

Filling gAps iN daTA about BiodiverSiTy with NaTural History Collections in the Western GHATS from IFP and Baro Herbarium

Programme:BIFA
Project ID: BIFA6_015
Project lead organization:French Institute of Pondicherry
Project implementation period:1/9/2021 - 28/2/2023
Report approved: 21/6/2022

Narrative Midterm report

Executive Summary

On 31/05/2022, the project is going on as planned. All the questions about data capture and data curation are solved, and the data capture is going in a good rhythm. As planned, we organized the main meeting between both teams as soon as possible and this meeting was successful in taking major decisions about the workflow. As planned, Baroda's team already realized a mission to Dangs herbarium.

The French Institute of Pondicherry possesses 27094 herbarium sheets
Mid-May 14000 have been digitized (51%)
Mid-May 15000 have been captured (55%) in the database and 880 (3%) have been also curated and published as a test sample before mid-term report.
This small number is explained by our workflow. We are building tools for bulk curation, not day-to-day, one-by-one curation. The next bulk (of 15 000 occurrences) will be processed very soon but we decided to focus on a small sample for the mid-term.

Baroda and Dangs Herbaria possess around 50 000 herbarium sheets
Mid-May 30 025 have been digitized (60 %)
Mid-May 32421 (64 %) have been captured in the database and 6211 (12 %) have been also curated. A set of 3598 occurrences (which is a special collection from Dangs Herbarium) has been published on the Gbif.

The important volume of data to be captured (both scanning the herbarium sheet and capturing data in the database) make it complicated to fulfill the objectives in the frame of 18 months, but we are working as fast as possible. In the financial report You'll find a request to transfer mission budget into salary to get 3 more months for Baroda team to be sure of the complete curation.

Progress against milestones

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

Dataset published:

| Dataset | DOI |
|--|---|
| BARO Herbarium, The M.S University of Baroda | https://doi.org/10.15468/ky5p3k |
| Herbarium of the French Institute of Pondicherry | https://doi.org/10.15468/un3c1t |
| | |
| | |

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

Name of the workshop participant:Julien Andrieu

Certification obtained: Advanced Badge

Report on Activities

Activity progress summary

When we applied for the grant we thought that it should be efficient to divide the workflow into two subsets of data based on the date of collection (pre and post-2012) because the data collected in the last 10 years is complete (for example always possess precise geographical coordinates from GPS).

However, once faced the organization of both the herbaria, such intent has been discarded and the whole data is being captured and digitized without this subdivision in periods.

Such division had been applied to the list of activities, therefore we can't declare, for this mid-term report that the data post-2012 has been completely captured and digitized.

The advancement of the project remains in the same proportions and we are confident in our engagement toward the Gbif.

Completed activities

Activity name: Data curation flow conception

Description: The whole team reunited in Pondicherry to discuss the main points of the data curation flow. Decisions were taken on the use of SPECIFY 6 and Google refine and a lot of technical decisions.

Start Date - End Date: 1/9/2021 - 31/3/2022

Verification Sources: Cf. Report on data curation workflow.

Activity name: Half of the data is captured

Description: previously entitled 'Post-2012 data entry' but the workflow has changed.

Start Date - End Date: 1/9/2021 - 31/5/2022

Verification Sources: Report on advancement

Report on Deliverables

Deliverables progress summary

As planned in the original Gantt Chart none of these 3 deliverables is concluded at this midterm. However, as soon as the midterm report will be validated, I will start the organization of the seminar on open data with my colleagues from Baroda, immediately after, I will focus on the redaction of the data paper. Of course, the data curators are working continuously toward the readiness of the data for publication...

Progress towards deliverables

Title: Data publication through Gbif

Type: Dataset

Status update: Not yet

Dataset scope: Open biodiversity data

Expected number of records: 27000

Data holder: IFP

Data host institution: IFP

Sampling method: None.

% complete: 3

DOI: <https://doi.org/10.15468/un3c1t>

Expected date of publication:

Title: Data paper

Type: Data Papers

Description: Description of the dataset

Sources of verification: Not started yet. Still planned to be ready in 9 months but only published after the data is online.

Title: Seminar on open data

Type: Other

Description: We planned to also work on the upscaling of the publication of open access data in India. We'll organize a seminar on this.

Sources of verification: Not yet started. It was planned for the 2nd half of the project.

Events

Mission to Pondicherry

Dates: 2021-11-08 - 2021-11-11

Organizing institution: IFP

Country: India

Number of participants: 9

Comments: This mission has been a success in finalizing the workflow between both teams.

Website or sources of verification: Report attached.

Events

Mission to Dangs

Dates: 2022-03-02 - 2022-03-12

Organizing institution: Baroda

Country: India

Number of participants: 2

Comments: This mission planned from the start has been a success in botanical verification of some herbarium sheets with experts.

It is also part of our engagement of communication and dissemination of open data practice with other indian institution.

Website or sources of verification: Report attached.

Communications and visibility

Some communication is regularly done as per the communication plan of our institutions.

First, we did the necessary to be registered on the Gbif webpage.

https://www.gbif.org/project/BIFA6_015/filling-data-gaps-with-the-natural-history-collections-of-the-western-ghats

Then we created a page on IFP's website :

<https://www.ifpindia.org/projects/filling-data-gaps-with-the-natural-history-collections-of-the-western-ghats/>

And the project has been presented in various newsletters eg.:

<https://www.ifpindia.org/patrika/patrika-issue-55/>

<https://www.ifpindia.org/patrika/patrika-issue-55/focus/>

Monitoring and evaluation

Monitoring and evaluation findings

The data curation workflow has indeed been checked and validated during the mission of Baroda's team to IFP in November 2021.

Every month (sometime more than that) each institution holds a meeting on the progression of digitization.

Every 5 months we've held a general meeting on the progression of digitization and if some decision has to be taken by the team.

The 3 members in charge of data capture and the PI have very regular zoom meetings to keep in touch and resolve various little matters.

Such monitoring is fine and the workflow is fluid.

Impact of COVID-19 pandemic on project implementation

NA

GBIF leads the Biodiversity Information Fund for Asia (BIFA), a programme funded by the Ministry of the Environment, Government of Japan. The programme provides supplementary support for activities addressing the needs of regional researchers and policymakers through mobilization and use of biodiversity data.

