

# 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:** 3/5/2023

## Narrative Final report

---

### Executive Summary

---

The team of FANTASTIC\_Ghats is very pleased to report the full completion of the project, on time, with all due deliverables. A total of 52,353 occurrences have been published on the Gbif Portal on the date of 27/02/2023. More precisely, the Dang Herbarium dataset (3,598 records) and Baroda (21,668 records) have been published on 10/02/2023 and the herbarium of the French Institute of Pondicherry (27,087 records) has been published on 27/02/2023.

The Online seminar has been held on the 14th of February 2023 at 10:30 am. It has been a success with more than 80 participants from various countries in South Asia. A data paper is ready to be submitted to Biogeographia.

Since the beginning, the approach has been collaborative amongst partners (IFP and MSU) and within each (Botany Researchers and professors, Engineers in IT, Research assistants, etc). This approach has proven efficient. We soon understood that some methods would be more efficient for some institutions than for others so we have not applied the exact same workflows. Obviously, the curation and the final formatting of published data have been fully aligned.

No particular problem has occurred and we have been able to simply stick to the roadmap, the calendar, the list of activities, and deliverables.

### Progress against milestones

---

**Has your project completed all planned activities?: Yes**

**Has your project produced all deliverables?: Yes**

## Report on Activities

---

### Activity implementation summary

- A data curation flow has been set up before the end of March 2022 in order to validate geographic coordinates and taxonomy before the entry of occurrence data. It included erratic data filtering or corrections in geographical coordinates and in taxonomy. It included building a process to enter data in the Darwin Core standard format.
- Every Dangs, Baroda and IFP sheet has been scanned according to standards before February 2023.
- Data of every Dangs, Baroda and IFP sheet has been captured following the Occurrence Standards of the Gbif and the Darwin Core before February 2023.
- The only slight difference between the provisional calendar and the realized one is that we

overestimated the difference before “recent” and “Old” data. When planning for the project we expected an easy curation for the post-2012 data for 2 reasons, the use of GPS at the moment of data collection and fewer changes in species names. It happens that curation has basically been made the same way for all the data without real distinction according to the date of observation. So the capture and curation have been a continuous flow for 18 months.

- Training on open data, data curation challenges, and standards for natural history collections with other Indian institutions working on the Western Ghats.
- Before 31/01/2023 the PI of FANTASTIC\_Ghats Julien Andrieu and Riya Engineer from Baroda gained certification at "BIFA Capacity Enhancement Workshop". The following year Srilatha R., at IFP, also gained certification at BIFA Capacity Enhancement Workshop.
- The new dataset accessible on the Gbif will be presented in the form of a data paper, collectively written, in *Biogeographia* (e-ISSN: 2475-5257).

## Completed activities

### Activity name: Data curation flow conception

---

**Description:** We followed the training of Gbif, and trained also on other tools and software (Specify), we met in order to fix a common frame and the Data flow curation has been fully set up in time.

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

**Verification Sources:** See the validated data set on the Gbif

### Activity name: Scan sample sheets from herbaria

---

**Description:** At the end of the project, 100% of the herbarium sheets have been scanned.

**Start Date - End Date:** 1/9/2021 - 28/2/2023

**Verification Sources:** See the dataset on the Gbif portal

### Activity name: Data capture

---

**Description:** 100% of the herbarium sheets have been object of data capture following Darwin core for occurrence data standards.

**Start Date - End Date:** 1/9/2021 - 28/2/2023

**Verification Sources:** See data published

### Activity name: Data validation

---

**Description:** 100 % of the captured data have followed rigorous data curation from the producer's point of view (curation of names and geographical coordinates) then from the publisher's point of view (Google Refine Gbif verification tools) etc.

**Start Date - End Date:** 1/9/2021 - 28/2/2023

**Verification Sources:** See published data

### Activity name: Trainings

---

**Description:** First, the project has been the occasion for major internal training and capacity improvement.

Then, following our planned list of deliverables, a webinar has been organised.

**Start Date - End Date:** 1/9/2021 - 28/2/2023

**Verification Sources:** See report on the webinar

### Activity name: Data paper preparation

---

**Description:** Julien Andrieu, PI of the project is writing the data paper to be submitted to biogeographia. The data paper will not be published before the data is shared through the Gbif portal.

**Start Date - End Date:** 1/2/2023 - 28/2/2023

**Verification Sources:** See the draft paper in the report section

### Activity name: Data publication through Gbif

---

**Description:** Before the 31/01/2023 the PIs of FANTASTIC Ghats (after having gained certification at BIFA Capacity Enhancement Workshop) has published and shared, through the Gbif, using Integrated Publishing Toolkit the complete occurrences data set combining the Natural History collections from the IFP and the Baro Herbarium.

**Start Date - End Date:** 1/2/2023 - 28/2/2023

**Verification Sources:** See published data

## Report on Deliverables

---

### Production of Deliverables - Summary

The four deliverables are successfully following the calendar. The data has been digitized, captured, curated, and published as an occurrences dataset with herbarium sheets as supplementary material. The webinar has been organized. The data paper will follow shortly (respecting the obligation to wait for the data to be published).

#### Production of deliverables

##### **Title: Digitisation of herbarium sheets**

---

**Type:** Dataset

**Status update:** 100%

**Dataset scope:** Open biodiversity data }

**Expected number of records:** 26331

**Data holder:** IFP and Baroda

**Data host institution:** GBIF

**Sampling method:**

**% complete:** 100

**DOI:**

**Expected date of publication:**

##### **Title: Data publication through Gbif IFP**

---

**Type:** Dataset

**Status update:** 100%

**Dataset scope:** Open biodiversity data }

**Expected number of records:** 25023

**Data holder:** IFP

**Data host institution:** GBIF

**Sampling method:**

**% complete:** 100

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

**Expected date of publication:**

##### **Title: Paper in Biogeographia**

---

**Type:** Data Papers

**Description:** Draft well advanced

**Sources of verification:** See draft in report section

##### **Title: Seminar on open data**

---

**Type:** Other

**Description:** The seminar was held on the 14th Feb. It has been a success.

**Sources of verification:** See report in the report section

##### **Title: Data publication through Gbif Baroda**

---

**Type:** Dataset

**Status update:** 100% of occurrences Media file has a bug. Mentors of BIFA are searching for a solution

**Dataset scope:** Open biodiversity data }

**Expected number of records:** 25470

**Data holder:** Baroda

**Data host institution:** GBIF

**Sampling method:**

**% complete:** 100

**DOI:** <https://doi.org/10.15468/hdwj5g>

**Expected date of publication:**

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

---

None

## Events

---

### 'Natural History Collections & Open Science

---

**Dates:** 2023-02-14 - 2023-02-14

**Organizing institution:** IFP

**Country:** India

**Number of participants:** 79

**Comments:** See report

**Website or sources of verification:** <https://www.ifpindia.org/announcements/online-seminar-on-natural-history-collections-open-science/>

### Communications and visibility

---

The Online Seminar and the data paper are the two main efforts in communication. In addition, the PI, Julien Andrieu, has accepted to become an Open Biodiversity Data Ambassador.

### Monitoring and evaluation

---

#### Final Evaluation

The project has been well prepared and we have been quite simply following the calendar and budget except for a very few adaptations. It has been very useful to have been allowed to amend the budget to add a few months of salary to the Baroda team.

We have been happily surprised by the success of the webinar. Let's hope that it will help India to invest in the GBIF or at least in open biodiversity data initiatives.

#### Best Practices and Lessons Learned

Communication is very important and the PI has maintained a regular communication flow with all the members of the project. It has been important to let the 2 institutions manage a part of free will in the application of the workflow (as long as the output was following the commonly decided format).

#### Post Project Activity(ies)

The data paper will be submitted shortly after the end of the project respecting the necessary order (publishing data first), it will require attention until full publication (peer review...)

### Sustainability plans

---

Both collections are not only historical collections but are getting completed with regular activities. This project has been the basis of the conversion to the open access approach of our data but the institution will update their dataset on the Gbif regularly.

---

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.

