BIFA4 – Data mobilization grant

MID-TERM NARRATIVE REPORT

Guidelines on how to complete this report are included in italics.

Use the information included in your BIFA Full Proposal (reproduced in Annex 1 of your BIFA contract) as a baseline from which to complete this report.

Remember that this report will be made available on your project page on the GBIF website and therefore should not include any contact information, unless you have permission from all mentioned in the report that their contact information can be published.

Contents

Project information

Main contact person and role:	Wei Song HWANG, Project Lead
Institution/network/agency affiliation:	Lee Kong Chian Natural History Museum, National University of Singapore
BIFA Project ID:	BIFA4_050
Project title:	Digitizing butterflies of Singapore & Peninsular Malaysia: contributions to the Butterflies of the Southeast Asian Islands Consortium Project
Start date and end date of the reporting period:	01/06/19 - 31/10/19 (Mid-term progress report)
Country in which the activities take place:	Singapore

Executive summary

Provide a brief explanation of the project and its implementation, the context and the approach taken for the mid-term evaluation, and a summary of progress of your project's original objectives. Also include any additional objectives that have been defined since the implementation of the project and explain why they've been added to the project's initial objectives. In the event of unexpected challenges which have prevented you to reach a planned project objective, please provide explanations and indicate how you plan to reach these objectives within the project implementation period and if you would like to receive the help of a specific project mentor. (Max. 350 words.)

This project aims to mobilize ~10,000 museum butterfly specimen housed in the Lee Kong Chian Natural History Museum which represents ~90% of the species present in Singapore and Peninsular Malaysia. Data mobilization for this project involves data transcription of original labels, specimen imaging and georeferencing of individual specimens that were already identified. Funding from BIFA allowed for the hiring of 1 part-time intern for 12 months to assist in the data mobilization process.

Currently we have a total of 3,631 specimens transcribed, which includes unique specimen ID assignment (=OccurrenceID), original locality and identification data. Among the transcribed specimens, 1,893 specimens have completed the image-vouchering process. The most complete set of specimens (725 records) come from Doggett's collection of Hesperiidae (skippers), which also have all localities georeferenced. This most complete set of specimens was further cleaned up to be the first published dataset for this project as required for the mid-term progress report.

The projected average specimen handling time of 20 minutes per specimen turned out to be a challenge to meet due to several factors. Familiarity with specimen handling by interns took on a steeper learning curve, primarily for imaging purposes as some optimization was required to allow for easier specimen and label manipulation/preparation. Due to the delay in getting the part-time intern (Vivian Feng, hired July 2019) on board, the funds now also allow for a second intern to be hired (Qian Yi Ho, hired Sep 2019), with any additional manpower funds to be topped up by the museum collections budget. The operations can now be scaled up with certain division of labour when executing the imaging pipeline. A student volunteer with some familiarity with local butterfly taxonomy has also been recruited and he has offered to commit time from mid-November until the end of the project to assist with the work.

One unplanned positive development that is helping to increase productivity is the use of a machine-learning algorithm by the intern to rename image files automatically by recognizing the assigned unique specimen ID label captured within the image of each specimen. Currently this procedure is being optimized to ensure higher accuracy, with the ultimate aim to save time.

Progress against expected milestones

Please indicate the status of the expected milestones as outlined in Annex 3 of the contract, at the time of mid-term reporting. Please provide links in the sources of verification.

In the event of unexpected delay(s), please provide detailed explanatory notes.

Expected milestones	Completed? Yes/No	Explanatory notes	Sources of verification
Gain certification at BIFA Capacity Enhancement Workshop **(Mid-term report milestone)**	Yes	Obtained Biodiversity Data Mobilization Advance Badge	https://openbadgepassport.com/app/profile/58952
At least one dataset has been published to GBIF.org **(Mid-term report milestone)**	Yes	Hosted by GBIF BIFA IPT	https://www.gbif.org/dataset/cc353435-511d-4ba2-b88d-b874d22c7a0a

Activities

Please indicate the status of the activities as outlined in the BIFA Full Proposal (section 4.4), at the time of mid-term reporting.

The table below should be completed in the same way as in the BIFA Full Proposal, but should include information and updates on the status of each activity. Please indicate relevant source(s) of verification and provide links, or a copy of the source of verification mentioned if no links are available. Attachments should be provided in the Annex.

In the event of unexpected delay(s), please provide detailed explanatory notes and indicate new planned completion date(s) within the project implementation period.

In the event of any additional activities having being completed, please add rows as required.

Activity name	Expected deliverable	Contribution to goals listed in table 4.3 of the BIFA Full Proposal	Status of activity as of mid-term reporting Provide % complete and state either on-track or delayed	Explanatory notes, inc. new planned completion date if necessary	Source(s) of verification
			Example: 50% and on-track		
Register and upload georeferenced butterfly collection data to GBIF. Specimen data and images for Peninsular Malaysia and Singapore based on existing collections.	Registering of 10,000 butterfly specimens, including georeferencing of localities.	Exact match to goal listed in table 4.3	Approximately 36% complete. Slight delay.	A single intern processing specimens and data transcription turned out to be more time-consuming than planned. A second intern is thus hired to scale up the operations now that most of the start-up trouble-shooting and optimizations are resolved. Completion date remains unchanged: 30 April 2020.	Please see budget expenditure on intern hire.

Deliverables

This section should summarize the project deliverables as outlined in the BIFA Full Proposal (section 4,5), at the time of mid-term reporting. Please highlight any changes from the original plans.

If no result has been achieved on a specific point, please clearly indicate this and provide expected date of completion.

In the event of unexpected delay(s), please provide detailed explanatory notes and indicate new planned completion date(s) within the project implementation period.

In the event of any additional deliverables having being completed, please add rows as required.

a. Data

Details of datasets published and/or pending publication as an outcome of the project. The table below should be completed in the same way as in the BIFA Full Proposal but should include information and updates on the status of each dataset. Please provide links in sources of verification.

If the dataset is not yet published, please provide a detailed explanation, % of digitization and expected date of publication. Add rows as required.

Title of dataset	Data type (checklist/occurrences/sampling event) [1]	Estimated number of records (specimens)	Status of dataset: Published or % of digitization/pending publication and expected date of publication	Data holding institution agreed to publish its data via GBIF.org (Yes/No)	Explanatory notes	Source(s) of verification: DOI or URL

---

[1] Dataset classes: http://www.gbif.org/publishing-data/summary#datasetclasses

Title of dataset	Data type (checklist/occurrences/sampling event) [1]	Estimated number of records (specimens)	Status of dataset: Published or % of digitization/pending publication and expected date of publication	Data holding institution agreed to publish its data via GBIF.org (Yes/No)	Explanatory notes	Source(s) of verification: DOI or URL
			Examples: *Published* *or* *50% digitized, expected date of publication January 2020*			
Butterflies_of_Singapore_and_Peninsular_Malaysia	Occurrences	725	Published	Yes	Occurrence data with georeferences added. Image files not yet associated in first version.	https://www.gbif.org/dataset/cc353435-511d-4ba2-b88d-b874d22c7a0a

b. Other deliverables

Describe other deliverables (e.g. publication of data papers, analysis, reports etc.) produced and/or planned to be produced/completed as part of the project. Please provide percentage complete and status of deliverable, together with an indicative dates/estimated time for completion.

Please provide links in the sources of verification. Attachments should be provided in the Annex.

Name and type of deliverable	Status of deliverable Provide % complete and on-track/ delayed	Explanatory notes	Source(s) of verification
	Example: 50%, on-track	*Example: expected to be completed January 2020*	
NA			

Events organized as part of the project

Please indicate the status of the events as outlined in the BIFA Full Proposal (section 4.6), at the time of mid-term reporting.

The table below should be completed in the same way as in the BIFA Full Proposal, but should include information and updates on the status of each event. Please provide links to any documents or web pages documenting the event(s) in the sources of verificaiton. Attachments should be provided in the Annex

In the event of unexpected delay(s), please provide detailed explanatory notes and indicate new planned completion date(s) within the project implementation period.

In the event of any additional events having being completed, please add rows as required.

Event title	Organizing institution	Date held / expected dates	Number of participants	Explanatory notes	Source(s) of verification
NA					

Calendar of activities

The calendar should be completed in the same way as in the BIFA Full Proposal (section 4.7) but should also clearly indicate and include any changes (e.g. use of colour-coding to indicate expected changes and/or delays). Please provide explanations for any changes in the "Notes" column and/or the explanatory notes section.

Activity	2019							2020				Notes
	M	J	J	A	S	O	N	D	J	F	M	A	
Activity: Register and upload georeferenced butterfly collection data to GBIF	■	■	■	■	■	■	■	■	■	■	■	■	Original activity listed in BIFA Full Proposal (section 4.7)
Mandatory attendance of a project team member to the BIFA Capacity Enhancement Workshop (expenses for attending the workshop are covered by the BIFA programme)			X										
Deadline for mid-term report – 31 October 2019 Mandatory milestones attached to the mid-term report: - Gain certification at BIFA Capacity Enhancement Workshop - At least one dataset has been published to GBIF.org						X							
Activity: Specimen Imaging			X	X	X	X	X	X	X	X	X		Revised activity timeline

Activity	2019								2020				Notes
	M	J	J	A	S	O	N	D	J	F	M	A	
Activity: Specimen label transcription and georeferencing			X	X	X	X	X	X	X	X	X		Revised activity timeline
Activity: Specimen data cleaning and processing for uploading as GBIF datasets												X	Revised activity timelinedf
Deadline for final reporting – 30 April 2020 Mandatory milestones attached to the final report: - All deliverables have been produced - Sustainability and next steps have been documented												X	

a. Explanatory notes

A more detailed breakdown of the work processes involved in the digitization workflow is provided here to improve visualization and planning. The decoupling of the transcription and georeferencing operations from the imaging operation was one key difference to how the operations has been optimized. A final data cleaning and editing phase is also included for the final month in preparation of full submissions of datasets for uploading onto GBIF. Although the amount of work may appear to be more, the addition of one more intern allows for work to be split and operations can also scale up for the imaging procedure.

Project communications and visibility

Describe the plans and how the results of your project have been and will be communicated and shared with the project stakeholders and broader GBIF community.

Please also review the page describing your project available from http://www.gbif.org/programme/bifa and highlight any additional documents, events, news items or links that you would like to add to your page and provide links/attachments in the Annex.

Sharing of the project has been built into the project implementation from the start. We are documenting each of the steps taken to build up the project such that a protocol can be shared with the butterfly collection community via a journal publication or similar means. Results from the project will be fed through our museum's established communications channels for wider public consumption, such as via our museum webpage (http://www.lkcnhm.nus.edu.sg), facebook page (https://www.facebook.com/lkcnhm/), and twitter account (https://twitter.com/lkcnhm). Our university's communications office can also be tapped to reach out to the local news media outlets. This publicity plan has not changed and we are on track to have the material available for use.

Internally within the museum, a training workshop will be arranged to share among the collections staff the aims and goals of GBIF and BIFA and how our project was conducted. This was conceptualized after the July BIFA engagement meeting/training workshop as there are clearly many benefits for the museum's collection staff to adopt the common practices and encourage to contribute towards the GBIF community.

Mid-term evaluation findings and recommendations for the remaining project implementation period

This section should cover for example:

- *An evaluation of the project activities by mid-term and their outputs/deliverables*
- *Comments on the project implementation, and its efficiency and effectiveness, strength and weaknesses etc.*
- *Any feedback on the project's relevance from the partners and stakeholders*
- *Indications and reasons for any changes which have been made to the project's original plans, and actions to follow-up*
- *The management arrangements for the project, including support from the GBIF Secretariat*
- *Areas of success to build on during the remaining project implementation period*
- *Any request for guidance and/or assistance in the form of a specific project mentor*

   (Max. 300 words)

The project got off to a rocky start with administrative delays that led to a lag in results being produced. Now that the operations have reached a steady-state after multiple rounds of optimization, and we have the means to scale up with an additional crew member, the output should increase to a rate that will meet the final deliverables. Having reached 36% transcription rate within ~3.5 months of data capture work by one part-time intern, the remaining 6 months with 2 interns should get us to the minimal target output.

The project implementation thus far has been a valuable learning journey for all members of the team and the lessons learned are being documented for the purpose of sharing with the community. Similarly, we have benefited from resources shared by earlier curators performing digitization projects, although some effort was required to tweak steps to make

feasible in our local collection context. Management of the team thus far has focused on making sure the personnel have the appropriate resources to carry out the project deliverables in terms of workspace, equipment, toolkits, specimens etc. For example, acquiring the right equipment for imaging different sizes of butterflies are more complex than initially planned. Thankfully, the interns hired have been very resourceful and able to work independently to solve a number of teething issues. The development of the file-renaming algorithm using image capture technique is a novel offshoot that will hopefully increase efficiency and become a worthy tool to share with the community.

The second half of the project timeline will be focusing on churning out the data in as smooth a pipeline as possible. Being a lean team of 3 (+1 incoming volunteer in mid-November), communications thus far has been effective and will leverage upon the strengths of each team member to help reach the target goals. We hope this project will be a good showcase of what we can achieve with a small outfit and limited resources.

Thus far, we have received very timely and helpful support from GBIF and BIFA staff and we will continue to engage them as we fine tune and polish up our datasets to be published.

Sustainability plans

Please provide a description of how the partners involved will build on the results of this project in their future work. This could include future collaborative activities, such as plans to complete any unfinished project activities and how the future impact of the project could be monitored and/or measured. (Max. 200 words).

The museum's commitment to digitize our collections has not changed since the full proposal. Our newly acquired collections management system is currently in the development phase and is slated to be operational in May 2020. The data generated from this project will feed very well into the museum collection system and the future plan is to help improve compatibility between the museum system and GBIF publishing criteria. The successful completion of this project will be useful to promote the cause. Future extensions for further digitization of the butterfly collection and rest of the insect collection will be supported by the museum and future sourcing of funds.

The consortium project led by Naturalis will also be building upon the datasets here to collate a more comprehensive species inventory list for the whole of Southeast Asia. Plans has not changed since the full proposal but an updated appraisal will be available at a later date.

Annex – Additional sources of verification

Sources of verification are for example links to relevant digital documents, news/newsletters, brochures, workshop related documents, pictures, etc.

First published dataset

https://cloud.gbif.org/bifa/resource?r=butterflies_of_singapore_and_peninsular_malaysia

Biodiversity Data Mobilisation, Advanced Badge

https://openbadgepassport.com/app/badge/info/162347

GBIF regional engagement meeting and data mobilization workshop in Asia

https://www.gbif.org/event/uE5VYX0rEqMYImhfft0nZ/gbif-regional-engagement-meeting-and-data-mobilization-workshop-in-asia

Presentation at the GBIF regional engagement meeting and data mobilization workshop in Asia:

Digitizing butterflies of Singapore & Peninsular Malaysia: Contributions to the Butterflies of Southeast Asian Islands Consortium Project

https://assets.ctfassets.net/uo17ejk9rkwj/4Xvo0iTKkDbp42Me45Zpue/5b576c8d3dc0410173eb5ed15f8ec127/13-BIFA4_050_Singapore_butterflies_HWS.pdf