How the Citizen Science Platform iSpot Ensures Data Accuracy During and After Collection

Michael Dodd ‡

‡ Open University, Milton Keynes, United Kingdom

Corresponding author: Michael Dodd (michael.dodd@open.ac.uk)

Abstract

The iSpot citizen science platform has been collecting biodiversity data since 2009 and includes ~900,000 observations, about half of these are in the British Isles. Our system to ensuring data accuracy, especially of species identification, uses metadata uploaded with photographs for time and sometimes location, a reputation system, an active user community and curation by an experienced ecologist.

Taxon identification/resolution – On iSpot, anyone can enter an observation and anyone can enter an identification (ID) but the ID that becomes 'likely' depends on the reputation of the user who provided it, combined with the reputation of other users who agree with it. Users gain reputation by entering IDs that other users with existing reputations agree with and which then become the likely ID. The accuracy of the system has been checked by passing a sample of observations with likely ID to the United Kingdom (UK) national system of verification, where experts in all groups of organisms check the IDs and other aspects of observation records. The observations in Table 1 were verified by national or regional taxonomic experts on the irecord system. For some taxonomic groups with organisms that are relatively easy to identify by photographs, such as plants, animals and some types of invertebrates, the proportion accepted is higher than the average. Whereas for other groups, such as fungi, the proportion is lower. Observations are usually rejected because the expert thinks it is not possible to provide an accurate ID better than to a genus or family level. However in other cases, the expert has simply not looked at the images provided on the iSpot system, even though they are given the link to the images. For example, recently an expert rejected an observation because they said that the species does not occur in that part of the country, but they had not looked at the actual observation with its images. The images clearly showed that the correct species does occur in all the surrounding areas, so it is quite likely that the experts themselves made a mistake in rejecting the record. Experts are now under huge pressure to validate or otherwise reject IDs, given the very large increase in observations coming in.

Observer expertise – In some ways this aspect is less relevant since the ID is often provided by others but still the observer has to give accurate location, time and provide

good images. Feedback is given on these aspects by the community, especially to new users of the system.

Clarity and resolution of digital recordings – There is advice on taking suitable images and the community often provides comments and hints for improvement. This is especially an issue with fungi, which often require an image of the underside of the fruiting body as well as overall shots. Some members of the community provide example observations with 10 or more images of the specimen to illustrate all relevant aspects.

Spatial accuracy – From one point of view, we want to leave the observations that are wrongly located to get a measure of the overall accuracy of the dataset. However the community complains if observations are obviously wrongly positioned and not corrected and it is not good to pass on wrongly located data to other organisations. There are articles and forum topics asking users to check their observation locations and suggest how to do this; comments are posted on the wrongly located observations asking for them to be corrected. As a last resort, if the user does not provide a corrected location, then the curator will take this responsibility, moving ~0.2% of observations. Errors range from simply missing a minus sign on the coordinate to making mistakes with mouse or pointer. Some of these can be easily corrected; others require more detective work or are impossible to correct and may be deleted from the system.

To assess the accuracy of locations, the first 1007 observations were selected from a buffer of 3 miles off the United Kingdom coastline (currently there are approximately 4500 observations in this buffer). This area was chosen as it is an area where errors may be easier to spot and where the observations were from all around the coast by many different users. The observations were examined individually, looking at the location name provided and checking if they were within ~2 miles of the coastline (880), greater than 2 miles away (37), or the location name was too ambiguous to tell (90).

The observations classed as greater than 2 miles from where they should be were mapped to see if there were any common issues (Fig. 1). The red dots are from the observation coordinates and the corresponding green dot from the place name provided. In some cases there are multiple observations at the same place. The most obvious issue is the cluster of observations wrongly mapped to south east Scotland, a known problem due to information from image files being wrongly read by the system. Images from a wide range of different cameras are submitted to the system, each with slight differences in how they contain date/time and location information, even though they are all .jpg files. The other long distance errors were from just 4 users, who may have used their home location by mistake or did not zoom in on the map to manually enter the correct location.

Observations from irecord have rarely if ever shown any query over location in terms of name of the location being different from the coordinates.

Temporal accuracy – The system automatically records many aspects of the observation including when it was uploaded but date of upload is often not the date of observation. So the date of observation is specifically asked for. There has not been a lot of checking of

dates but via phenology and other aspects of the image or date of species appearance it is very rare for the date of observation to be wrongly recorded.

Community involvement in curation – The community is involved in giving identifications and agreements, and in writing comments on observations and in the forum. They also set up and run projects on particular localities or taxonomic groups, and looking back through all the existing observations in these areas or taxa and checking them all. It is possible for the community to achieve the correct identification but for other aspects such as wrong locations, they try to ask the original observer and if they don't respond then ask the curator to move the observation to the correct place, if that is possible to deduce. It is important for the curator and ideally programmers to be involved with the community so this is a two way-process.

Keywords

biodiversity, online community, social media, reputation system, social learning

Presenting author

Michael Dodd

Presented at

TDWG 2022

Conflicts of interest

Figure 1.

Red dots are from the observation coordinates and the corresponding green dot from the placename provided. In some cases there are multiple observations at the same place.

Table 1.

The number of iSpot observations verified by national or regional taxonomic experts on the irecord system in all taxonomic groups.

Row Labels	Total
Accepted	20495
Queried	109
Rejected	975