Mobilizing biodiversity data from natural history collections in Fiji

Programme:BID
Project ID: BID-PA2020-006-NAC
Project lead organization: The University of the South Pacific, Institute of Applied Sciences
Project implementation period:1/7/2021 - 30/6/2023
Report approved: 10/10/2022

Narrative Midterm report

Executive Summary

To-date we have achieved the mandatory milestones, that is, one dataset has been published and at least one person from the project team has received certification from the BID capacity enhancement workshop. We finally were able to host our first workshop, albeit virtual, in April 2022 and conduct an engagement survey online. These were activities proposed for 2021 and it brought the team huge relief to finally accomplish these tasks despite the delays and setbacks and having to adjust our activities slightly than originally planned.

We had a number of teething issues at the beginning of the project in 2021 that set us back considerably, particularly in terms of losing time to meet deliverables. Some of these were administrative issues where there was a delay in signing our contract with GBIF and our request for an extension on the project timeframe being denied. Other issues were pandemic-related.

We have weekly virtual meetings to check in on each other's mental well-being and how we're progressing with our individual assigned tasks. These meetings also provide a space for us to discuss any problems with our datasets and keep discussions flowing regarding project activities and deliverables. Minutes are recorded which help with monitoring our progress.

We have requested for a budget amendment separately. We also realized that we proposed to publish two marine datasets for the period June 2022-May 2023. This is incorrect. It should only be one dataset. We wrongly entered duplicate deliverables for our marine dataset. In our proposal we also mistakenly used the word 'digitization' instead of 'data capturing' or 'transcribing' for our project activities. There is also a possibility that we might not publish more than 55,000 occurrence records as originally proposed.

Progress against milestones

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

Dataset published:

Dataset	DOI
Aquatic biodiversity records at The University of the South Pacific	https://doi.org/10.15468/w8jpdj

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

Name of the workshop participant: Yumiko Baba

Certification obtained: Basic Badge

Report on Activities

Activity progress summary

1) Initially we had proposed to conduct an introductory theory session with our project partners to familiarize them with the importance of data mobilization and the details of our BID project. After many lengthy discussions in 2021 we decided that it would be best to also include other local biodiversity data holders and practitioners and include a practical session that involved training them in specimen collection and management. Unfortunately due to pandemic restrictions on any form of gatherings where there was a limit to the numbers that could be physically present and the uncertainties of when these restrictions would be lifted we had to keep postponing our workshop dates. We then altered our plans for a collaboration workshop and engagement survey to accommodate for delays and setbacks brought about by the pandemic and lockdown restrictions. The adjusted plan included splitting the initial proposed collaboration workshop into two parts, a theory session delivered virtually and a practical session to be delivered face-to-face once restrictions eased up. The theory part was delivered on 28th April 2022 via Zoom, and was received well by the participants. It was a great learning experience for all and this was evident during the Q & A session where discussions revolved around sensitivity of sharing data and data ownership. The practical session will be held later this year.

2) Prior to the virtual workshop we also conducted an engagement survey using google forms to assess the situational needs of biodiversity data holders. In the "report attachments" a copy of the survey form, a copy of the virtual workshop programme, a list of the participants names, and a summary of responses from the survey is attached.

3) We have identified Specify to be used by both the herbarium and the marine collection. This is after virtual discussions in 2021 with the collection managers from National Institute of Water and Atmospheric Research (NIWA) in New Zealand, who were kind enough to give us a brief introduction on the usability and functionality of Specify. We are currently in discussion with the USP's IT department to identify someone to assign the task of installing the database. There have also been discussions at the possibility of upgrading and maintaining the herbarium's local database but Specify may be the better option in the long run. Either way we have to work with our IT department's processes to install any database.

4) Data-capturing and mobilizing data has progressed at a snail's pace but is progressing nonetheless. Scanning of marine catalogue books took longer than expected with limited access to material and facilities during lockdown in 2021. Being on lockdown also meant that access to herbarium specimens and data was restricted. The herbarium database can only be accessed via the university's network therefore cross-checking data from our excel sheet to what's on the database or cross-checking against physical records was not possible. Now that we are no longer working from home and able to properly access the herbarium database and it's exported data, we have discovered duplicate records and missing data for a number of fields from a dataset containing more than 30,000 records. To clean the data we have decided to work with small chunks of data at a time, starting off with approximately 300 records of ferns. Once we have established a working system with the ferns, the work flow will be a little smoother with the rest of the dataset. This may mean we will have a smaller dataset to publish than we initially anticipated. But we will at least be publishing records that has been thoroughly scrutinized and has greater data integrity.

Completed activities

Activity: Engagement survey

Description: Engage and survey endusers via a situational and needs online assessment of biodiversity data practitioners and their organizations. The form was made available online for at least two weeks.
Start Date - End Date: 26/4/2022 - 6/5/2022
Verification Sources: A sample of the survey form and a summary of the responses received is in the "Report Attachments" section.

Activity: Collaboration Workshop

Description: Introductory workshop with our project partners from government and other non-government biodiversity data holders, introduce/familiarize them with data mobilization, what GBIF is and the details of our project. Theory part only (two hour session via Zoom).
Start Date - End Date: 28/4/2022 - 28/4/2022
Verification Sources: In the "Report Attachments" is a list of participants that attended the virtual theory session and a copy of our programme.

Deliverables progress summary

1) We have published 821 new marine records of the 10,000 proposed records to be published (https://doi.org/10.15468/w8jpdj).

2) There has been a considerable delay in publishing any of our plant records. We are working through the data cleaning slowly but steadily. We have discovered duplicate records and missing data for a number of fields from a dataset containing more than 30,000 records. There have been a number of discussions on how to move forward with our dataset and we have decided to work with small chunks of data at a time, starting off with approximately 300 records of ferns. Once we have established a working system with the ferns, the work flow will be a little smoother with the rest of the dataset. This may mean we will have a smaller dataset to publish than we initially anticipated. We hope to publish our first plant dataset, before the end of August.

Progress towards deliverables

Dataset deliverables

Publish new plant records
Dataset type: Occurrences
Dataset scope: Increase occurrence records of plants from Fiji.
Number of records: 24,000
Data holder: South Pacific Regional Herbarium
Data host institution: University of the South Pacific
% complete: 0%
Status update: Still data cleaning.
DOI:
Expected date of publication: 2022-08-31

Publish new plant records
Dataset type: Occurrences
Dataset scope: Occurrence records of plants of agricultural crops from Fiji
Number of records: 4,000
Data holder: Ministry of Agriculture
Data host institution: Ministry of Agriculture and South Pacific Regional Herbarium
% complete: 0%
Status update: Yet to collate data.
DOI:
Expected date of publication: 2022-08-31

Publish new marine records
Dataset type: Occurrences
Dataset scope: Occurrence records of Pacific marine collections based on physical specimens from Fiji.
Number of records: 10,000
Data holder: USP Marine Collection
Data host institution: University of the South Pacific
% complete: 8%
Status update: 821 records published of the 10,000 proposed.
DOI: https://doi.org/10.15468/w8jpdj
Expected date of publication:

Other deliverables

Events

GBIF Data Mobilization (virtual) Workshop

Dates: 2022-04-28 - 2022-04-28
Organizing institution: University of the South Pacific
Country: Fiji
Number of participants: 21
Comments: We altered our plans for a collaboration workshop and engagement survey to accommodate for delays and setbacks brought about by the pandemic and lockdown restrictions. The adjusted plan included splitting the initial proposed collaboration workshop into two parts, a theory

session delivered virtually and a practical session to be delivered face-to-face once restrictions eased up. The theory part was delivered on 28th April 2022 via Zoom, and was received well by the participants. Prior to the virtual workshop we also conducted an engagement survey via google forms to assess the situational needs of biodiversity data holders. In the "report attachments" a copy of the survey form, a copy of the virtual workshop programme, and a list of the participants names is attached.

Website or sources of verification: In the "report attachments" a copy of the survey form, a copy of the virtual workshop programme, and a list of the participants names is attached.

Communications and visibility

Given that we've had a few teething problems trying to kick off our project and meet our deliverables, we haven't had much to share regarding our achievements to-date. But we did write a brief media release that's published on the IAS-USP webpage (https://www.usp.ac.fj/the-institute-of-applied-sciences/news/the-institute-of-applied-science-receives-fjd-90000-for-data-mobilization-project/) and we introduced details of our project and GBIF as a platform to mobilize data to our fellow biodiversity practitioners in our first virtual workshop held on 28 April 2022. For your reference, our programme for that workshop can be found in the "Report Attachments" section.

We endeavor to be more visible with our progress and plan on tweeting updates via the South Pacific Regional Herbarium twitter account (@SPRH1933). This appears to be the more popular and easiest means to communicate our work with the general public.

Monitoring and evaluation

Monitoring and evaluation findings

1) Weekly meetings are held via Zoom or Teams to monitor progress. Through these discussions we have learned how much effort is required for the data cleaning process of the plant records. We are working through the dataset in smaller working sets.

2) Due to the delayed delivery of the engagement survey and collaboration workshop we decided that holding the workshops of the exact nature twice, in a short span would be redundant. Therefore, we decided not to deliver the second-year engagement/collaboration workshop. Additionally, it became apparent that our project partners prefer for their Suva-based staff to be trained. We will keep it open for the Suva-based staff to attend the second session of our collaboration workshop. We have submitted a budget amendment request to reflect these changes with suggestions of how the funding can be otherwise spent.

3) Our project partners and other biodiversity data holders that we engaged in the first virtual collaboration workshop were receptive to the idea of mobilizing data, which was very encouraging. That first virtual workshop held on April 28 was an opportune time to introduce GBIF to our participants.

4) There is a possibility that we might not publish 55,000 occurrence records as initially proposed. This may only apply to our plant datasets where we are finding a number of duplicate records, which was not obvious in the beginning. This will become a clearer over the next few weeks after we have had the chance to publish our first plant dataset.

5) Support from GBIF secretariat has been great. They are always responsive to emails to any of our queries. They were also very sympathetic when one of our project members could not complete the BID enhancement capacity workshop due to the untimely demise of a loved one.

Impact of COVID-19 pandemic on project implementation

Fiji went into a second lockdown from April 2021 due to the pandemic. We were not able to properly return to work until 4th October 2021. The uncertainty, the constant adjustments to changing restrictions and curfews imposed on us, the mental burnout from being forced out of our usual routines and cooped up indoors, trying to stay COVID-free, enduring grief from losing loved ones, and a myriad of other factors related to our second wave took a heavy toll on us individually and in different ways, and working from home wasn't always ideal to say the least. Project members could not access the collection facilities and the university's internal network that delayed our plans to organize ourselves for the project. Additionally, there was uncertainty around when we could resume physical gatherings and there were also strict measures in place that limited the number of people that could be physically present in an enclosed space at any given time. This hampered our progress significantly. We were unable to conduct proposed workshops, we are unable to access material that would help us progress

with publishing datasets. One team member had two computers crash while working from home during lockdown which added to further delay of working towards deliverables. We were not permitted an extension to our project timeframe to make up for time lost during the lockdown but there may be a small possibility that we may not be able to publish 55,000 occurrence records as initially proposed. We've also had to make adjustments to our project activities that has been mentioned above and through a separate budget amendment request.

GBIF