OPEN ACCESS

Conference Abstract

Utilising the Crowd to Unlock the Data on Herbarium Specimens at the Royal Botanic Garden Edinburgh

Sally King[‡], Juliette Pinon[§], Robyn Drinkwater[‡]

‡ Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom
§ Muséum national d'Histoire naturelle, Paris, France

Abstract

Digitisation of specimens at the Royal Botanic Garden Edinburgh (RBGE) has created nearly half a million imaged specimens. With data entry from the specimen labels on herbarium sheets identified as the rate-limiting step in the digitisation workflow, the majority of specimens are databased with minimal data (filing name and geographical region), leaving a need to add further label data (collector, collecting locality, collection date etc.) to make the specimens research ready. We are exploring a number of different ways to complete data entry for specimens that have been imaged. These have included Optical Character Recognition (OCR), to identify meaningful specimen groupings to increase the speed of data entry and more recently citizen science platforms to provide accurate crowd-sourced transcriptions of specimen label data. We sent specimen images of the Australian flowering plants held at RBGE herbarium to DigiVol (https://volunteer.ala.org.au/institution/index/21309224), the citizen science platform developed alongside The Atlas of Living Australia. In 29 expeditions, 156 citizen scientists completed collection label data entry for RBGE's 41,000 specimens of Australian flowering plants.

We found that 95% of the transcriptions were completed by less than a third (27%) of the volunteers. Of the four volunteer experience levels in DigiVol we found that the middle two,

Collection Managers and Scientists, transcribed fewer specimens, but also made fewer mistakes. We found that by removing the filing name from the information provided with the expedition the number of errors in the Museum Details section of the transcription decreased, as the filing name was often added as the label name, regardless of whether this is the case. The feedback we provided for each expedition was used to highlight common errors to try and reduce their occurrence as well as to inform the volunteers of what their transcriptions had revealed about this part of the collection. We explore the citizen science transcription workflow, its rate-limiting steps and how we have worked to include the citizen science and OCR data on our online herbarium catalogue.

Keywords

crowdsourcing, OCR, Australia, citizen science

Presenting author

Sally King & Robyn Drinkwater

Presented at

Biodiversity_Next 2019