

Report

Mission to IFP

A visit to IFP as planned under the GBIF's BIFA program, Biodiversity Data Mobilization project titled "Filling gAps iN daTA about BiodiverSiTy with NaTural History Collections in the Western GHATS from IFP and Baro Herbarium" was scheduled from 8th November 2021 to 13th November 2021. This visit was primarily scheduled to discuss work flow so that both the teams can be on same track throughout the project. The meeting comprised of members of both the teams including senior botanists (Dr. Padamnabhi Nagar, Dr. Ayyappan, Dr. Balachandran), PI (Dr. Julien Andrieu), IT expert (Srilatha, Dr. Balachandran) and Junior Research Fellows (Pradeep Chandrasegarane, Riya Engineer and Krishna Rajput).

On the day 1 one of the meeting, basics of data digitization was discussed. What kind of data both the teams already have and what is lacking was discussed and based on that how both teams will work was decided. Hierarchy of classification for GBIF occurrence dataset was discussed. In what way the scientific names of the specimens will be resolved and what all tools and applications can be used was talked about. Along with the above discussions, Capacity Enhancement Workshop for BIFA programme organised by GBIF was simultaneously going on in which Dr. Julien and Riya has participated and all that was learned about digitizing data during the workshop was also discussed with other team members during the meet.

On day 2, types of datasets supported by GBIF and Darwin Core Archive was discussed. For our dataset which is to be published in the form of occurrence type of dataset, following Darwin Core fields were discussed: occurrenceID and how to create it, format of eventDate and what should be done if date is missing from the data, format of scientificName, institutionCode, collectionCode and catalogNumber, verbatimLocality and what steps should be taken to determine missing location from the data, occurrenceRemarks and eventRemarks, organismID and few other fields. Also, what are required fields and what are strongly recommended fields for occurrence type of dataset was discussed. Few examples of datasets which were already published on GBIF were studied for the reference and once all the Darwin Core terms were fully understood, which fields are relevant to our dataset and which are to be kept were decided by both the teams.

Day 3 observed major discussion on data curation, cleaning and management and tools and applications that can be used to do so. Learning to use Open Refine for data cleaning was an important task of that day. Different functions and features of open refine was explored and how to manage bulky data without spending huge amount of time was sought out.

Day 4 was all about publishing cleaned and curated data using IPT. Also, timeline of achieving different milestones of the project such as midterm report and different ways of working out to achieve timely targets was discussed.

This mission has setup a foundation for how to work out throughout the rest of the project for both the teams. Baroda team got an opportunity to visit the herbarium of IFP. Both the teams got to learn many new things about data digitization and it's

importance of publishing it on an open access website and what challenges one faces during the whole process.

