Biodiversity Information Science and Standards 10: e184058
doi: 10.3897/biss.10.184058

Conference Abstract

AI-Assisted Quality Checks for Herbarium Sheet Digitisation

Naifeng Zhang‡, Arianna Salili-James‡, Sanson T.S. Poon‡, Ben Hartley‡, Robyn Drinkwater§, Milo Phillips§, Rob Cubey§, Elspeth Margaret Haston§

‡ Natural History Museum, London, United Kingdom
§ Royal Botanic Garden Edinburgh, Edinburgh, United Kingdom

Corresponding author: Arianna Salili-James (arianna.salili-james@nhm.ac.uk)

Received: 02 Jan 2026 | Published: 14 Jan 2026

Citation: Zhang N, Salili-James A, Poon STS, Hartley B, Drinkwater R, Phillips M, Cubey R, Haston E (2026) AI-Assisted Quality Checks for Herbarium Sheet Digitisation. Biodiversity Information Science and Standards 10: e184058. https://doi.org/10.3897/biss.10.184058

Abstract

It is estimated that there are nearly 400 million herbarium specimens held across approximately 3,500 herbaria worldwide (Davis 2023). Over the past decade, many institutions have embarked on large-scale digitisation initiatives to robustly and efficiently digitise herbarium sheet collections. While these processes can be highly productive, reaching daily rates of 600–1,000 herbarium sheets in some programmes, there remains a risk and cost associated with re-digitisation, often caused by complications during the imaging stage. To address these challenges, the Natural History Museum, London (NHM) partnered with the Royal Botanic Garden Edinburgh (RBGE) to develop an Artificial Intelligence (AI)-based workflow aimed at providing enhanced quality control of herbarium sheet digitisation. This initiative was supported by the UK Government's Department for Culture, Media and Sport (DCMS) through the NHM–DCMS AI Pilot Programme (Poon et al. 2026).

The aim of this project was to develop an AI-assisted quality control framework (Zhang et al. 2025). Trained on collections at RBGE, the tool enables automated quality checks of digitised herbarium sheets. This is done by detecting missing objects (such as colour reference cards or barcodes), identifying cropping errors, and highlighting focus issues (Fig. 1). The pipeline integrates several computational techniques, including object detection, decision trees, classical statistics, and deep learning methods. In addition to

© Zhang N et al. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

post-digitisation analysis, a user-friendly interface was also developed to support *in situ* quality checks during the imaging stages.

Figure 1.

Examples of standard (a) and flawed images (b to f), indicating five frequent issues we selected in this project: b) abnormal bars; c) focus errors; d) missing elements, such as missing barcodes; e) under-cropping; and f) over-cropping (Zhang et al. 2025). © 2025 Naifeng Zhang. Licensed under CC BY-4.0.

Model evaluation on an independent test set demonstrated the high accuracy and robustness of the AI-assisted framework (Zhang et al. 2025). The object detection model achieved an overall precision of 0.98, recall of 0.98, and mean Average Precision scores of 0.99 (mAP50) and 0.94 (mAP50–95) across all object categories. Key reference elements, including colour reference cards, barcodes, and rulers, were detected with near-perfect accuracy, ensuring precise object localisation. We note that as model development was done on RBGE collections alone, this meant that reference elements were similar across many herbarium sheets in the sample. Lastly, the issue detection

models also performed strongly, achieving a precision of approximately 1.00 for out-of-focus detection, 0.98 for over-cropping, and 0.93 for under-cropping.

The proposed workflow underwent further evaluation with herbarium digitisers at RBGE, which will inform refinements to the model, software interface, and the overall pipeline. The next step is to expand on this pilot project by testing the model with collections across different institutions, leading to more refinements for generalisability. This can also help us evaluate and estimate the total proportions of digitised herbarium sheets with issues. Finally, our global aim is to provide an AI-based quality control service that can be integrated into digitisation workflows not only at RBGE but across institutions worldwide.

Keywords

quality control, artificial intelligence, herbaria, computer vision, collection digitisation

Presenting author

Arianna Salili-James

Presented at

Living Data 2025

Funding program

DCMS AI Pilot Programme

Hosting institution

Natural History Museum, London

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Davis C (2023) The herbarium of the future. Trends in Ecology & Evolution https://doi.org/10.1016/j.tree.2022.11.015
- Poon S, Salili-James A, Addis C, Leong A, van Schalkwyk H, Zhang N, Hartley B (2026) Scaling AI Across the UK Culture and Heritage Sector: The NHM-DCMS AI Pilot

Programme. Biodiversity Information Science and Standards https://doi.org/10.3897/biss.9.182096

- Zhang N, Poon S, Salili-James A, Scott B, Phillips M, Drinkwater R, Cubey R, Haston E (2025) DCMS x NHM AI Pilot Project - RBGE - Enhancing herbarium digitisation with AI-driven computer vision for image quality monitoring. Zenodo https://doi.org/10.5281/zenodo.16910416