Digitization of natural history collections: A guideline and nationwide capacity building workshop in Malaysia

ABSTRACT

Natural history museum collections are the most important sources of information on the present and past biodiversity of our planet. Most of the information is primarily stored in analogue form, and digitization of the collections can provide further open access to the images and specimen data to address the many global challenges. However, many museums do not digitize their collections because of constraints on budgets, human resources, and technologies. To encourage the digitization process, we present a guideline that offers low-cost and technical knowledge solutions yet balances the quality of the work and outcomes. The guideline describes three phases of digitization, namely preproduction, production, and postproduction. The preproduction phase includes human resource planning and selecting the highest priority collections for digitization. In the preproduction phase, a worksheet is provided for the digitizer to document the metadata, as well as a list of equipment needed to set up a digitizer station to image the specimens and associated labels. In the production phase, we place special emphasis on the light and color calibrations, as well as the guidelines for ISO/shutter speed/aperture to ensure a satisfactory quality of the digitized output. Once the specimen and labels have been imaged in the production phase, we demonstrate an end-to-end pipeline that uses optical character recognition (OCR) to transfer the physical text on the labels into a digital form and document it in a worksheet cell. A nationwide capacity workshop is then conducted to impart the guideline, and pre- and postcourse surveys were conducted to assess the confidence and skills acquired by the participants. This paper also discusses the challenges and future work that need to be taken forward for proper digital biodiversity data management.