

# Improving Biological Collections Data through Human-AI Collaboration

Alan Stenhouse<sup>‡</sup>, Nicole Fisher<sup>‡</sup>, Brendan Lepschi<sup>§</sup>, Alexander Schmidt-Lebuhn<sup>‡</sup>, Juanita Rodriguez<sup>‡</sup>, Federica Turco<sup>‡</sup>, Emma Toms<sup>‡</sup>, Andrew Reeson<sup>l</sup>, Cécile Paris<sup>¶</sup>, Pete Thrall<sup>‡</sup>

<sup>‡</sup> CSIRO, Canberra, Australia

<sup>§</sup> Australian National Herbarium, Centre for Australian National Biodiversity Research, Canberra, Australia

<sup>l</sup> Data61, CSIRO, Canberra, Australia

<sup>¶</sup> Data61, CSIRO, Sydney, Australia

Corresponding author: Alan Stenhouse ([alan.stenhouse@csiro.au](mailto:alan.stenhouse@csiro.au))

## Abstract

Biological collections play a crucial role in our understanding of biodiversity and inform research in areas such as biosecurity, conservation, human health and climate change. In recent years, the digitisation of biological specimen collections has emerged as a vital mechanism for preserving and facilitating access to these invaluable scientific datasets. However, the growing volume of specimens and associated data presents significant challenges for curation and data management. By leveraging human-Artificial Intelligence (AI) collaborations, we aim to transform the way biological collections are curated and managed, unlocking their full potential in addressing global challenges.

We present our initial contribution to this field through the development of a software prototype to improve metadata extraction from digital specimen images in biological collections. The prototype provides an easy-to-use platform for collaborating with web-based AI services, such as [Google Vision](#) and [OpenAI](#) Generative Pre-trained Transformer (GPT) Large Language Models (LLM). We demonstrate its effectiveness when applied to herbarium and insect specimen images. Machine-human collaboration may occur at various points within the workflows and can significantly affect outcomes. Initial trials suggest that the visual display of AI model uncertainty could be useful during expert data curation. While much work remains to be done, our results indicate that collaboration between humans and AI models can significantly improve the digitisation rate of biological specimens and thereby enable faster global access to this vital data.

Finally, we introduce our broader vision for improving biological collection curation and management using human-AI collaborative methods. We explore the rationale behind this approach and the potential benefits of adding AI-based assistants to collection teams. We also examine future possibilities and the concept of creating 'digital colleagues' for seamless collaboration between human and digital curators. This 'collaborative intelligence' will enable us to make better use of both human and machine capabilities to

achieve the goal of unlocking and improving our use of these vital biodiversity data to tackle real-world problems.

## **Keywords**

curation, digital curator, named entity recognition, OCR

## **Presenting author**

Alan Stenhouse

## **Presented at**

TDWG 2023

## **Conflicts of interest**

The authors have declared that no competing interests exist.