunities include improved timeliness, a reduction in the costs and organisation related to data collection, potentially finer spatial and temporal granularities and more transferability between countries. Challenges of relevance to our work include a reliance on data produced by commercial actors and the implications of reusing data without the explicit consent of users. Fraisl et al., 2020 in a recent desktop study reported that big data is "already contributing" to the monitoring of five SDG indicators, and that citizen science "could contribute" to 76 indicators. These potential indicators were not evenly spread across all SDGs, with the most potential being suggested for SDGs 11 and 15 (*Sustainable cities and communities* and *Life on land* respectively). However, most examples of the use of such data remain experimental (Van Halderen et al., 2021), relying on bespoke methods and datasets.

Interest in the potential of big data, or more generally new data sources, for indicator production has been fuelled by an explosion of interest in data driven research, made possible by both increased accessibility to a wide variety of novel data sources and increasing ease of use of technology to build workflows incorporating, for example, classification tools trained on very large datasets and running externally in the cloud (Yang et al., 2017).

In Geographic Information Science, Goodchild's seminal 2007 paper (Goodchild, 2007) set out a definition of what he termed Volunteered Geographic Information (VGI). Goodchild described VGI as a special case of what had already been termed User-Generated Content (UGC), where private citizens created geographic information, something which had previously almost exclusively been the role of state. Goodchild's definition was very broad, and encompassed not only projects where citizens actively worked together towards a shared aim (e.g. OpenStreetMap (Haklay and Weber, 2008)), but also those where geographic information was essentially a byproduct and data were shared for personal or social purposes (e.g. Flickr (Spyrou and Mylonas, 2014)). These differences led to many papers exploring the motivations of contributors to VGI (e.g. Larson et al., 2020; Measham and Barnett, 2008), and are important to our work for two reasons. Firstly, since our aim is to create an indicator recording sightings of a specific species using different forms of VGI, we assume that data produced by experts in citizen science projects is of sufficiently high quality that we can use it to train models (Kosmala et al., 2016). Secondly, since our indicator species, the red kite, is highly visible, relatively easy to identify and has cultural significance, we assume that images uploaded to social media platforms include records describing this species.

The approach taken to creating an indicator is highly dependent on the data used and the ways in which they were produced. For example, where citizen scientists have actively recorded sightings of a species, and these sightings overlap with an indicator need (Robbins et al., 2020), then questions with respect to data quality might relate to completeness (e.g. are sightings related to accessibility, and thus biased towards populated areas). If the data produced are found to be broadly comparable to some gold standard, then they can in principle be used directly for indicator production. However, where the data were produced for other purposes, as is the case for example with social media data, additional preprocessing steps are required. These can include querying for records in a particular spatial region and over a given temporal interval, before extracting only relevant records. In our case, where a particular species is of interest, typical approaches analyse data using either simple keyword extraction (e.g. by compiling lists of potentially relevant terms) applied to metadata, followed by some form of disambiguation or classification to remove records using matched terms in another sense. Progress in off-the-shelf image classification approaches

has led to analysis based solely on keywords assigned to images by machine learning classifiers (Burke et al., 2022) especially where the identified classes are relatively broad (e.g. presence of water, green space or trees). Where classifications are more fine-grained, common approaches still involve human annotation of metadata and image content, for example relating activity types to cultural ecosystem services (Pickering et al., 2020). Many current approaches to analysing image based social media focus on analysing either textual metadata or image content, with surprisingly few attempting to combine information derived from both modalities simultaneously, even though existing work suggest potential improvements in performance (You et al., 2016).

All of these advances have led to a plethora of applications of direct relevance to environmental indicators using, mostly, social media data and occasionally also leveraging content produced in citizen science projects. Toivonen et al., 2019 set out in a review the potential value of big data for biodiversity monitoring and conservation science. Many applications are tailored towards understanding of human-nature interactions in space and time, for example by identifying visitor hotspots (Tenerelli et al., 2016). Analysing and understanding the preferences of visitors (Hausmann et al., 2018) can provide important information for policymakers seeking to design and manage attractive and sustainable nature reserves and protected areas. Data traces found on social media can also identify, for example, hot-spots of invasive alien species (Daume, 2016) and illegal trade in wild-life (Di Minin et al., 2019), potentially driving reallocation of resources to improve protection of endangered species and their environments.

A second group of biodiversity applications focus on mapping of species distributions. For instance, Jeawak et al., 2018 show how alternative data sources, such as observations from Flickr, can complement traditional data sources used for species distribution mapping. Mapping of species distribution has also historically taken advantage of citizen science projects, with for example the *Christmas Bird Count* having a history dating back more than 100 years. Digitisation has eased access to data from such projects, and public concern and interest in cataloguing wildlife means that projects such as iNaturalist and eBird have been very successful. La Sorte et al., 2018 reviewed the potential of big data for ornithology, and underlined how new data sources could benefit species distribution modelling, for example by analysing temporal variations such as geographical range shifts or modified migration trajectories due to climate change. Moving beyond citizen science data created by experts, Leighton et al., 2016 investigated the spatial patterns in phenotype traits of the colour morphs of black bears, barn owls and black sparrowhawks as well as the distribution of the hooded and carrion crows using Google Images, one of the largest collections of images found on the internet. As a cost-effective alternative to traditional methods like field surveys, Leighton et al., 2016 found good agreement, suggesting that UGC may be an effective source for environmental indicators. More recent work by Burke et al., 2022[p.1] takes this a step further, by aiming at “estimating the fraction of images within multiple unverified datasets that potentially depict a specified target fauna”. Their proposed workflow relies on general-purpose image classifiers, such as those made available by Google Cloud Vision or Microsoft Azure Computer Vision, to extract descriptive text-tags from an authoritative image dataset. The computed frequencies of the returned tags allow for ‘fingerprinting’ of their species data which can be used to estimate the share of relevant data in other, unverified datasets.

2.1. Research gap and research questions

Despite a wide range of works discussing and demonstrating the potential of new data forms in the creation of indicators, we could not identify previous research which explicitly created a workflow designed to integrate data from different sources and of different modalities. Furthermore, although the properties of different forms of UGC are relatively well understood, they have not been effectively used to develop reproducible workflows. Finally, most studies evaluate the

quality of extracted information in isolation through metrics such as precision and recall, but do not explore the added value of integrating data. These gaps lead to three research questions, which we address in what follows. The first two questions relate to the development of our workflow:

1. How can a generic workflow be developed which leverages expert contributions from citizen science data to extract content from social media posts related to a given species, which can easily be used in practical applications?
2. How can multiple modalities (e.g. text and images) be used to extract relevant information and does such a combination result in higher recall and/or precision?

The third research question relates to the added value of our workflow in integrating data from different sources.

3. What added value with respect to coverage, type, overlap and volume can we identify using a combination of citizen science and social media data?

3. Methods

3.1. Target species and study area

We selected the medium-sized bird of prey red kite (*Milvus milvus*) as our exemplary target species. In the early nineteenth century, red kites were abundant throughout Europe, but they experienced a rapid decline in numbers at the end of the same century, mostly due to persecution (Davies and Davis, 1973; Evans and Pienkowski, 1991). As a consequence, policy regulations and protections were put in place in various countries. In Great Britain, extinction was averted by what is claimed to be the longest continuous conservation project in the world (RSPB, n.d.), and red kites were protected under Annex 1 of the EEC Bird Directive, Schedule 1 of the Wildlife and Countryside Act 1981, and the Convention on the Conservation of Migratory Species of Wild Animals (CMS) (Evans and Pienkowski, 1991). Red kites are highly visible birds, and are an excellent example of a cultural ecosystem service in the form of visible wildlife enjoyed by locals and visitors alike. Importantly for our research, red kites have a number of distinguishing features, in particular a forked tail, making them generally straightforward to identify in images (Fig. 1). Our choice of red kites was motivated by one other curiosity. Their English name is ambiguous, in that the two words ‘red’ and ‘kite’ may refer to either our species of interest, or a red-coloured toy flown in the wind by children. Such ambiguities are an important, though oft-ignored challenge in extracting information from UGC.

To encourage population growth of red kites, reintroduction programs have taken place at sites across Europe. We chose one of these sites, the Chilterns, a 1700 km² Area of Outstanding National Beauty, as a study area for this project (Fig. 3). In the Chilterns, red kite viewing is actively advertised as a tourist attraction (Board, n.d.), and given the area’s proximity to London it is an important location for recreation and enjoyment of nature for a very large potential population.

3.2. Workflow overview

Since our workflow is designed to be generic, take advantage of the text and image data and combine records from citizen science reports with social media data, it uses a combination of a simple rule based-approach, existing pre-trained models and a model trained specifically for our target species (Fig. 2). Our approach is designed to take advantage of what we assume to be high quality data collected by citizen scientists with an interest in ornithology, use off-the-shelf models where possible, and reduce the initial number of social media posts in a given region to a manageable size for manual verification. Our workflow thus:



```
{
  "PhotoID": 4728797008,
  "Latitude": 51.661076,
  "Longitude": -0.906007,
  "Title": "Red Kite 11,
  "Username": "kitmasterbloke,
  "UserID": "58415659@N00,
  "DateTaken": "6/23/2010 15:05:42,
  "UploadDate": "6/23/2010 22:13:45,
  "Views": 340,
  "Tags": "birds;birdsofprey;redkite;stokenchurch;
  "Description": "Red Kites hunting over Stokenchurch;
  Buckinghamshire; just off the M40/A40
  junction [...] At one point;
  I had 8 adults in a single frame. In 30
  minutes; I saw 25-35 birds;
  truly remarkable and one of the great
  wildlife spectacles of Southern England."
}
```

Fig. 1. Exemplary red kite Flickr post from the Chilterns including its meta-data. Credit to Steve Knight, (CC BY 2.0). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

1. Identifies all geotagged social media records in our study area.
2. Extracts all records with the *Latin* name of our target species *Milvus milvus*. We assume these users are experts, and that these records are thus assigned to our final set of relevant records.
3. Uses an existing model to identify images containing birds. We retain only images which are classified here as birds (p_B 50%) are worth further processing.
4. Checks whether or not images classified as birds also have titles, descriptions and tags indicating that they are red kites in six European languages (English: *red kite*, German: *Rotmilan*, *Gabelweih*, *Königsweihe*, French: *milan royal*, Italian: *nibbio reale*, Spanish: *milano real* and Dutch: *rode wouw*). We assume that such images are highly likely to be of our target species, and include these in our final set.
5. For candidate images classified as birds, we run a second classifier trained on citizen science records, to identify those likely to be red kites (p_{RK} 50%). These are also added to our final set of candidate images.
6. Merges all candidate images identified in the social media data with those from our citizen science sources.
7. Manually verifies the final candidate image output at step 6, to identify social media images containing red kites by hand. Our assumption here is that the workflow has greatly reduced the number of candidate images, and that a trained annotator can quickly identify true positives returned by the workflow.
8. Analyses the properties of red kite records from the three integrated datasets. We investigate four dimensions of the merged data:
 - Spatial: all images retrieved in the study area are mapped and patterns of contribution are qualitatively explored

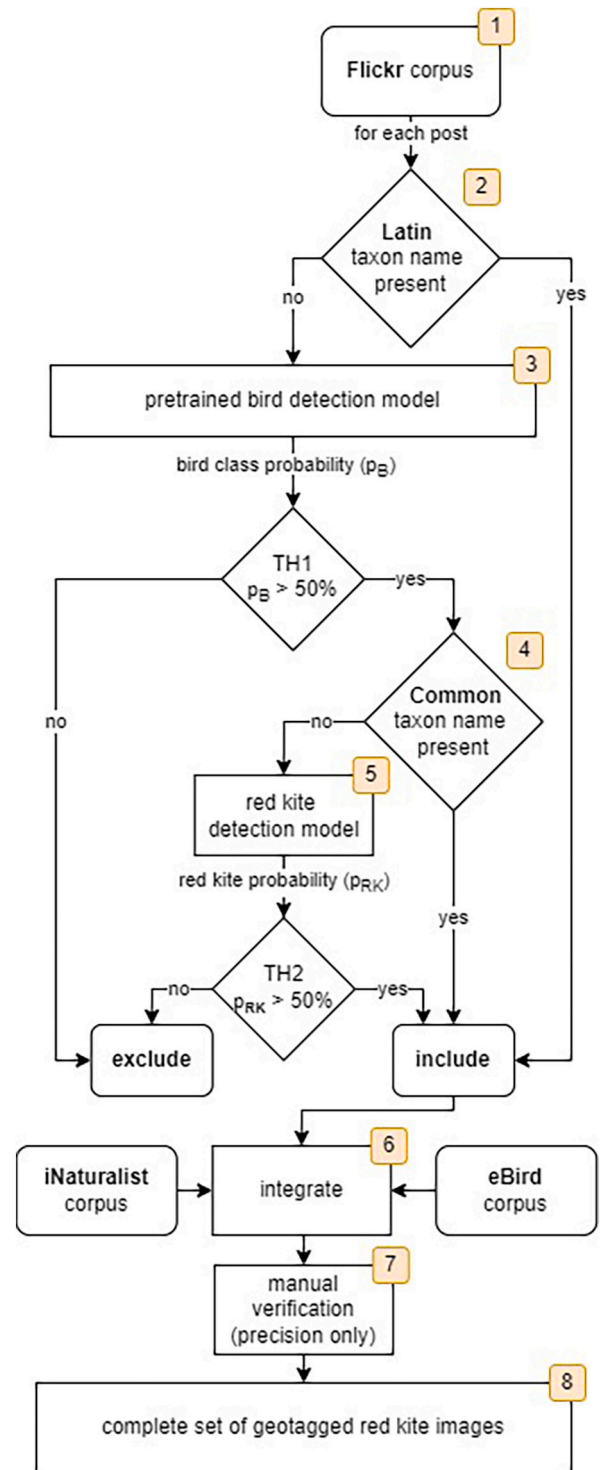


Fig. 2. Initial workflow to extract red kite relevant posts from Flickr to subsequently integrate them with observations in eBird and iNaturalist. TH1 and TH2 specify the bird model and red kite model threshold respectively (section 3.2). However, for the final results presented in this paper, textual and visual data was parsed separately (final workflow), and their results were merged. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

- Temporal: we explore how patterns of citizen science and social media contributions change over time (both in terms of absolute time and seasonal patterns of contribution)

- Users: we investigate how patterns of contribution vary within and between the three merged collections
- Data quality: finally, we explore image data quality by manually labelling all images for (1) image quality, (2) number of red kites per image and (3) whether the red kites are sitting or flying

In the following we describe the data used and image classification steps in more detail.

3.3. Data

Our workflow required three distinct datasets. These were:

1. All Flickr posts (our social media data) in the bounding box of our study area.
2. All citizen scientist records referring to red kites, irrespective of their locations.
3. A set of citizen science data capturing other birds for use in model training.

Flickr is a social media site, where individuals can upload photographs and metadata including tags and locations in the form of coordinates. Flickr usage has declined in recent years, but it remains very popular in research, mostly because of its well documented and easy to use API, which allows querying using search terms and bounding boxes. Our citizen scientist data came from two platforms: iNaturalist and eBird. iNaturalist allows participants to upload images of organisms such as plants and insects to the platform and use its community to crowdsource the taxonomic identification. Currently according to their website (<https://inaturalist.org>), iNaturalist hosts nearly 100 Million observations of over 375,000 species and is therefore one of the largest and most successful citizen science projects to date (Unger et al., 2020). eBird has similar features to iNaturalist but as a platform is exclusively specialised in bird observations. Their website states (<https://ebird.org>) that “eBird is among the world’s largest biodiversity-related science projects, with more than 100 million bird sightings contributed annually”. It predominantly hosts observation location data, but also corresponding bird images as well as bird sounds (Sullivan et al., 2009; Wood et al., 2011).

We downloaded a total of 604,951 geotagged Flickr posts in the bounding box of the Chilterns as depicted in Fig. 3 by querying the Flickr API. Each Flickr post consists of a photograph and additional system- and user-generated information. In our case, this metadata (Fig. 1) includes the physical location and time the image was taken, as well as textual information added by the author, such as a post title, description and tags. This rich textual information together with the images are used by our red kite detection workflow.

To train our red kite classification model (step 5 in Fig. 2) we needed a reliable dataset containing images of birds composed of true positives (images of red kites) and true negatives (images of other bird species). Since we assume that citizen science data are reliable compared to captions from social media data, we created this dataset using eBird and iNaturalist records of red kites. Both eBird and iNaturalist moderate content, for example filtering and checking unusual sightings or volumes of contribution.

To access data from eBird, we obtained the metadata for given search terms from the official website (eBird., 2021) and downloaded all associated “Specimen Page URLs”, which represent individual images. For iNaturalist we used a dedicated API,¹ similar to the one provided by Flickr and a community-driven open source Python library called pyinaturalist (Noé, 2022).

Our true positive dataset thus consisted of citizen science records of red kites sourced from eBird and iNaturalist, manually filtered for poor

examples (e.g. feathers or birds not visible to the human eye). True negatives were eBird records for 11 common European bird species, without any filtering for poor examples. The 11 species were: Black-headed Gull (*Chroicocephalus ridibundus*), Eurasian Blue Tit (*Cyanistes caeruleus*), Eurasian Magpie (*Pica pica*), Eurasian Jay (*Garrulus glandarius*), Eurasian Blackbird (*Turdus merula*), House Sparrow (*Passer domesticus*), Rock Pigeon (*Columba livia*), Mallard (*Anas platyrhynchos*), Eurasian Coot (*Fulica atra*), Barn Swallow (*Hirundo rustica*) and Mute Swan (*Cygnus olor*). A break-down of the data records by data source and type is visible in Table 1.

Through this approach, we created a balanced dataset consisting of 24,675 true positive red kite images, and 31,922 true negatives. These data were then randomly split into data for training, validation (used to select the most performant model variant discussed below) and testing (used to evaluate the performance of the final red kite classification model) as seen in Table 2.

3.4. Image classification models

In our workflow (Fig. 2) two convolutional neural networks (CNNs) were used. The first was the bird object detection model, more specifically a ResNet101 downloaded from the TensorFlow 2 Detection Model Zoo.² We initialised our ResNet101 object detection model, which uses 640×640 pixel input images, with pretrained COCO weights. This implies that the model was already able to detect 80 generic object classes such as ‘truck’, ‘chair’, ‘apple’ and most importantly ‘bird’ with a mean average precision across all classes of 0.318 and an inference time of 55 ms (Martín Abadi et al., 2015). Being able to successfully detect the class ‘bird’ as a filtering step before running our red kite model was central to our procedure.

The second CNN was our red kite image classification model. Red kites do not belong to the generic objects detectable by pretrained, off-the-shelf image classifiers. Thus, we needed to train an image classification model capable of performing this task. For this we applied transfer-learning to a ResNet50 pretrained on ImageNet. We slightly adapted the original architecture using Keras (Chollet et al., 2015) as a high-level API build on top of TensorFlow. We replaced the final 1000-way classification layer of the model with two additional, fully connected layers with ReLU non-linearity activation followed by a 2-way logistic regression classifier, trained via standard cross-entropy loss. These layers add 1 million untrained parameters to the network which need to be tuned to detect red kites on 224×224 pixel input images. The pretrained network was frozen during the transfer-learning process, and only the newly added layers were trained. ResNet50s are commonly used in practice, especially for transfer-learning applications, since they offer a good trade-off between inference time and accuracy. Similarly, de Lutio et al., 2021; Miao et al., 2019; Nguyen et al., 2017 successfully applied ResNet50s to train a plant, and two wildlife detection models respectively.

The two CNNs differ in size and complexity. The ResNet101 object detection model has double the amount of layers (101) but was used off the shelf, whereas the ResNet50 red kite model was trained using transfer-learning on the citizen science data described in Table 2. These differences are important when interpreting the performance of the individual models within the workflow, since more layers are associated with higher precision at the cost of longer inference times and higher computational demand.

The custom red kite model was trained using the Google Colab Pro infrastructure³ which allows Jupyter notebooks to be run in a GPU enabled runtime environment which is an affordable alternative to acquiring the hardware needed for model training. The model was trained with a NVIDIA TESLA T4[®] graphic card for 500 epochs (i.e. 500

¹ <https://inaturalist.org/pages/apireference>, accessed: 14.06.2021

² <https://bit.ly/38cqLYX>

³ <https://colab.research.google.com>, accessed: 17.06.2021

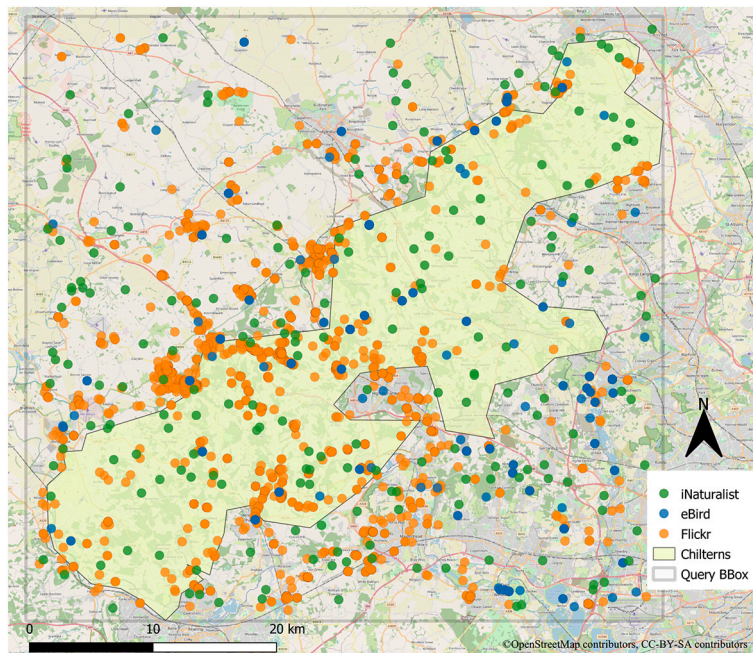


Fig. 3. Locations of extracted red kite observations from the three integrated UGC sources Flickr, eBird and iNaturalist within the Chilterns. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1

Break-down of the data records by their platform and type. NA indicates an unknown quantity of records at the time of acquisition. Red kite images from citizen science data were manually verified before use as training and validation data in the CNN part of the workflow (step 5 in Figure 2).

| Data records | Flickr | eBird | iNaturalist |
|---|---------|--------------------------|-------------|
| red kites (verified) | NA | 7790 | 16,885 |
| 11 other European bird species (not verified) | – | 32,986 (3 k per species) | 0 |
| total | 604,951 | 40,776 | 16,885 |

Table 2

Datasets used to train, validate and test the red kite classifier.

| Class | Training | Validation | Testing | Total |
|---|--------------|------------|-----------|--------|
| true positives (red kites) | 19,929 (81%) | 3796 (15%) | 950 (4%) | 24,675 |
| true negatives (11 other European bird species) | 27,453 (83%) | 4422 (13%) | 1110 (4%) | 32,986 |

passes over the entire training set) and a batch size of 64, with an ADAM optimiser and a learning rate of 10^{-5} . Image augmentation was applied in the form of horizontal flip, 0.2 degree counter clock wise shear and a random zoom between 0 and 0.2 - all leading to 224×244 pixel RGB input tensors. Data was normalised to ImageNet mean values, and the pixels values were rescaled in the range of [0, 1]. Model training took roughly 4 days. The best model was selected based on minimal validation loss that occurred at epoch 448. This model showed a training loss of 0.256, a training accuracy of 0.899, a validation loss of 0.298 and validation accuracy of 0.891. We evaluated the red kite model performance based on an independent test set of 2060 images (as described in 3.3). 950 of these images were true positive red kites images and the

remaining 1110 were true negative red kite images consisting of an even mix of the 11 common European bird species. Table 3 provides a complete overview of the final red kite model performance. We assessed model performance as good, with an F1-score of 0.839.

4. Results and interpretation

The aim of our workflow was to extract relevant images of red kites from Flickr data and to use these to complement citizen science records from eBird and iNaturalist. In the following, we therefore explore three aspects of the results we obtained:

Table 3

Confusion matrix of the red kite model with the resulting performance metrics based on the test set.

| | test set | predicted label | | | |
|---------------------|----------|-----------------|--------------|-----------|-------|
| | | red kite | non red kite | | |
| red kite images | 950 | 754 | 196 | accuracy | 0.860 |
| non red kite images | 1110 | 93 | 1017 | precision | 0.890 |
| | | | | recall | 0.794 |
| | | | | F1 | 0.839 |

- How effective is our workflow at extracting relevant red kite images, and how much added value is obtained through the use of both text and image content (Research questions 1 & 2)
- What are the properties of the extracted records within our study area, and do the social media data complement the citizen science platforms (Research question 3)

4.1. Workflow performance

The workflow (Fig. 2) returned 3065 candidate images, down-sampling the original dataset by 99.5%. These images were then individually inspected to identify true and false positives, and allow us to calculate precision. Images were marked as true positives if a red kite was clearly identifiable in an image. This meant that images had to be sufficiently clear, such that distinctive features of red kites (e.g. their forked tails or red-brown colouring) were visible. Images where a bird was visible, but not unambiguously identifiable, images showing feathers or pellets and images which were obviously irrelevant were all marked as false positives. A total of 2017 records were thus identified as true positives, with 1048 false positives, and a resulting precision for the complete workflow of 0.658.

To understand the benefits of text and image analysis, we ran the components of the workflow individually, and annotated any additional images extracted (Table 4).

1. In the textual workflow setting, records were returned if either the Latin name or a common name for red kite (in six language variations) were detected. This approach identified 2215 posts of which 1946 were true positives, and 269 false positives, resulting in a precision of 0.879.
2. In the visual workflow setting, only visual information was considered. A post was considered relevant and included if both the bird model and red kite model return a probability above 50% for the given image. This approach returned 2763 included posts of which 1723 were true positives and 1040 were false positives, giving a precision of 0.624.

Next, we analysed the added benefit that each of these data streams provide which requires knowledge about the overlap between them. We found 1419 Flickr posts that were included by both settings of which 1407 were true positives. This means that by only retaining candidate records identified by both textual and image-based information we can achieve an almost perfect precision of 0.992. We then checked for records that were exclusively identified by either text or image analysis. 539 posts were only detected by the textual analysis (point 1 in list above) and 316 were only detected by the visual analysis (point 2 in list above). Combining these results leads to a total of 3559 records of which 2262 are true positives and 1297 are false positives, and a precision of 0.636. Looking back at the performance of our initial integrated workflow (Table 4:), we note that 245 (12%) additional true positive red kite posts were extracted by merging the results of separately performed textual and visual analysis. This increase in recall is at the cost of a very slight reduction in precision of 0.02. Summarising these findings, 62% of true positives were found using either text or image analysis. 24% are only correctly classified by textual data and 14% are missed if no visual analysis is performed.

Table 4

Precision using different combinations of the components in the workflow.

| Approach | Included posts | True positives | Precision |
|-------------------------------------|----------------|----------------|-----------|
| Only text data | 2215 | 1946 | 0.878 |
| Only visual data | 2763 | 1723 | 0.624 |
| Initial workflow (Fig. 2) | 3065 | 2017 | 0.658 |
| text + visual data (final workflow) | 3559 | 2262 | 0.636 |

4.2. Data integration

Having identified relevant records from Flickr data, we then merged them with eBird and iNaturalist photographs labelled as red kites, found in our study area, to explore the benefits of integrating these three sources in terms of spatial and temporal coverage and user diversity.

Our merged dataset consists of all geotagged photographs found in the Chilterns depicting identifiable red kites (for Flickr) and all geotagged records reported as being red kites in the Chilterns in eBird and iNaturalist.

The spatial distribution of the three sources is shown in Fig. 3. We record a total of 2732 observations, including 2262 (83%) Flickr, 271 (10%) eBird and 199 (7%) iNaturalist observations inside the query bounding box. Flickr therefore increases the total data volume by a factor ~10 compared to the citizen science datasets. Fig. 3 shows how the underlying three datasets overlap but also complement each other in their spatial coverage. Flickr data is more strongly clustered around urban settlements (e.g. Watlington or Princes Risborough) while eBird and iNaturalist show a more homogeneous spread over the region as a whole. Nevertheless, Flickr does not show a significant increase in data in the south-east around the London suburb Slough, suggesting that red kite data is indeed specifically centred on the Chilterns. We also see that Flickr adds asymmetrically more data in the south-east compared to the north-west, directly affecting the spatial coverage in these areas. Overall, combining the three datasets significantly boosts spatial coverage and data volume.

We then explored the temporal distribution of observations. In a first step, we summarise observations in yearly bins for all three datasets (Fig. 4 a). We observe that Flickr has the highest and most consistent coverage, ranging from the year 2005 (only one observation) to 2021. There is a peak at around 2011 with declining values on both sides. The eBird data ranges from 2004 to 2021, with four gaps, namely 2005, 2007–8 and 2010. A similar pattern is observed for iNaturalist but with a later first recorded observation in 2013. Flickr observations on the other hand have declined considerably since 2011. Between 2007 and 2016, Flickr provided nearly all the available red kite data (with images) for the Chilterns from all datasets combined. The year 2020 marks the only year where another data source besides Flickr provided more contributions - namely iNaturalist. At the end of our time frame in the year 2021 Flickr again provides the majority of data but its share has shrunk to 44% (101) while the share of eBird and iNaturalist rose to 31% (71) and 25% (57) respectively.

In a second temporal analysis we aggregated observations by months to identify seasonal trends of red kite observations in the Chilterns (Fig. 4 b). Flickr shows a clear bias towards spring and summer months with a decline in observations during autumn and winter. iNaturalist shows a similar seasonal trend but the signal for eBird is less distinct.

Using multiple datasets can improve coverage in additional ways, beyond spatial and temporal coverage. The representativeness of user-generated content is strongly dependent on the number of users contributing to the data and what share of the population they represent. Fig. 5 shows the number of unique users per platform with their respective share of observations. We found that in the case of eBird, there are a total of 71 observers (unique users) of whom five contribute 100 out of all 271 eBird observations, with the top contributor adding 30 observations alone. Flickr contributions are more evenly distributed, with 2262 records contributed by 487 unique users. The top contributor shared 88 observations, while 231 contributors each added a single observation. iNaturalist also contains a relatively diverse set of 135 unique users, with the top contributor having added 11 observations. At least in the given region, iNaturalist seems to be able to attract a diverse pool of users, possibly due to a strategy more focused on the platform's approachability via its own mobile app that incorporates gamification elements (Unger et al., 2020).

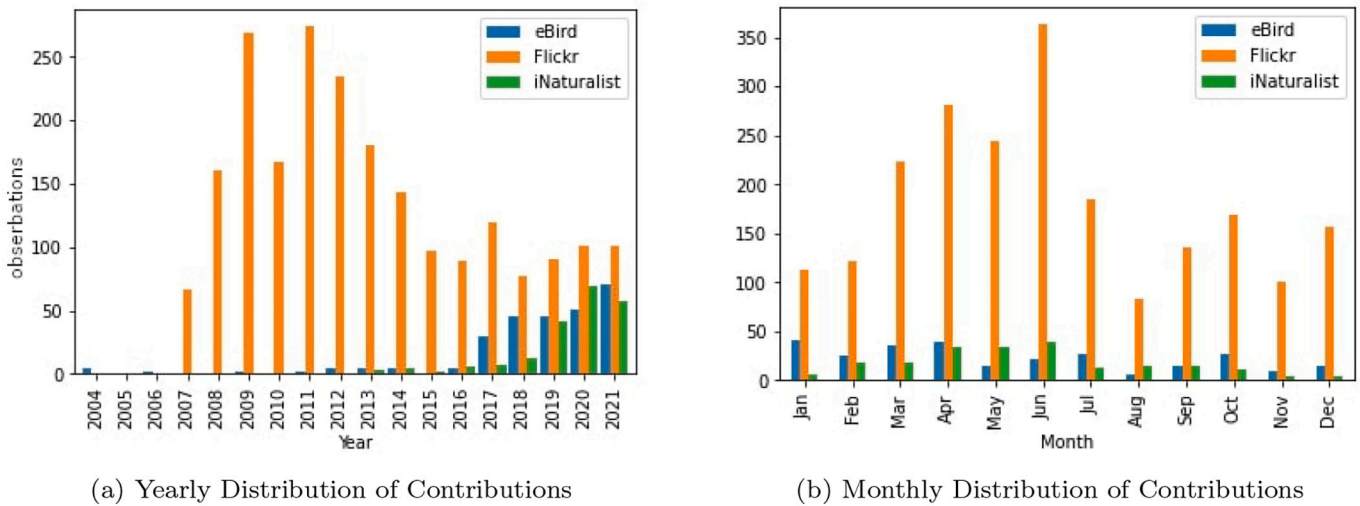


Fig. 4. The figures show the yearly (a) and monthly (b) temporal distribution of red kite observations in the Chilterns. Flickr adds distinctively to the temporal coverage in the years 2007 to 2016 after which eBird and iNaturalist gain in popularity. Strong seasonal patterns can be observed that are skewed towards the spring and summer months. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

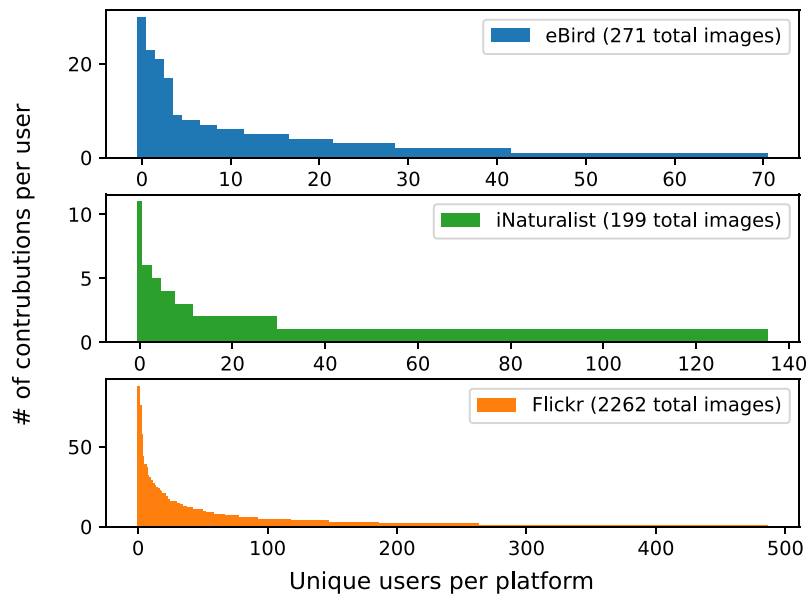


Fig. 5. The histograms visualise the contributions per unique user among the three UGC sources. eBird data is characterised by the largest influence of users with high individual contributions while also having the smallest number of unique users.

4.3. Data quality

As a final part of the data integration analysis we investigated differences in the image properties of the associated three platforms (Fig. 6). We manually labelled all images based on:

1. Image quality quantified by image resolution and how visible red kites were. Flickr images have an average of 0.75 M pixels (MP) and are therefore of significantly higher resolution than images from eBird and iNaturalist with 0.29 MP and 0.19 MP. We considered blurry images or images of red kites in the far distance as low quality. According to our analysis, eBird offers the least low quality images with 16%, followed by 24% for Flickr and 38% for iNaturalist.
2. The number of red kites per image (together with the amount of photo sessions per platform) is an important attribute for inferring population sizes, since normally one bird per image is assumed. We

found that 19% of eBird, 17% of Flickr and 10% of iNaturalist images contained multiple red kites.

3. If a red kite is depicted as sitting or flying can be important for multiple reasons, such as for interpreting the generalisation capabilities of the CNN models, since the shape of the bird differs drastically between modes. Additionally, sitting birds are often harder to spot and unambiguously identify than flying individuals. We hypothesise that more sitting bird images hint at a user base better trained at detecting these birds and who are actively looking for these individuals. We recorded by far the most sitting birds in eBird with 46%, followed by Flickr with 18% and iNaturalist with 11%.

During the annotation process we suspected that multiple images contained the same red kite. We therefore decided to investigate how many images per user and location (within a 500 m radius) we could find - we defined such clusters as *photo sessions*, as an approximation for unique observations. 74% of Flickr, 73% of eBird and 0% of iNaturalist

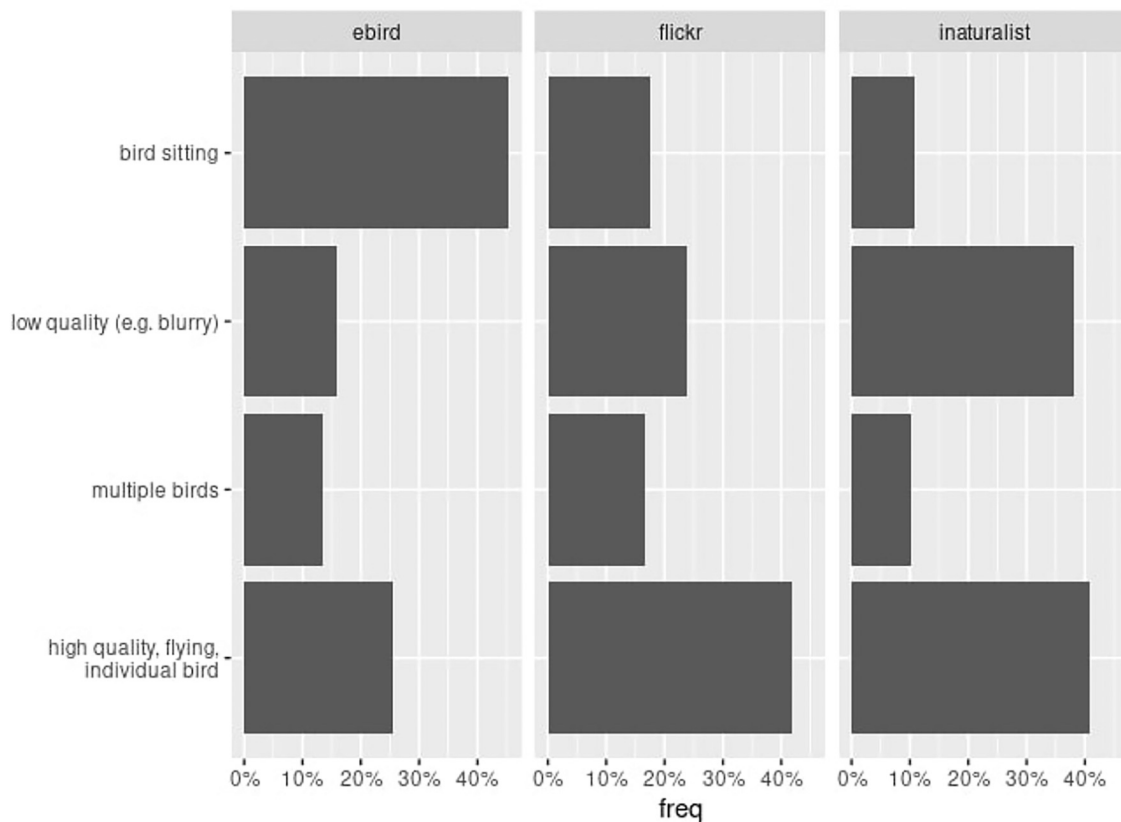


Fig. 6. Analysis of image properties for the three UGC platforms Flickr, eBird and iNaturalist.

images were part of a photo session with two or more images. On average the photo sessions contained 2.29, 2 and 1 image(s) for Flickr, eBird and iNaturalist respectively. This difference is important, since it suggests that in iNaturalist, citizen scientists behave differently, recording only single observations, while in Flickr and eBird multiple images of the same birds are more likely (and thus potentially more contextual information can be explored). Using eBird and Flickr records without filtering for photo sessions would likely result in higher counts than iNaturalist records.

5. Discussion

In this study we developed a workflow which leveraged citizen science data to extract further relevant records from social media posts in the same region. The workflow functions as a data filter enabling downsampling of an initially very large dataset into a human analysable subset - in our case containing 0.5% of the original posts. By massively reducing data volumes, it becomes realistic to analyse the remaining data by hand to select true positives, with around one hour required for the 4000 or so candidate posts we identified. Our workflow thus addresses the research gap identified by [Burke et al., 2022](#), using generalisable methods to extract target data from various, unverified sources to enrich data.

Our initial data extraction workflow ([Fig. 2](#)) assumed that word ambiguity would play a dominant role in identifying relevant posts in the Flickr corpus. We explored the influence of text and image analysis by analysing each component individually. Here, three points are important. Firstly, simple keyword matching delivered high precision with little evidence of ambiguity with respect to the use of red kite in our study area. Image analysis, performed using an off-the-shelf model to identify birds, followed by a red kite model leveraging citizen science data for training, returned more potential candidates than textual analysis, but with lower precision. Secondly, by only retaining posts

identified by both textual and visual analysis, we could achieve almost perfect precision (0.992), at the cost of lower recall. Thirdly, by combining the two approaches, we increased the extracted data volume by almost 14%, while still downsampling the original dataset by around 99.5% and with a precision 0.636. Recall for the workflow overall is not known, since this would require manual analysis of over 600,000 posts. Our results thus go further than previous work by demonstrating how images retrieved using text and image metadata can be combined to achieve very high precision, or merged to increase recall while still filtering initial datasets very effectively. To better understand our results and their transferability to other species, we looked more closely at data quality. Flickr users often used relevant textual descriptions to label their red kite sightings with their captured images. This high textual quality together with the small impact of word ambiguity explains the equally high textual precision. Detecting red kites in images is not as straightforward as looking for specific keywords. Even for experts it can be hard to differentiate between similar raptors e.g. red kites and buzzards. An expert or in our case two sequential CNNs must make a prediction based on a single image whereas the textual description was formulated by the author as a result of the entire red kite observation, potentially hearing and seeing the bird from multiple angles and distances.

Our approach was similar to that taken by a naive human - first we classified images as birds, and only then did we consider whether or not these birds might be red kites. The bird detector was a pretrained, off-the-shelf classifier with well tested performance, and this approach appeared to work well. However, we did not use object classification in our red kite detector, but rather an image classifier, which could only classify entire images and not detect isolated objects. Although object detection models are more performant, they need greater volumes of manually annotated training data. We opted for a faster model creation process, based on categorised photographs from trusted sources (eBird and iNaturalist) to train layers in an image classification model without

further manual annotation, allowing our study to be more transferable to other targets. Since in practice the image properties of these photographs varied (Fig. 6) with eBird having much higher quality images than iNaturalist, future work could benefit from selecting only high quality images for training purposes. Improving the image detection models, e.g. by choosing more performant off-the-shelf architectures as these emerge, or collecting more training data (e.g. including explicitly sitting and flying birds) would be a straightforward way to increase the overall workflow performance.

It is important to highlight a constraint that we laid upon the data. Firstly, we considered only observations with images. Flickr posts, by their nature, almost always include images. eBird and iNaturalist however, are dominated by observations without images – eBird for instance hosts 204,151 world-wide red kite observations of which only 6109 (3%) have images. This restriction was made to ensure consistency of the extracted data across all platforms and to amass a geotagged image corpus. Validating observations based on images was crucial to our analysis.

The visual distribution of points on the map in Fig. 3 shows how the different sources complement one another. The locations of Flickr posts resemble patterns familiar from social media analysis, clustering around urban areas and points of interest along the existing road network, correlating with accessibility (Hausmann et al., 2019). This suggests that the Flickr observations are often taken opportunistically e.g. during a walk without an initial intent to seek out and photograph red kites. eBird and iNaturalist on the other hand show a different pattern. The observations are more heterogeneously allocated and show less obvious relationships to known spatial features. This suggests that birdwatchers go out with a clear intention to observe birds and seek a variety of locations for that purpose.

We analysed the temporal coverage of red kite observations in the Chilterns on a yearly and monthly scale. Aggregating data over years revealed that the pattern shown by Flickr clearly sets itself apart from the one's of eBird and iNaturalist. Year on year changes appear to be more driven by underlying platform dynamics such as user-base and popularity changes (Wu et al., 2016). We suspect that the rapid drop of Flickr observations from 2012 onward represents a decrease in Flickr popularity (Stuart, 2019) rather than a decline in red kites in the Chilterns. On the other hand, we observe a strong increase for eBird and iNaturalist from the year 2016 onward. This could be the result of increased popularity, increased interest in red kites, or increased visits to the study region. Looking at monthly temporal scales shows a trend towards the warmer spring and summer months between March and June. These results may suggest higher visitation rates to the Chilterns in warmer periods but could also be influenced by specific red kite behavioural patterns.

Investigating the number of unique users per data source revealed that representativeness varies between platforms. eBird data was contributed by the fewest individuals whereas Flickr and iNaturalist offered a more diverse user base. This observation could be attributed to higher platform popularity and overall larger user bases of the latter two. Having knowledge of the share of the population represented by an UGC based analysis is crucial for policy makers to make adequate decisions that reflect the people's opinion (Wang et al., 2019).

As a final dimension of our data integration we explored platform specific image properties. Photo sessions may be useful as a proxy for observed individuals, since they suggest multiple images of the same individual and can be considered similar to the very popular notion of photo user days (Keeler et al., 2015). The lack of photo sessions found in iNaturalist may be a result of the generally small corpus or specific behaviour of participants. These observations point to the importance of understanding platform dynamics and focussing on the use of such data as either relative proxies, or additional evidence of presence, for use in species distribution mapping (ElQadi et al., 2017).

The image quality analysis revealed clear differences between social media data in Flickr and citizen science data in eBird and iNaturalist. For

Flickr we observed significantly higher image resolution and quality. We hypothesise that Flickr users are primarily interested in capturing scenic and visually pleasing images whereas eBird and iNaturalist users are more concerned about capturing the target species itself as a proof of observation and less about the image quality. This discovery may point to the potential usefulness of social media data for identification and tracking of individuals (Pace et al., 2019). Visual quality of images in terms of blurriness and visibility of the red kites correlates strongly with whether birds were sitting or flying. iNaturalist which had the largest share of flying red kites also has the most images of low quality. We conclude that the high image resolution and image quality for Flickr indicates that it is a valuable data source not only in terms of quantity (volume) but also in terms of quality.

It is not unexpected that by combining datasets we produce a larger volume of data - but having more data does not always mean there is more information. It may simply be more noise. Or as Boyd and Crawford, 2012 [p.668] put it "increasing the size of the haystack does not make the needle easier to find". It is therefore important to analyse different dimensions - besides volume - that characterise added value and quantify them. Our findings suggest that the datasets complement one another in terms of spatial and temporal coverage - filling data gaps existing in the individual datasets. The increase in the number of unique users suggests that not only the data is complementary, but also that representativeness, in terms of observers, has improved.

6. Conclusion

Our aim in this paper was to develop a generic workflow, leveraging citizen science records to identify complementary data in social media. Our approach can be easily transferred to visible species which are likely to be photographed by citizen scientists and nature and landscape photographers - in other words it can be used as a potential indicator of a cultural ecosystem service in the form of nature appreciation.

We illustrated the use of our workflow in an area known for red kites; a classic indicator species appreciated in many European countries. Our results demonstrate that leveraging citizen science data is an effective approach to increasing data volume and representativeness, potentially filling demographic, spatial and temporal data gaps. We also analysed how well textual and visual data components such as descriptions and images could be used to improve the data integration, extending previous work that used only one of these components (Jeawak et al., 2018; Lopez et al., 2020). The high quality of user provided text from Flickr yielded a higher precision than the visual analysis of images. Nonetheless, analysis of images allowed us to extract 14% more red kite observations from Flickr using a custom red kite image classifier trained on eBird and iNaturalist data that was able to generalise to unseen Flickr images. We found that if we included only Flickr posts identified by textual and image-based information, that with the retained candidate records we can achieve an almost perfect precision of 0.992.

Our workflow does not make manual verification of extracted data obsolete, but rather allows the down-scaling of large initial data volumes to manageable subsets. Ultimately, our work highlights the importance of using multiple data sources to obtain more complete and less biased datasets.

Data availability

The code and data are available at <https://gitlab.vgiscience.de/integration-quality/integration-quality>

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Maximilian C. Hartmann and Ross S. Purves report financial support was provided by Swiss National Science Foundation. Alishiba Dsouza

reports financial support was provided by German Research Foundation. Moritz Schott reports financial support was provided by German Research Foundation. Yannick Metz reports financial support was provided by German Research Foundation.

Acknowledgements

We would like to acknowledge a number of important contributions to our work. Prof. Michael Ying Yang from the University of Twente gave useful inputs in the early stages of the research. Dr. Alexander Dunkel from TU Dresden provided us with access to data for our research area. Prof. Sven Lautenbach from Heidelberg University gave helpful and constructive feedback on a draft of the manuscript. Steve Knight gave us permission to use his image in the paper (Fig. 1). Of course our work would not have been possible without the citizen scientists who contributed data to eBird and iNaturalist, and those who contributed images of red kites to Flickr.

Maximilian C. Hartmann and Ross S. Purves' work on this project was supported by the Swiss National Science Foundation project EVA-VGI 2 [grant number 186389]; Alishiba Dsouza, Moritz Schott, Yannick Metz's work was funded by the German Research Foundation projects WorldKG, IDEAL-VGI and BirdTrace respectively [grant numbers 424985896, 424966858, 314671965].

References

- Abadi, Martín, Agarwal, Ashish, Barham, Paul, Brevdo, Eugene, Chen, Zhifeng, Citro, Craig, Corrado, Greg S., Davis, Andy, Dean, Jeffrey, Devin, Matthieu, Ghemawat, Sanjay, Goodfellow, Ian, Harp, Andrew, Irving, Geoffrey, Isard, Michael, Jia, Y., Jozefowicz, Rafal, Kaiser, Lukasz, Kudlur, Manjunath, Levenberg, Josh, 2015. TensorFlow: Large-scale Machine Learning on Heterogeneous Systems. Software available from. tensorow.org/ github.com/tensorow/models/blob/master/research/object/detection/g3doc/tf2_detection_zoo.md.
- Board, C. C. (n.d.). The Chilterns Area of Outstanding Natural Beauty - Red Kites. <https://www.chilternsaonb.org/about-chilterns/red-kites.html>.
- Boyd, D., Crawford, K., 2012. Critical questions for big data. *Inf. Commun. Soc.* 15 (5), 662–679. <https://doi.org/10.1080/1369118X.2012.678878>.
- Burke, H.M., Tingley, R., Dorin, A., 2022. Tag frequency difference: rapid estimation of image set relevance for species occurrence data using general-purpose image classifiers. *Ecol. Inform.* 69, 101598 <https://doi.org/10.1016/j.ecoinf.2022.101598>.
- Chollet, F., et al., 2015. Keras. <https://keras.io>.
- Daume, S., 2016. Mining twitter to monitor invasive alien species — an analytical framework and sample information topologies. *Ecol. Inform.* 31, 70–82. <https://doi.org/10.1016/j.ecoinf.2015.11.014>.
- Davies, P.W., Davis, P., 1973. The ecology and conservation of the red kite in wales. *Br. Birds* 66 (183–224), 241–269.
- Di Minin, E., Fink, C., Hiippala, T., Tenkanen, H., 2019. A framework for investigating illegal wildlife trade on social media with machine learning. *Conserv. Biol.* 33 (1), 210–213. <https://doi.org/10.1111/cobi.13104>.
- eBird, 2021. eBird: An Online Database of Bird Distribution and Abundance. <https://media.ebird.org/catalog?taxonCode=redkit1%5C&mediaType=photo>.
- Edwards, T., Jones, C.B., Corcoran, P., 2022. Identifying wildlife observations on twitter. *Ecol. Inform.* 67, 101500 <https://doi.org/10.1016/j.ecoinf.2021.101500>.
- ElQadi, M.M., Dorin, A., Dyer, A., Burd, 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. <https://doi.org/10.1016/j.ecoinf.2017.02.006>.
- Evans, L.M., Pienkowski, M., 1991. World status of the red kite. *Br. Birds* 84, 171–187.
- Figueroa-Alfaro, R.W., Tang, Z., 2017. Evaluating the aesthetic value of cultural ecosystem services by mapping geo-tagged photographs from social media data on panoramio and iocr. *J. Environ. Plan. Manag.* 60 (2), 266–281. <https://doi.org/10.1080/09640568.2016.1151772>.
- Fraisil, D., Campbell, J., See, L., Wehn, U., Wardlaw, J., Gold, M., Moorthy, I., Arias, R., Piera, J., Oliver, J.L., Masó, J., Penker, M., Fritz, S., 2020. Mapping citizen science contributions to the un sustainable development goals. *Sustain. Sci.* 15 (6), 1735–1751. <https://doi.org/10.1007/s11625-020-00833-7>.
- Ghermandi, A., Sinclair, M., 2019. Passive crowdsourcing of social media in environmental research: a systematic map. *Glob. Environ. Chang.* 55, 36–47. <https://doi.org/10.1016/j.gloenvcha.2019.02.003>.
- Goodchild, M.F., 2007. Citizens as sensors: the world of volunteered geography. *GeoJournal* 69, 211–221. <https://doi.org/10.1007/s10708-007-9111-y>.
- Haklay, M., 2013. Citizen science and volunteered geographic information: Overview and typology of participation. In: Sui, D., Elwood, S., Goodchild, M. (Eds.), *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (Vgi) in Theory and Practice* (Pp. 105–122). Springer Netherlands. https://doi.org/10.1007/978-94-007-4587-2_7.
- Haklay, M., Weber, P., 2008. Openstreetmap: User-generated street maps. *IEEE Pervasive Comp.* 7 (4).
- Hamstead, Z.A., Fisher, D., Ilieva, R.T., Wood, S.A., McPhearson, T., Kremer, P., 2018. Geotagged social media as a rapid indicator of park visitation and equitable park access. *Comput. Environ. Urban. Syst.* 72, 38–50. <https://doi.org/10.1016/j.compenurbysys.2018.01.007>.
- Hausmann, A., Toivonen, T., Slotow, R., Tenkanen, H., Moilanen, A., Heikinheimo, V., Di Minin, E., 2018. Social media data can be used to understand tourists' preferences for nature-based experiences in protected areas. *Conserv. Lett.* 11 (1), e12343 <https://doi.org/10.1111/conl.12343>.
- Hausmann, A., Toivonen, T., Fink, C., Heikinheimo, V., Tenkanen, H., Butchart, S.H., Brooks, T.M., Di Minin, E., 2019. Assessing global popularity and threats to important bird and biodiversity areas using social media data. *Sci. Total Environ.* 683, 617–623. <https://doi.org/10.1016/j.scitotenv.2019.05.268>.
- Jeawak, S.S., Jones, C.B., Schockaert, S., 2018. Mapping wildlife species distribution with social media: augmenting text classification with species names. *Leibniz Intern. Proceed. Inform LIPICs* 114 (45), 1–6. <https://doi.org/10.4230/LIPICs.GIScience.2018.34>.
- Keeler, B.L., Wood, S.A., Polasky, S., Kling, C., Filstrup, C.T., Downing, J.A., 2015. Recreational demand for clean water: evidence from geotagged photographs by visitors to lakes. *Front. Ecol. Environ.* 13 (2), 76–81. <https://doi.org/10.1890/140124>.
- Kosmala, M., Wiggins, A., Swanson, A., Simmons, B., 2016. Assessing data quality in citizen science. *Front. Ecol. Environ.* 14 (10), 551–560. <https://doi.org/10.1002/fee.1436>.
- La Sorte, F.A., Lepczyk, C.A., Burnett, J.L., Hurlbert, A.H., Tingley, M.W., Zuckerberg, B., 2018. Opportunities and challenges for big data ornithology. *Condor* 120 (2), 414–426. <https://doi.org/10.1650/CONDOR-17-206.1>.
- Larson, L.R., Cooper, C.B., Futch, S., Singh, D., Shipley, N.J., Dale, K., LeBaron, G.S., Takekawa, J.Y., 2020. The diverse motivations of citizen scientists: does conservation emphasis grow as volunteer participation progresses? *Biol. Conserv.* 242, 108428 <https://doi.org/10.1016/j.biocon.2020.108428>.
- Lee, H., Seo, B., Koellner, T., Lautenbach, S., 2019. Mapping cultural ecosystem services 2.0 – potential and shortcomings from unlabeled crowd sourced images. *Ecol. Indic.* 96, 505–515. <https://doi.org/10.1016/j.ecolind.2018.08.035>.
- Lee, H., Seo, B., Cord, A.F., Volk, M., Lautenbach, S., 2022. Using crowdsourced images to study selected cultural ecosystem services and their relationships with species richness and carbon sequestration. *Ecosyst. Serv.* 54, 101411 <https://doi.org/10.1016/j.ecoser.2022.101411>.
- Leighton, G.R.M., Hugo, P.S., Roulin, A., Amar, A., 2016. Just Google it: assessing the use of Google images to describe geographical variation in visible traits of organisms. *Methods Ecol. Evol.* 7, 1060–1070. <https://doi.org/10.1111/2041-210X.12562>.
- Lopez, B., Minor, E., Crooks, A., 2020. Insights into human-wildlife interactions in cities from bird sightings recorded online. *Landscape Urban Plan.* 196 (May 2019), 103742. <https://doi.org/10.1016/j.landurbplan.2019.103742>.
- de Lutio, R., She, Y., D'Aronco, S., Russo, S., Brun, P., Wegner, J.D., Schindler, K., 2021. Digital Taxonomist: Identifying Plant Species in Citizen Scientists' Photographs. <https://arxiv.org/abs/2106.03774>.
- MacFeeley, S., 2019. The big (data) bang: opportunities and challenges for compiling sdg indicators. *Global Policy* 10 (S1), 121–133. <https://doi.org/10.1111/1758-5899.12595>.
- Masó, J., Serral, I., Domingo-Marimon, C., Zabala, A., 2020. Earth observations for sustainable development goals monitoring based on essential variables and driver-pressure-state-impactresponse indicators. *Intern. J. Digital Earth* 13 (2), 217–235. <https://doi.org/10.1080/17538947.2019.1576787>.
- Measham, T.G., Barnett, G.B., 2008. Environmental volunteering: motivations, modes and outcomes. *Aust. Geogr.* 39 (4), 537–552. <https://doi.org/10.1080/00049180802419237>.
- Miao, Z., Gaynor, K.M., Wang, J., Liu, Z., Muellerklein, O., Norouzzadeh, M.S., McInturff, A., Bowie, R.C., Nathan, R., Yu, S.X., Getz, W.M., 2019. Insights and approaches using deep learning to classify wildlife. *Sci. Rep.* 9 (1), 1–9. <https://doi.org/10.1038/s41598-019-44565-w>.
- Nguyen, H., Maclagan, S.J., Nguyen, T.D., Nguyen, T., Flemons, P., Andrews, K., Ritchie, E.G., Phung, D., 2017. Animal recognition and identification with deep convolutional neural networks for automated wildlife monitoring. In: *Proceedings - 2017 International Conference on Data Science and Advanced Analytics, DSAA 2017, 2018-Janua* (Figure 1), pp. 40–49. <https://doi.org/10.1109/DSAA.2017.31>.
- Noé, N., 2022. Pynaturalist Github. <https://github.com/niconoe/pynaturalist>.
- Pace, D.S., Giacomini, G., Campana, I., Paraboschi, M., Pellegrino, G., Silvestri, M., Alessi, J., Angeletti, D., Cafaro, V., Pavan, G., Ardzzone, G., Arcangeli, A., 2019. An integrated approach for cetacean knowledge and conservation in the central mediterranean sea using research and social media data sources. *Aquat. Conserv. Mar. Freshwat. Ecosyst.* 29 (8), 1302–1323. <https://doi.org/10.1002/aqc.3117>.
- Pickering, C., Walden-Schreiner, C., Barros, A., Rossi, S.D., 2020. Using social media images and text to examine how tourists view and value the highest mountain in Australia. *J. Outdoor Recreat. Tour.* 29, 100252 <https://doi.org/10.1016/j.jort.2019.100252>.
- Robbins, J.R., Babey, L., Embling, C.B., 2020. Citizen science in the marine environment: estimating common dolphin densities in the north-East Atlantic. *PeerJ* 8, e8335. <https://doi.org/10.7717/peerj.8335>.
- RSPB. (n.d.). Red Kite Conservation. <https://rspb.org.uk/birds-and-wildlife/wildlife-guides/bird-az/red-kite/conservation>.
- Spyrou, E., Mylonas, P., 2014. An overview of iocr challenges and research opportunities. In: *2014 9th International Workshop on Semantic and Social Media Adaptation and Personalization*, pp. 88–93. <https://doi.org/10.1109/SMAP.2014.19>.

- Stuart, E., 2019. Flickr: organizing and tagging images online. *KO Knowl. Organ.* 46 (3), 223–235. <https://doi.org/10.5771/0943-7444-2019-3-223>.
- Sullivan, B.L., Wood, C.L., Iliff, M.J., Bonney, R.E., Fink, D., Kelling, S., 2009. eBird: a citizen-based bird observation network in the biological sciences. *Biol. Conserv.* 142 (10), 2282–2292. <https://doi.org/10.1016/j.biocon.2009.05.006>.
- Tenerelli, P., Demšar, U., Luque, S., 2016. Crowdsourcing indicators for cultural ecosystem services: a geographically weighted approach for mountain landscapes. *Ecol. Indic.* 64, 237–248. <https://doi.org/10.1016/j.ecolind.2015.12.042>.
- Tenkanen, H., Di Minin, E., Heikinheimo, V., Hausmann, A., Herbst, M., Kajala, L., Toivonen, T., 2017. Instagram, iCr, or twitter: assessing the usability of social media data for visitor monitoring in protected areas. *Sci. Rep.* 7 (1), 1–11. <https://doi.org/10.1038/s41598-017-18007-4>.
- Toivonen, T., Heikinheimo, V., Fink, C., Hausmann, A., Hiippala, T., Järvi, O., Tenkanen, H., Di Minin, E., 2019. Social media data for conservation science: a methodological overview. *Biol. Conserv.* 233, 298–315. <https://doi.org/10.1016/j.biocon.2019.01.023>.
- Trojan, J., Schade, S., Lemmens, R., Frantál, B., 2019. Citizen science as a new approach in geography and beyond: review and reflections. *Moravian Geogr. Reports* 27 (4), 254–264.
- UNEP, 2019. Measuring Progress, Towards Achieving the Environmental Dimension of the SDGs. <https://wedocs.unep.org/bitstream/handle/20.500.11822/30672/SDG/brief/008.Pdf>.
- Unger, S., Rollins, M., Tietz, A., Dumais, H., 2020. Inaturalist as an engaging tool for identifying organisms in outdoor activities. *J. Biol. Educ.* 1–11. <https://doi.org/10.1080/00219266.2020.1739114>.
- Van Halderen, G., Bernal, I., Sejersen, T., Jansen, R., Ploug, N., Truszczynski, M., 2021. Big Data for the SDGS : Country Examples in Compiling SDG Indicators using Non-traditional Data Sources. <https://hdl.handle.net/20.500.12870/3442>.
- Wang, Zijian, Hale, Scott, Adelani, David Ifeoluwa, Grabowicz, Przemyslaw, Hartman, Timo, Fabin, Flöck, Jurgens, David, 2019. Demographic Inference and Representative Population Estimates from Multilingual Social Media Data. Association for Computing Machinery, New York, NY, USA, pp. 2056–2067.
- Wood, C., Sullivan, B., Iliff, M., Fink, D., Kelling, S., 2011. eBird: engaging birders in science and conservation. *PLoS Biol.* 9 (12) <https://doi.org/10.1371/journal.pbio.1001220>.
- Wu, B., Cheng, W.-H., Zhang, Y., Mei, T., 2016. Time matters: Multi-scale temporalization of social media popularity. In: Proceedings of the 24th ACM International Conference on Multimedia, pp. 1336–1344. <https://doi.org/10.1145/2964284.2964335>.
- Yang, C., Yu, M., Hu, F., Jiang, Y., Li, Y., 2017. Utilizing cloud computing to address big geospatial data challenges [geospatial cloud computing and big data]. *Comput. Environ. Urban. Syst.* 61, 120–128. <https://doi.org/10.1016/j.compenvurbsys.2016.10.010>.
- You, Q., Cao, L., Jin, H., Luo, J., 2016. Robust Visual-Textual Sentiment Analysis: When Attention Meets Tree-Structured Recursive Neural Networks, pp. 1008–1017. <https://doi.org/10.1145/2964284.2964288>.
- Zaccheddu, G., 2019. The Territorial Dimension in Sdg Indicators: Geospatial Data Analysis and its Integration with Statistical Data. https://un-ggim-europe.org/wp-content/uploads/2019/05/UN/GGIM_08_05_2019-The-territorial-dimension-in-SDG-indicators-Final.pdf.