EXECUTIVE SUMMARY:

As of March 2024, the Digitisation Project is on course to deliver in time and within the agreed budget of £29.3M. We expect to conclude all barcoding and imaging work by October 2025 with all transcription work to conclude by December 2025 providing contingency for both workstreams through to project closure on the 31st March 2026 if required. The Supplier Contract (Max Communications) will be extended from the 1st April 2024 to 31st March 2026 in line with the BofT's Approval in June 2022. Discussion are being held with Defra for Comprehensive Spending Review funding to cover 2024/25 and plans are in place to source funding for 2025/26

Digitisation Project Mid-Point Review - Q4 2023/24
Project Background

The aim for the Kew Science Collections Digitisation Project is to create an Open Herbarium and Fungarium through the digitisation of the specimen collections and the development and deployment of a user-focused open access data portal, through which digital access will be provided to 7 million preserved specimens of vascular plants and 1.25 million fungi held at Royal Botanic Gardens, Kew (Kew). The development of an Integrated Collections Management System (ICMS) will enable efficient management of the collections and the data describing them, more efficient tracking of the use of specimens (helping RBG Kew to meet legislative requirements concerning access to genetic resources and benefit sharing, including the Nagoya Protocol of the Convention on Biological Diversity, and plant health legislation).

Kew provides an unparalleled resource, estimated to hold representatives of around 70% of species, many of which are unavailable elsewhere. The key benefit to the UK public for providing this digitised collection is an open access resource which will be used by scientists in UK and globally, putting the UK on a level position with other world leading institutions.

As of March 2024, the Digitisation Project is on course to deliver in time and within the agreed budget of £29.3M. We expect to conclude all barcoding and imaging work by October 2025 with all transcription work to conclude by December 2025 providing contingency for both workstreams through to project closure on the 31st March 2026 if required.

The project is operating 6.5% behind the projected end of Quarter 3 (2023/24) KPI (Max and In-house) for barcoding and imaging but is expected to be back on course by September 2024, an extended time has been provided to allow the new target for recapture work (Images which were rejected from initial capture) to be bedded in. Outcomes will be achieved as Max Communications (Max), the digitisation supplier, has increased the number of Digitisation Stations to support capture and re-capture work and the in-house Fungarium Team have an additional workspace increasing daily outcomes as evidenced in through January.

Between January and September 2023, we ran two pilot value for money projects modelling approaches using volunteers and in-house Kew staff for imaging and quality control (QC) to ascertain the relative cost effectiveness of this approach compared with that we have deployed over years 1-2 (mostly using a 3rd party digitisation supplier for imaging and in-house staff for QC with a more modest volunteer contribution). These concluded that the 3rd party supplier approach was cost effective, could be delivered in time, while minimising the impact on other core work taking place within the Herbarium.

Considering this, in October 2023 the Digitisation Board approved seeking to extend our contract with the digitisation supplier (Max Communication Ltd) for a further 2 years starting April 2024. Max

have shown their desire to continue to work with Kew on the project, on similar consistent terms to the original contract.

The Integrated Collections Management System (ICMS) (now holds more than 2M specimen records) and we are on track to conclude all the implementation work by March 2025 and by the end of the overall project will provide management and access to over 9M records. A new data portal was launched as an alpha release in October, further work is planned during 2024/25 to advance the functionality of this system.

Two key focus areas in the coming quarter (January – March 2024) are (i) to increase the transcription quality assured acceptance levels and (ii) to develop an automatic transcription ingestion into ICMS. At present there is an existing backlog of quality assured produced Max transcription data, which is now predicted to be cleared by the end of March 2024, after which we expect to keep pace with all new incoming data.

1. Overview of Delivery to Date:

- **Integrated Collection Management System (ICMS)** - The ICMS implementation work is on track to conclude by March 2025. Over 2M specimen records are now managed from this system together with all transactions (loans & borrows) between Kew and other institutions and external researchers are managed. The Team will add the Economic Botany, Living Collection and Microscope Slides data into ICMS over the next 15 months. It will also provide the ability for the Library's Illustration and artefact data to be managed within the system as it is produced. The ICMS will hold and manage more than 9m specimen records by March 2026.

- **Data Portal** - A new Data Portal has been developed replacing the outdated HerbCat and HerbTrack systems and went live as a soft release Alpha Version in October 2023. The Data Portal system currently supplies access to 1.5M specimens records, with an expected 9M specimen records to be accessible by March 2026.

- **Digitisation Project** – Over 3M of the 6.75M specimens will have been fully transcribed by 31st March 2024. Based on these data, the Project Board is confident that the project will meet its final KPI by the intended project end date of 31st March 2026

2. KPI Performance & Projections to March 2026
Assessment of Performance - Looking Ahead to March 2026
Capture and Transcription
Based on the current monthly performance of Max, the In-house teams and volunteer activity we are confident that all capture work (barcoding imaging and metadata) will be completed by October 2025 with transcription work projected to conclude by December 2025 allowing contingency if required.

Quality Assurance & Challenges
On receiving the supplier's software, it was apparent many changes were required to support Kew's quality assurance work processes. The time required to make and test these changes in collaboration with the supplier resulted in a backlog of 700k images and 1.98m transcription data (to date) yet to be fully quality assured.
To manage and address this issue, an external Quality Assurance Supplier has been procured to manage the 700k backlog of images which are planned to be cleared by Quarter 2 of 2024, and additional QA staff employed. This will allow volumes to increase for both image and transcription data.

Data Cleaning & Importing Workstreams

Data cleaning has been increasing in pace during Q3 & 4 with the implementation of more efficient software scripts increasing the volume of data available to be imported into the Integrated Collection Management System (ICMS).

3. Project Budget

In November 2021, the Defra approved budget was £29.3m running between 1st April 2022 to 31st March 2026. By the 31st March 2024 the project expenditure to date will be £14.5m with the total projected cost to the end of the project remaining at £29.3m. In this period the project has made savings on the Digitisation Supplier coming in at a lower tender price (£0.5m), and staff costs set out within the original budget.

These savings have been reallocated to

- appointing 11 FTE staff to support an increase in supervision of staff, project activities such as imaging and quality assurance work.
- Cost of living (CofL) salary increase for all Kew project staff in 2023/24 and over the next two years.

- A contingency of £0.4m is available within the budget between April 2024 to March 2026 for any further cost increases. This contingency is 3% of the projected cost £14.05m for the final 2 years of the project.

- Max have provisionally agreed to keep the same unit costs for the final 2 years adding no added cost increases on their contract price.

4. Funding

The first two years of the project have been fully funded by Defra and an application for year 3 funding (2024/25) has been made to Defra as part of the business planning process. UK government funding AHRC for digitisation of UK natural history collections is likely to become available in 2026/27, but it is unclear if Kew will have access to this within our planned timeframe. We are likely to also seek funding for year 4 from Defra. c£1m of Kew's £2m fundraising target has been raised to date.

Kew Board of Trustees have signed off the Defra Loan should it be needed while further work takes place to secure funding to ensure the project can be delivered on schedule.

5. Making Digitised Data Available

As aforementioned the ICMS currently holds 1.94m specimen records which include nearly all data from Kew's core collections, and the Data Portal providing external global access to 1.5m Herbarium and Fungarium unrestricted records drawn from ICMS.

On reviewing the current delivery outcomes, we plan to add over 3.1m records Herbarium and Fungarium records within ICMS by March 2025 taking it to a total of 5.06m records and circa 9m records (Herbarium, Fungarium and all other core collection data) by the end of March 2026 accessible to Scientists and Researchers at Kew.

All unrestricted Herbarium and Fungarium specimen records will be drawn from the ICMS and made available within the Data Portal providing worldwide accessibility to Kew data, this will be a build on the current c1.5m specimens records currently available with new digitised data to be added from the end of Quarter 2024/25.

	Pre 31/03/24	April 2024	April 2025	April 2026	
ICMS	1.94m	2.64m	5.06m	c9m	Available to Kew Science staff

Global Access to Herbarium & Fungarium Data

	Pre 31/03/24	July 2024	April 2025	April 2026	

Data Portal	1.5m	2.2m	4.62m	c8m	Available to scientists globally
GBIF	1.05m*	2.2m	4.62m	c8m	Available to scientists globally

Just Herbarium data – Fungarium data will be added from July 2024

In addition to these numbers, we continue to monitor the number of cites to illustrate usage of data as part of the projects Benefits Realisation.

For advice or guidance on pursuing your own digitisation project, please contact project manager John Adcock: j.adcock@kew.org