FINAL 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, PI**
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 - 31/08/20**
Country in which the activities take place:	**Singapore**

Executive summary

Please provide a brief summary of your project's objective(s), its implementation and achievements. You should also describe the context and the approach taken for the final evaluation of your project, and also the key best practices and lessons learnt identified. Remember to include any additional objectives that were defined during the implementation of the project and explain why they were 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 if and how you plan to reach these objectives post project.

(Max. 350 words.)

This project aims to mobilize 10,000 museum butterfly specimens 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. Co-funding from the museum allowed for an additional 3 part-time interns to join the team at different stages of the project. The Covid-19 pandemic disrupted operations at the start of 2020, but the 4-month extension granted allowed us to catch up and achieve the project goals.

Final outcome of the project has 12,702 specimens transcribed, 25,245 images taken, 1,435 species and subspecies recorded, 187 unique localities georeferenced. All specimen data have been packaged and published as 6 datasets on GBIF portal, each representing one butterfly family. Project outcomes were shared at GBIF Asia Virtual Summit 2020, within the museum community, and also timed for public release in September 2020.

Execution of the project relied on a high level of automation given the large quantity of files to process and lean staffing. This includes implementing an OCR algorithm for image file re-naming, batch processing to resize, watermark and redact specimen images. A decision was made to withhold locality information for species identified as rare or very rare for conservation protection reasons, which led to additional post hoc redaction of locality information in almost one-third of the species present. This exercise alerted us to the fact that a significant proportion of the butterfly species already has a restricted distribution and low in population numbers by the end of the last millennium, and their current status is a cause of concern.

Progress against expected milestones

Please indicate the status of the expected milestones as outlined in Annex 3 of the contract, at the time of final 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
All deliverables have been produced **(Final report milestone)**	Yes	All datasets published on GBIF portal	https://www.gbif.org/project/7msmpErXC8uaGYzhJpyHoX/digitizing-butterfly-specimens-of-singapore-and-peninsular-malaysia#datasets
Sustainability and next steps have been documented	Yes	Impact of project on museum digitization	Participation of museum staff involved in other digitization projects in sharing session and training workshop. https://bit.ly/2EH50ro https://bit.ly/2YOcUXi

Expected milestones	Completed? Yes/No	Explanatory notes	Sources of verification
(Final report milestone)		effort is positive	

Activities

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

Please include any additional activities completed during the implementation of the project that were not originally outlined in the BIFA Full Proposal.

Add rows as required.

Activity name	Expected deliverable	Contribution to goals listed in table 4.3 of the BIFA Full Proposal	Status of activity at final reporting: Completed? Yes / No (inc. % complete)	Explanatory notes, including new planned completion date if necessary	Source(s) of verification
			Example: *No (90% complete)*		
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	Yes (127% Complete)	Datasets complete and published, with locality information withheld for rare/very rare species. Images as Associated Media prepared but undergoing redaction of label images before publishing in next version.	https://www.gbif.org/project/7msmpErXC8uaGYzhJpyHoX/digitizing-butterfly-specimens-of-singapore-and-peninsular-malaysia#datasets

Deliverables

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

If no result has been achieved on a specific point, or in the event of unexpected delay(s), please provide detailed explanatory notes and indicate new planned completion date(s).

Please include any additional deliverables completed during the implementation of the project, that were not originally outlined in the BIFA Full Proposal.

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 indicate this as "not published" and 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 / not published and % of digitization 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
Butterflies_of_Singapore_and_Peninsular_Malaysia (Hesperiidae)	Occurrences	2,613	Published (100%)	Yes	Image files not yet associated in first version, undergoing redaction.	https://doi.org/10.15468/rcghqm
Butterflies_of_Singapore_and_Peninsular_Malaysia (Riodinidae)	Occurrences	192	Published (100%)	Yes	Image files not yet associated in first version, undergoing redaction.	https://doi.org/10.15468/rw7ppj
Butterflies_of_Singapore_and_Peninsular_Malaysia (Papilionidae)	Occurrences	481	Published (100%)	Yes	Image files not yet associated in first version, undergoing redaction.	https://doi.org/10.15468/r3h2zy
Butterflies_of_Singapore_and_Peninsular_Malaysia (Pieridae)	Occurrences	771	Published (100%)	Yes	Image files not yet associated in first version, undergoing redaction.	https://doi.org/10.15468/pmqfpa
Butterflies_of_Singapore_and_Peninsular_Malaysia (Nymphalidae)	Occurrences	3,438	Published (100%)	Yes	Image files not yet associated in	https://doi.org/10.15468/pjebbq

[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 / not published and % of digitization 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
					first version, undergoing redaction.	
Butterflies_of_Singapore_and_Peninsular_Malaysia (Lycaenidae)	Occurrences	5,207	Published	Yes	Image files not yet associated in first version, undergoing redaction.	https://doi.org/10.15468/7446rf

b. Other deliverables

Describe other deliverables (e.g. publication of data papers, analysis, reports etc.) produced by the project. In the event of unexpected delay(s), please provide detailed explanatory notes and indicate if the deliverable(s) are to be produced post-project, and if so, indicate expected completion date(s).

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

Name and type of deliverable	Status of deliverable: Completed? Yes / No (inc. % complete)	Explanatory notes	Source(s) of verification
	Example: Not completed, 90% complete.	*Example: expected to be completed September 2020*	
Data paper	No (30% complete)	Expected to be completed October 2020, not in original proposal.	https://drive.google.com/file/d/14cpybDd8UG_MTN28mYgmKcRBQGBx0zoo/view?usp=sharing

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 final 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 verification. 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).

Please include any additional events organized as part of the project that were not originally outlined in the BIFA Full Proposal.

Add rows as required.

Event title	Organizing institution	Date held / expected dates	Number of participants	Explanatory notes	Source(s) of verification
Museum Sharing Session: Float like a butterfly, Adventures in digitising our Lepidoptera collection	LKCNHM	25th Aug 2020, 3-4pm	32	Internal sharing among museum staff and volunteers	https://bit.ly/2EH50ro Recorded Zoom session
Museum Training Workshop: Tools that can make your databasing life easier	LKCNHM	31st Aug 2020, 3-5pm	14	Internal sharing among museum staff and volunteers	https://bit.ly/2YOcUXi Recorded Zoom session

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. Please provide general comments on the calendar of your activities in the 'General comments on project implementation' section.

Activity	2019								2020								Notes
	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	
Activity: Register and upload georeferenced butterfly collection data to GBIF	⬛	⬛	⬛	⬛	⬛	⬛	⬛	⬛	⬛	⬛	⬛	⬛	🟥	🟥	🟥	🟥	Original activity listed in BIFA Full Proposal (section 4.7), project extension in red.
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	X	X	X	X	X	Revised activity timeline from mid-term report
Activity: Specimen label transcription and georeferencing			X	X	X	X	X	X	X	X	X	X					Revised activity timeline from mid-term report

Activity	2019							2020								Notes
	M	J	J	A	S	O	N	D	J	F	M	A	M	J	J	A	
Activity: Specimen data cleaning and processing for uploading as GBIF datasets													X	X	X	X	Revised activity timeline from mid-term report
Activity: Museum Sharing Session																X	
Activity: Museum Training Workshop																X	
Deadline for final reporting – 31 Aug 2020 Mandatory milestones attached to the final report: - All deliverables have been produced - Sustainability and next steps have been documented																X	Publication of 6 datasets on GBIF portal

a. **General comments on project implementation**

(Max. 200 words)

Despite initial administrative delays and the onset of the Covid-19 pandemic, this project was able to overcome challenges encountered to achieve its objectives thanks to the extension granted by the GBIF BIFA managing committee. Relying on a lean crew to carry out this project required a lot of forward planning and the campus-wide ever-changing work schedules as the pandemic evolves means ongoing adjustments even to the end of the project. The high levels of automation implemented at various stages of the specimen imaging process greatly enhanced our capabilities to capture as much specimen images as possible. Constant optimization of work flow

enabled best resource utilization at each stage of the project. Upon completion of the datasets, a decision to withhold sensitive locality information for conservation reasons requires some re-working of both the core dataset as well as the images associated. Once protocols were set up and optimized for each task, production of data was smooth and efficient, and the achievement of mobilizing more specimens than specified was the outcome.

Project communications and visibility

Describe the way the results of your project have been and will continue to be communicated and shared with the project stakeholders and broader GBIF community.
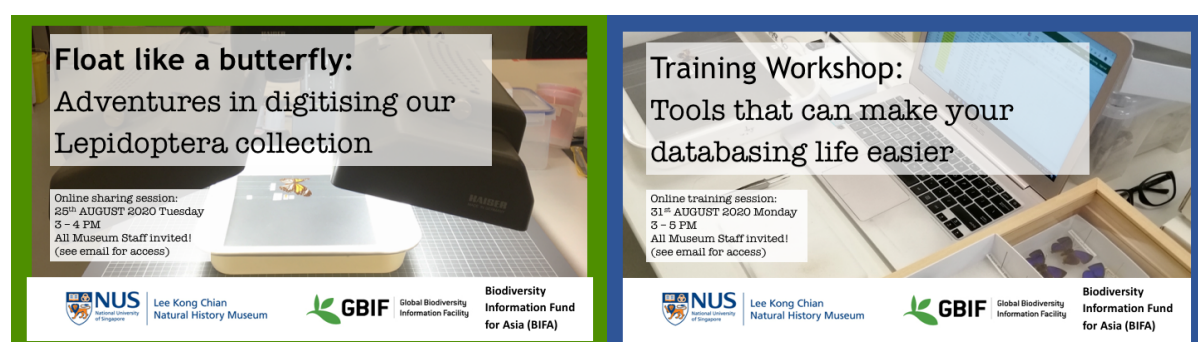
Please also review the page describing your project available from http://www.gbif.org/programme/bifa and highlight any additional documents, events, news items or links that you would like to add to your page and provide links/attachments in the Annex.

(Max. 300 words)

Sharing of the project has been built into the project implementation from the start. 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/), twitter account (https://twitter.com/lkcnhm) and Instagram account (https://www.instagram.com/lkcnhm/). An Instagram story media release is slated in September 2020 to showcase this project (Draft examples https://drive.google.com/drive/folders/1RMsmr0va9P5qnAMQ8kRwufw6rGDk85oJ?usp=sharing).

This will be followed by a museum blog post describing the behind-the-scenes to mobilize the butterfly specimens (drafting stage). We have reached out to our university's communications office and the local news journalists to communicate the results of this project to the general public.

Internally within the museum, a sharing session and a training workshop has been conducted to share among the museum staff and volunteers the aims and goals of GBIF and BIFA and how our project was conducted, together with a quick tutorial on some useful tools for data cleaning. 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. We have documented 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 (drafting stage).

Monitoring, evaluation and lessons learned

Please indicate the results of the monitoring and evaluation plans, as outlined in your BIFA Full Proposal (section 4.9), at the time of final reporting.

Ideally this section should also include the following:

- *An evaluation of the project activities and their outputs/deliverables*
- *An assessment of the overall outcomes, impacts of the project and how it contributes to the overall objective of the BIFA programme*
- *Comments on the project implementation and completion, 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*
- *Areas of success to build on, after the project's implementation period and best practices that could be useful for the community*
- *Main lessons from the project experience that could be applied in other contexts*
- *Conclusions from your experience during the implementation of the project and recommentations for the GBIF Secretariat or the community to reinforce the success of the project*

(Max. 400 words)

The weekly monitoring of specimen output to evaluate project progress as stated in the BIFA Full Proposal was effective in addressing issues encountered and optimizing processes along the way. Once the protocols were established, churning out the products was a matter of having sufficient manpower and time. Working with part-time interns requires flexibility in time management and irregular workable hours are to be expected on a weekly basis. At the later stage of the project this February, due to control measures instituted by the university to minimize chances of Covid-19 infection, various work schedules were implemented that were disruptive to our pipeline, to the extent of temporary cessation due to work-from-home orders from end April to beginning of June 2020. Recovery thereafter was slow as there was a limitation of staff allowed back into the museum to work so work schedules had to be staggered in various ways. The project deadline extension allowed for the make-up of lost time due to the abnormal disruptions and the project objectives were eventually achieved with some additional results.

The outcome of the project is that we now have a virtual collection of these butterfly specimens. These datasets with georeferenced localities and high-resolution images will be valuable for the scientific community to study variation across spatial and temporal gradients within the geographic region. Given the fragility of the actual specimens, this virtual collection has made itself more accessible for scientific studies, having a multiplier effect that is hard to match on the ground. We hope that our contribution to GBIF here is a valuable addition in terms of species representation coming from Southeast Asia.

The strength of the project lies in the adoption of automation in helping to process the images which was a huge time-saver. Having a dedicated team to work on this project solely was essential to minimize time management issues. Weaknesses in this project includes the lack of a ready-to-go set of equipment, which required additional initial effort to set up and optimize before we could scale up the processes. The manpower needs of the project was underestimated, and although difficult to find interested candidates, we were able to boost up the personnel on the team to four eventually, with the support of the museum co-funding the additional staffing. This was helpful especially during the disruption period due to Covid-19. Documentation of best practices, experience gained here, will be shared within the museum and wider scientific community, and contribute towards the planning of the next data mobilization project.

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 new collections management system has become operational during the interim of this project. We were able to align the fields in the system with Darwin Core terminology which is useful when exporting specimen data that can be published as GBIF datasets.

The museum has been very supportive of this project and the senior management has expressed strong satisfaction with the final deliverables obtained. Experience gained from this project will influence how future digitization projects be conducted within the museum, with an additional desire to make them into publishable GBIF datasets whenever possible. Impact of the project can be measured by the number of such datasets coming out from the museum in the future.

Annex – Additional sources of verification

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

Direct links to online resources have been shared throughout this document as sources of verification. Some of these are draft folders which may be deactivated when task is completed.