

## Enhancing accessibility of forest flora data in Tanzania

---

**Programme:** BID  
**Project ID:** BID-AF2020-038-INS  
**Project lead organization:** Tanzania Forestry Research Institute  
**Project implementation period:** 1/4/2021 - 30/3/2023  
**Report approved:** 29/3/2022

### Narrative Midterm report

---

#### Executive Summary

---

We started implementing this project late June 2021. In June we managed to conduct a training workshop whereby herbarium staffs and project members were trained on how to perform biodiversity data digitization, mobilization and data publishing. Furthermore, to date we have digitized more than 10,000 specimens. Among these, a data set of flowering plants (8,689 specimens) have been published on GBIF data portal. Currently, we are progressing with digitization work parallel with updating specimen taxonomy and labels.

#### Progress against milestones

---

**Has your project published at least one dataset through GBIF.org?: Yes**

**Dataset published:**

Dataset	DOI
Diversity and composition of preserved flowering plant specimens at Tanzania Forestry Research Institute (TAFORI) Herbarium	<a href="https://doi.org/10.15468/sktqsa">https://doi.org/10.15468/sktqsa</a>

**Has at least one member of your project team received certification following the BID capacity enhancement workshop?: Yes**

**Name of the workshop participant:** Thomas Corodius Sawe

**Certification obtained:** Basic Badge

#### Report on Activities

---

##### Activity progress summary

Project team is progressing with the implementation of the project activities as planned though at low pace than anticipated. This is because we didn't include some important activities during proposal write up. These include;

##### **(1) Data Capture and standardization following Darwin Core format**

This is core activity of digitizing data preserved in Herbarium. This activity involves extracting information from preserved specimen in to excel sheet following the Darwin Core Data standards. In the submitted proposal we allocated more time on image acquisition and processing. We didn't not consider Data capture as individual activity and therefore we did not allocate sufficient resources (time and budget) for its implementation.

##### **(2) Data Cleaning and Validation**

Captured data from specimen labels contain number of errors including spelling errors, wrong date formats, incorrect taxonomy e.t. c. Therefore we are investing significant amount of time to clean the data and update taxonomy using online resources.

### **(3) Geo-referencing**

Nearly all of the specimens collected before year 2000 in our herbarium do not have geographical coordinates. Therefore we are also investing substantial amount of time which we didn't account for during proposal submission to perform geo-referencing.

### **(4) Updating specimen labels**

One of the objectives in this project is to update taxonomy ranking of the preserved specimens since most of them were collected in 19s and their taxonomy status have changed. In addition, some specimens were labeled using synonyms. We have therefore been working to update the information in the captured data as well as in the specimen labels by creating new labels with updated taxonomy and geographical coordinates.

The above mentioned activities were necessary to be implemented before the carryout of some activities proposed in the submitted proposal including image acquisition. For this reason, some adjustment have been made on implementation of project activities as follow;

- Proposed activity number two (2) and seven (7) on the submitted proposal are currently integrated within new activities such that; (i) Pre-digitization begins, sequentially followed by the new activities. (ii) The captured information from the specimen is uploaded into the in the GBIF portal for Publishing. This means that we have already started implementing activity number seven in the submitted proposal which was planned to commence in March and end in September 2022. Due to these changes activities 2 and 7 in the submitted proposal are expected to be completed in September 2022 parallel to the new activities (above).
- Furthermore, proposed activity number 5 (Imaging/scanning each specimen individually) and 6 (Specimen image processing) in the submitted proposal will not take place until October 2022 and end in March 2023. This is because before image acquisition we need all specimens to have updated label.

### **Completed activities**

#### **Activity: TAFORI herbarium staff capacity building**

---

**Description:** Eight Herbarium staff and three other project members were trained on biodiversity data mobilization and digitization. Specifically, the training focused on;

- (1) Biodiversity Data standards
- (2) Biodiversity data transformation
- (3) Darwin Core Data Standards
- (4) Data publishing using IPT
- (5) Meta data preparation and data Mapping
- (6) Data uploading and publishing

**Start Date - End Date:** 13/6/2021 - 25/6/2021

**Verification Sources:** See Training report

### **Report on Deliverables**

---

#### **Deliverables progress summary**

Project team is working to attain project deliverables. We are working to complete a data set of 55,000 specimen documented in submitted proposal. However, we have revised this deliverable to increase efficiency and productivity in implementing the project. We are expecting to have four data sets instead of one as mentioned in the submitted proposal. This will increase usability of our herbarium data. The data sets will be as follows;

- (i) Data set of flowering plants (covering Dicotyledons and Monocotyledons) Specimens
- (ii) Data sets of Pteridophytes
- (iii) Data set of Gymnosperms
- (iv) Data set of Bryophytes

In addition, we have a new deliverable on Specimen labels. Most of our specimens are old and lack geographical coordinates. We are therefore working to update the nomenclature/taxonomy of

specimens that have changed their taxonomy. In addition we carrying out Geo-referencing to obtain geographical coordinates for specimens lacking coordinates. Finally we are preparing new specimen labels to include geographical coordinates and new taxonomical name for each label.

## Progress towards deliverables

### Dataset deliverables

---

#### Plant collection

**Dataset type:** Occurrences

**Dataset scope:** This data set was extracted from 8689 sheets of preserved Dicotyledon flowering specimens under class Magnoliopsida covering 111 families, 737 Genera, and 1985 species. Dominant families in this data set are; Asteraceae (12.5%), Fabaceae (9.3%), Apocynaceae (6.8%), Euphorbiaceae (6.7%), Lamiaceae (6.5%), Acanthaceae (5.7%) and Combretaceae (5.1%). Dominant species in this data set are *Brachystegia speciformis* (5.2%) and *Combretum* (2.7%) belong to Fabaceae and Combretaceae families respectively.

**Number of records:** 8,813

**Data holder:** Tanzania Forestry Research Institute

**Data host institution:** Tanzania Forestry Research Institute

**% complete:** 100%

**Status update:** Published

**DOI:** <https://doi.org/10.15468/sktqsa>

**Expected date of publication:** 2022-01-20

### Other deliverables

---

#### Herbarium staff capacity building

**Description:** Herbarium staff trained

**% complete:** 100%

**Status update:** Completed

**Sources of verification:** Training report attached

#### Specimen labels updating

**Description:** We have produced new labels for 9000 specimen with updated scientific name and geographical coordinates.

**% complete:** 100%

**Status update:** Complete

**Sources of verification:** Attached mid report

### Communications and visibility

---

Links of published data sets will be made available to TAFORI institution website to increase visibility and accessibility

### Monitoring and evaluation

---

#### Monitoring and evaluation findings

By the midterm of this project we had planned to complete number of project activities including herbarium staff training and purchase of imaging station. The project team has successfully managed to complete some of the planned activities and we are still working to complete others. So far we have successfully trained 8 herbarium staff on how to digitize herbarium data and to publish on GBIF data portal. In addition, we have managed to digitize 9000 specimens of flowering plants whereby occurrence dataset from 8813 preserved specimens have been published into GBIF data portal.

Despite the success, some of the activities that were planned to be completed within the mid report period in the submitted proposal are yet to be completed. For example, activity number two (2) in the submitted proposal on "Pre-digitization, specimen curation and staging" will not be completed until end of September 2022. The delays are due to the fact that we didn't allocate sufficient time for some of the planned activities also we completely didn't include important activities which are necessary in data digitization process. These activities include; (i) Data Capture and standardization following Darwin, (ii) Core format, (iii) Data Cleaning and Validation, (iii) Geo-referencing and (iv) Updating specimen labels. We have revised the project work plan such that these new activities are integrated within the activities

proposed in the submitted proposal. However, this has implications on the time line of other activities.

TAFORI and project management is ensuring working environment is good for the staffs and interns working on the project. On the other hand, GBIF secretariat is providing important support through Africa BID support contractor and node managers.

### Impact of COVID-19 pandemic on project implementation

---

No COVID-19 Impact

---

GBIF leads the Biodiversity Information for Development (BID), a programme funded by the European Union. The programme provides supplementary support for activities addressing the needs of regional researchers and policymakers through mobilization and use of biodiversity data.

