FINAL ACTIVITY REPORT

Guidelines on how to complete the activity report are included in italics.

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

Contents

Project information

Main contact person and role:	***Sabhrina Gita Aninta, Project Leader***
Institution/network/agency affiliation:	***Tambora Muda Indonesia***
BIFA Project ID:	***BIFA3_23***
Project title:	***The Biodiversity Theses Database***
Start date and end date of the reporting period:	***1 April 2018 and 31 August 2019***
Country in which the activities take place:	***Indonesia***

Executive summary

Provide a brief explanation of the project and its implementation, the context and the approach taken for the final evaluation, and a summary of the objectives achieved, lessons learned and conclusions..

Student theses provide a vast untapped source of biodiversity information in Indonesia as biodiversity assessment has been a common topic for research for undergraduate students in Indonesian universities. The Biodiversity Theses Database project—Biodiverskripsi—is a project that aims to collate biodiversity data originating from student theses and make it available to wider community. In this final report, we evaluate Biodiverskripsi against several objectively verifiable indicators that we planned on providing on our project proposal. These are mainly the deliverables of the project, such as the existing sampling-event data resulting from student theses, data enumeration protocol, catalogue of transcribed theses, custom web-based information system, and manuscript draft for the planned data paper. Additional reflection on project implementation, feedback from stakeholder, and overall evaluation narrative of project activities and outputs are also provided along with lessons learned for mobilizing this type of biodiversity data.

Biodiverskripsi has currently published ~14,000 occurrence data from 1,330 sampling events thanks to a team of dedicated volunteers from Tambora Muda Indonesia. A considerable number of these occurrence data contains fuzzy taxonomy, which we believe due to lack of updated taxonomy of most unpopular taxa. These occurrence data were also available in the current custom web-based application (biodiverskripsi.org) which we made in Indonesian language to promote biodiversity data mobilization to wider public. The portal was soft-

launched in 24th of March 2019 during our closing conference to get feedback from wider public. The data paper to assess the state of Indonesian student research was not yet submitted as we encounter challenging issues in the technical validation stage of our data. Publishing a reliable data set of species occurrences from secondary literature required a different type of validation steps unfamiliar to most experts working in taxonomy.

Despite the adjustments, all the planned activities are well-implemented and our closing conference has opened up exciting new opportunities for our new community of biodiversity informatics enthusiasts. The effort of mobilizing biodiversity data from student theses will likely continue due to successes in our project. In general, we came to an improved understanding towards biodiversity data management in Indonesia and able to recommend a more efficient and effective project management to mobilize biodiversity data from grey literature through volunteering scheme.

Project objectives

This section should include the list of objectives included in your original project proposal, stating for each how far you advanced towards their achievement. Also include any additional objectives that were defined during the implementation of the project. In the event of undexpected challenges prevented you to reach a planned project objective, please provide detailed explanations and indicate how you plan to reach these objectives post project.

The objective of The Biodiversity Theses Database project was to **make biodiversity-related researches produced by early research training in Indonesian higher education become integrated nationally and form a substantial source for global biodiversity research**. In achieving this objective, we assigned three major goals:

1. Ecological monitoring data from student theses of at least five Indonesian universities produced in 2000-2017 collated in the form of sampling-event data as an integrative format that could be used internationally.

   We have currently collated ~20,000 occurrence data from 1684 sampling events from 270 theses coming from 6 universities. Much of the collated occurrence data was still prone to non-standardized entries, fuzzy taxonomy, and invalid coordinates. Thus, at the end of project term on 31st of August 2019, we still only allow ~14,000 occurrence data from 1,330 sampling events in the GBIF and our data portal. We expect to publish the updated version of data set with improved quantity and quality as soon as possible and continuously afterwards whenever we have new contributors.

2. A sustainable platform to assist the process of sharing and publishing biodiversity data from student theses in sampling-event data format created to fit the need of Indonesian scientific community.

We managed to deliver the web-based information system as an independent platform within the project timeline online in biodiverskripsi.org. This portal was developed in Indonesian language and constantly evaluated for user experience. Making public aware on the existence of our data portal and increased their experience became our post-project activities to ensure more data usage and contribution.

3. The states of quality, discovery, accessibility, and archiving of student theses conducting biodiversity assessment produced in 2000-2017 from at least five universities from all over Indonesia compiled.

   We have not yet completely assessed the state of research in Indonesian student theses as the occurrence data has not been thoroughly validated. We intend to continue assessing the state of research, but this require additional awareness program as we came to realize the low familiarity of students to species database, even GBIF.

Activities

Please indicate the status of the activities as outlined in the project proposal, at the time of the final report. The table below should be completed in the same way as in the full proposal but should include information and updates on the status of each activity.

In the event of unexpected delay, please provide detailed explanatory notes and indicate planned completion date after the end of the project. Add as many rows as needed.

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

Description of activity	Partners involved	Contribution of activity to goals listed in table 4.3	Status of activity as of final reporting Completed? Yes/No	Explanatory notes, inc. planned completion date if necessary	Sources of verification
Mobilizing species occurrence and sampling-event data from observation networks and monitoring systems					
Kickoff meeting with at least 10 universities and	Tambora	Promote project and increase data input (Goal	Yes		See Annex 1.1.

other stakeholders		1 and 2)			
Theses bibliography sorting	Tambora	Assisting data enumeration (Goal 1)	Yes		See Annex 1.2.
Meeting with experts to discuss review mechanism	InaBIF, Tambora	Addressing the problem of various data quality (Goal 1)	Yes		See Annex 1.3.
Develop data enumeration protocol guidebook	InaBIF, Tambora	Assisting data enumeration (Goal 1)	Yes		See Annex 1.4.
Workshop on data enumeration	InaBIF, Tambora	Assisting data enumeration (Goal 1)	Yes		See Annex 1.5.
Data enumeration	Tambora	Ecological monitoring data available (Goal 1)	Yes		See Annex 1.6.
Preparing data papers					
Workshop for analyzing data from data portal	InaBIF, Tambora	Increase data use (Goal 2)	Yes		See Annex 2.2.
Data paper preparation and submission	InaBIF, Tambora	Evaluate data set (Goal 3)	Partially-completed	The technical validation part of the data paper manuscript was incomplete, we expect to submit within 2019-2020	See Annex 3.1.
Other activity types					

Construction of custom web-based information system	InaBIF, Tambora	Create sustainable platform for data sharing and publishing (Goal 2)	Yes		biodiverskripsi.org , See Annex 2.1.
Evaluation of custom web-based information system	InaBIF, Tambora	Create sustainable platform for data sharing and publishing (Goal 2)	Yes		See Annex 2.3.
Closing conference	InaBIF, Tambora	Increase data use and evaluate project (Goal 2 and 3)	Yes		See Annex 3.2.
Communicate project to national audience for wider awareness	InaBIF, Tambora	Increase data use and evaluate project (Goal 2 and 3)	Yes	.	See Annex 3.3.

Deliverables

This section should summarize the project deliverables completed by the final reporting date , with a description of the associated outputs. Please highlight any changes from the original plans provided in the full project proposal.

In the event of unexpected delay, please provide detailed explanatory notes and indicate planned completion date. Add as many rows as needed.

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

a. Data

Details of datasets mobilized and/or pending mobilization as an outcome of the project: Please use list from mid-term report and update this as at final reporting.

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

Title of dataset	Taxonomic/geographic scope	Approximate number of records (specimens)	Current format (e.g. undigitized, digitized)	Status of dataset: Published/not published – inc. date/expected date of publication	Explanatory notes	DOI or URL
Higher-Plants of Indonesia	Embryophyta/Indonesia	14,624	Digitized	Published – 27/03/2019	These datasets have been available on IPT since 26/03/2019 as one dataset and this is more valid as we collected them using the same transcription method. The partitions were made on	http://ipt.biologi.lipi.go.id/resource?r=biodiverskripsi
Vertebrates of Indonesia	Chordata/Indonesia		Digitized			
Marine Life of Indonesia	Echinodermata, Hemichordata, Cnidaria,		Digitized			

	Ctenophora, Mollusca, Porifera, Arthropoda, Mollusca, Actinopterygii, Sarcopterygii/Indonesia				the original proposal and mid-term report due to our misunderstanding of the nature of sampling event data. The updated data sets will be published within a few weeks after the end of the project term and sustainably updated with increasing number of contributors in the future.	
Freshwater Invertebrates of Indonesia	Phytoplankton, Zooplankton, freshwater Arthropods, Ecdysozoa, Lophotrochozoa/Indonesia		Digitized			
Terrestrial Invertebrates of Indonesia	Arthropoda, Annelida, Mollusca, Nematoda/Indonesia		Digitized			

b. Other deliverables

Describe other deliverables (e.g. publication of data papers, catalogues, reports etc.) produced and/or planned to be produced/completed as a post-project deliverable. Please provide indicative dates/estimated time for completion for planned post-project deliverables.

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

Name and type of deliverable	**Status of deliverable** **Published/not published – inc. date/expected date of publication or estimation of time for completion**	**Explanatory notes**	**Source(s) of verification**

Data Paper	In preparation	We have prepared the manuscript but it is not sufficient to get the receipt of article processing charge before 31st of August 2019. This will be a post-project deliverable to increase data usage.	
Transcription Guidebook	Published – 23/03/2019	Written in Indonesia	https://biodiverskripsi.org/pdf/Ikut-Kontribusi.pdf
Catalogue of Transcribed Theses	Published – 23/03/2019	Available as searchable metadata in biodiverskripsi.org	See Annex 1.2.
Infographic	Published – 07/2019	Available as few posts in our social media in Facebook, Instagram, and Twitter	See Annex 3.3.
Video of Project Progress	Not Published	We could not manage all materials visually documenting the transcription process and made a video out of it before 31st of August 2019. This will be a post-project deliverable to increase potential users' awareness on our data portal.	

Calendar of activities

The calendar should be completed in the same way as in the Full Project Proposal (4.6) but should include any changes. Please provide reasons for any changes in the Notes column in the table below.

Date	Event	Lead partner	Notes
April-May 2018	Kickoff meeting with at least 10 universities and other stakeholders	Tambora Muda Indonesia	Completed in 2nd of July with 5 universities due to long national leave term on May 2018.
May 2018	Theses bibliography sorting	Tambora Muda Indonesia	Completed in March 2019 and still continues as post-project activity as additional theses contributors were identified.
May 2018	Meeting with expert to discuss review mechanism	Tambora Muda Indonesia	Completed in August 2018 along with data enumeration workshop to train volunteers to review transcribed data, followed by several personal communications afterwards
May-June 2018	Develop data enumeration protocol guidebook	Tambora Muda Indonesia	Completed in March 2019 and still continues as post project activities as issues with transcribing data from student theses continues to emerge.
June 2018	Attendance of project team member at BIFA Capacity Enhancement Workshop	Tambora Muda Indonesia	Completed on schedule
June 2018	Workshop on data enumeration	Tambora Muda Indonesia	Completed in August 2018, corresponding to the delay in meeting universities.
July 2018-March 2019	Communicate project to national audience for wider awareness	Tambora Muda Indonesia	Completed in August 2019, corresponding to the delay in constructing web-based information.
July 2018	Construction of custom web-based information system	Tambora Muda Indonesia	Completed in March 2019, corresponding to the delay in

			obtaining first few data sets from the transcription.
August 2018	Evaluation of custom web-based information system	Tambora Muda Indonesia	Completed in March 2019, corresponding to the delay in constructing web-based information.
June 2018-August 2019	Data enumeration	Tambora Muda Indonesia	Completed in August 2019, but will be still continued by interested volunteers as post-project activity as additional theses contributors were identified
October-December 2018	Data paper preparation and submission	Tambora Muda Indonesia	Partially completed. We are still waiting for data cleaning and validation before submitting the manuscript on around 2019-2020.
October 2018	Workshop for analyzing data from data portal	Tambora Muda Indonesia	Completed in December 2018 using GBIF portal after we knew that the portal could not be finished in October 2018
March 2019	Closing conference	Tambora Muda Indonesia	Completed on schedule

a. General explanatory notes

We have completed all activities within the extended project term except for data paper submission. Data paper submission was postponed due to partially completed manuscript as the technical validation part of the manuscript was not finished.

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. Highlight any additional documents, events, news items or links that you would like to add to your page and provide links/attachment in the Annex..

This project has been continuously communicated by social media accounts of Tambora Muda Indonesia which are followed by thousands of conservationists and biological conservation enthusiasts including our stakeholders (Facebook, followed by 3,419 accounts per 27/08/2019, and Instagram, followed by 7,568 accounts per 27/08/2019) and the website of Tambora Muda Indonesia (ranked #342,299 nationally) in Indonesian language. The project page assigned by GBIF has helped us communicate our project to stakeholders and any interested parties in English, and it has also increased the accountability of our work.

Final evaluation findings and conclusions

This section of the report should cover for example:

- *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*
- *The management arrangements for the project, including support from the GBIF Secretariat*
- *Areas of success to build on, after the project's implementation period*
- *Conclusions from your experience during the implementation of the project*

Evaluation of project activities and their output

In the following sections, we evaluate project goals according to the implemented activities and describe what changes were made to the original plans along with areas of success to build on.

Ecological monitoring data collected

From of our first goal of collating sampling-event data sets from at least five universities across Indonesia, we managed to get around 14,043 occurrences from 1,330 sampling events ready for research use in GBIF. This data set still has several issues related to geographic coordinate, taxonomic information, and few non-standardized entries nonetheless. We have been working to improve the data quality by further cleaning and validating the data post-project. As most of our team members are working part-time in the Biodiverskripsi project or voluntarily as part of Tambora Muda Indonesia, the time-frame for data cleaning and validation was constantly adjusted.

We overestimated the capacity of our volunteers and operational team as the number of data issues build up with additional data contribution. Managing a pipeline of data transcription, cleaning, and validating that involve a number of volunteers who work remotely

from each other has its own complexity and challenges. Although we have transcribed additional theses so that we could update our published data up to ~20,000 occurrences in GBIF, we have not done so as we have not yet fully addressed the aforementioned data issues. Only by redistributing tasks and redelegate some activities among each other, our volunteer-based community project could sustainably run.

The transcription protocol guidebook to assist the transcription quality was in place, but we still need more inputs on how to make the guide as user-friendly as possible. We have difficulties in assuming the level of knowledge of people who wants to contribute their theses towards our portal and some terms may be unfamiliar, considering the small evaluation we conducted during our workshops. According to the volunteers' feedback, we assume too much familiarity on the topic of Darwin Core and taxonomy from the volunteers and assign less time than needed. From our workshops and closing conference, we found that most students who had participated was not familiar with using a web-based data portal to obtain species occurrence data. We also overestimated the experts' familiarity with sampling-event data and Darwin Core, led to the hesitance of several invited Indonesian taxonomists to be a part of the validation mechanism.

In addition to improvement in the data collation pipeline, we suggest that data mobilization projects need to invest wisely on time in establishing a sustainable collaboration and additional research on data sharing culture in Indonesian universities. Due to unfamiliarity to our project's impact, we only manage to sign MoUs with three partner universities instead of five according to our target by the end of our project term. Nonetheless, we took a great care in accommodating the three university partners and get two groups of independent researchers join our project to enable us getting occurrence data from a total of five universities' worth. We also decided to prolong the transcription process in one university which have more difficulties maintaining constant contribution by using the remaining budget previously assigned for additional collaborating universities.

Create sustainable platform for data sharing and publishing

Despite the delay, we have launched biodiverskripsi.org as our web-based data portal within the project term and get preliminary evaluation from potential users. Our continuous testing and correspondence with the IT team who mostly not familiar with biodiversity data have resulted in longer time-frame for our data portal development. In constructing a new type of data portal with particular database structure, we suggest an investment to professional database consultant dedicated to the project to speed up the process. We plan to continuously evaluate our data portal post-project and upgraded its features according to users' feedback whenever necessary.

Also, although the evaluation of the database (user experience in inputting data, extracting information, and navigating through the system) was initially expected to be conducted by at least 30 students from each partner university and other relevant stakeholders, maintaining separate remote workshops to evaluate the data portal was challenging and sending online questionnaires only gave us a total of 16 public respondents. We accommodate their feedback nonetheless and took a great care regarding users' familiarity to the data portal. The planned workshop to evaluate data portal was changed into a workshop to analyse data from GBIF to assess students' familiarity with database on species occurrences in general. As this workshop was delayed to December 2018 and InaBIF could not fund anything post November 2018, this workshop was an in-kind contribution from Tambora Muda Indonesia, taking benefit from our reputation and network. The lack of funding did not enable us to invite experts to give their insights so that we focus in taking student's feedback on using a data portal.

The state of student research

The manuscript reporting the evaluation of ecological monitoring data from Indonesian student theses because unfinished data cleaning and technical validation in the resulting dataset is currently in preparation. Our team agree to not only publish a metadata about our dataset but also gave the scientific community more insight about biodiversity data from student theses. Such manuscript took longer time to write, not to mention the problems in the resulting transcription. Most journals, however, concerned with the technical validation of our data and this increases the time for the manuscript reviewing process. Data paper submission will be a post project activity and GBIF will be notified as soon as our manuscript is submitted.

In communicating the state of student research in biodiversity and our project in general, we managed to held an event to promote people's awareness towards our project and our data portal. The event ended up more like an interactive seminar rather than a conference. Adopting the form of "world café", we got the participants also presenting their views on biodiversity data sharing and management. The participants were fewer than we planned, mostly because we could not ensure how many participants will really attend from the ~70 RSVPs. As soon as the number of participants registering to the event was more than 50, we stopped advertising our event. This resulted in 25 confirming participants, which three of them could not attend due to various reasons. We also invite university partners to send their representatives, but one university partner could not send their representative in our closing conference. This absence of representative from one university partner left us with additional

budget to spend, which we used to promote our project further and increase data quality from new volunteers.

We got people from various background participating in our world café sessions, increasing our perspective in how public perceive biodiversity data sharing and management. Most of them said that they would like this event to be held in universities to promote such effort in more universities. In implementing this event, we also built new network with our speakers and potential future collaboration for keeping the sustainability of this project.

Additionally, we continuously update our project progress in our website, tamboramuda.org, and sending updates to our project page, but writing content in two languages has been challenging for our content writer so that we are a bit delayed in updating both our website and our project page in GBIF. We should assign two different persons for different language to ensure communication of updated progress. Additional infographics were made to be shared in social media to further increase public's awareness towards our project and data portal. Posting promoted posts during the extended project term in our account increased public's awareness on our project so that we have various inquiries regarding contribution mechanism. We expect to collaborate with more universities/independent researchers in the coming years.

Conclusion

Despite the changes and delays, our team manage to mobilize thousands of species occurrences data from student theses from large time frame and collating data using sampling-event data within the project term. We managed to meet our targeted data collation within the extended project term although its validation would need more time due to the large amount of information involved. In doing so, we contribute to the overall objective of the BIFA programme in "*filling taxonomic and geographical gaps in the availability of accessible data on the occurrence of species in the region, including relative abundance and changes over time*".

Our project has enabled not only us but a number of communities to learned a lot about data sharing and research in Indonesia, particularly the ones related to biodiversity. Due to public enthusiasm for our project, we have several post-project activities including continuing biodiversity data mobilization, improving data quality, increasing data usage, and establishing a more permanent server for the data portal. We expect to keep meeting new collaborators along the way and focus on increasing public awareness on the importance of data sharing in biodiversity conservation.

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.

We intend to extend the data mobilization for the remaining theses while promoting our project to attract more universities/institutions to join our efforts. Afterwards, we intend to continue this initiative using our independent fund, working together with InaBIF to promote Indonesian biodiversity database. We previously intend to bridge Biodiverskripsi system to InaBIF to maintain the project under government funding and allow LIPI to do the curation of data, but such partnership require considerable human resource and dedicated staff which LIPI is currently lacking. Our plan to oblige students to put their data into the system as part of graduation requirements in each university also required a considerable change in paradigm about research dan data sharing. Fortunately, our closing conference has opened up new network and community who are the source of future collaborations and database maintenance.

At the closing conference, we got a person mobilizing biodiversity data using citizen science by collecting the data according to Darwin Core terms and also place the resulting data on a web-based information system. We intend to collate our data with theirs, forming a database of Indonesian flora and fauna, before we could make a solid MoU with the institution hosting InaBIF. We hope this could sustain our data portal after the end term of our current server in 2021. The planned MoU with InaBIF will be developed after the project term to sustain this initiative. We hope to incorporate a more solid review mechanism now that the data has been collected.

To evaluate the long-term impact of our database, we intended to provide interactive view counts and data set usage of the database in the web/IPT along with storing the resulting studies using Biodiverskripsi data portal. Nonetheless, incorporating such interactive features were beyond our budget and a static report will be provided instead in the publication part of our portal. When GBIF has finished indexed the datasets, we will link them to our portal to provide stats on data set usage. Tambora will also identify factors causing the use intensity of our database in the long run to improve its objective. The data portal has been developed to include user download information for the portal administrator to evaluate data usage. Although the MoU only got us one-year worth of collaboration, we intend to keep informing the university partners about the extent of usage of this database.

Recommendations and lessons learned

This section should describe your experiences that could help in designing and implementing biodiversity mobilization projects more effectively, including the best practices to adopt and the pitfalls to avoid.

We are confident to recommend our existing transcription pipeline to mobilize biodiversity data from similar types of sources with additional necessary modification. For example, some steps in our pipeline need to be modified to increase data quality, i.e. reducing duplicates, non-standardized entries, and fuzzy taxonomy. The existence of ~14,000 occurrence records along with the data portal proved that our approach could work in mobilizing biodiversity data from student theses.

To ensure data quality, however, we should not assume too much familiarity for the volunteers regarding their experience with data mobilization and Darwin Core terms. Reinforcing the volunteers' knowledge by repeating simple concepts and skills trained is safer than assuming that they already know parts of it. Strengthening their grasp on important concepts in transcribing data, such as defining a sampling event, georeferencing a locality, validating a scientific name, etc., is important to assist them mobilizing data from various types of research conducted by student theses.

Familiarity to a system is something that also should not be undermined in building a custom web-based application to present data. User experience is a large factor in determining the utility of our data portal in addition to many other relevant factors. As with the volunteers transcribing the data, we shall not assume too much familiarity for our user in using our planned data portal. Considering a community not yet familiar with a data portal is an important step and a research on user behaviour when using a data portal for the first time could be an option for anyone who intend to develop a similar system, starting from an established data portal such as GBIF. To develop an existing data portal afterwards, the evaluation should not rely on a passive questionnaire but in active request of feedback. This may require a separate dedicated project if one seek to ensure the quality of the development process.

In using a volunteer-based scheme, project coordinators should also expect longer time-frame from original plan. Transcribing the narrative in the student theses into Darwin Core terms to produce a decent amount of data simply takes time and focus. With a volunteering scheme, although with incentives, we recommend that we shall not overestimate the amount of data that could be mobilized in a time frame. Data capture and entry could be done with a few of very dedicated persons who understand the data and the purpose of the data capture,

instead of with many paid persons who did not understand the data and the purpose the data capture. We also suggest team building activities if anyone wants to use similar volunteer-based approach in their data mobilization project to sustain and reaffirm volunteers' motivation.

When at some point the team needs to do several types of tasks at once to deliver project activities and deliverables within the project's timeline, we should prioritise on the quality of data collected rather than the quantity. A data set of high quality will increase the potential data usage, trust, and future collaborations. Ensuring trust and validity of data is important to maintain the sustainability of the database. This could further address the issue of lack of trust from universities to join our effort in data mobilization.

In addition to the exclusive nature of science conducted in most of Indonesian academic institutions, the universities that we approached were hesitant in joining our initiative as we have focused too much on the data mobilization initiative and less towards the research we are trying to do. Promoting our research instead will enable people to see the bigger picture and understand the type and extent of data involved in the data mobilization effort, along the insights that could resulted from this. The publication of data paper, for example, was not familiar among our academics, but was an attractive incentive for publishing data.

A dedicated time frame that consider national calendar was also necessary to build the collaboration. As our project has showed, building the same perspective and understanding takes time and such process shall not be undermined for the sake of timeline. To keep everything within a project timeline though, we suggest a deep assessment on the research culture of the targeted collaborators and the available options of benefit for both parties.

Communicating the project to increase public awareness is an important part of the project. Any plan to communicate project's progress and results should aim to make public aware, engage, and form large part of public's concern towards biodiversity management. In the long term, an effective communication plan could further impact political agenda so that it accommodates data mobilization activities. We suggest two copywriters who could manage two languages and promote in both English and Indonesian to increase both international and national awareness.

Annex – Sources of verification

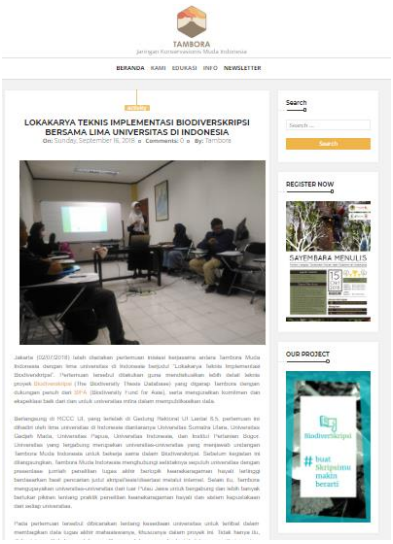
Sources of verification are for example links to relevant digital documents, news/newsletters, brochures, copies of agreements with data holding institutions, workshop related documents, pictures, etc.

Sources of verification for our deliverables and activities were arranged according to project goals as follows:

1. Ecological monitoring data collected

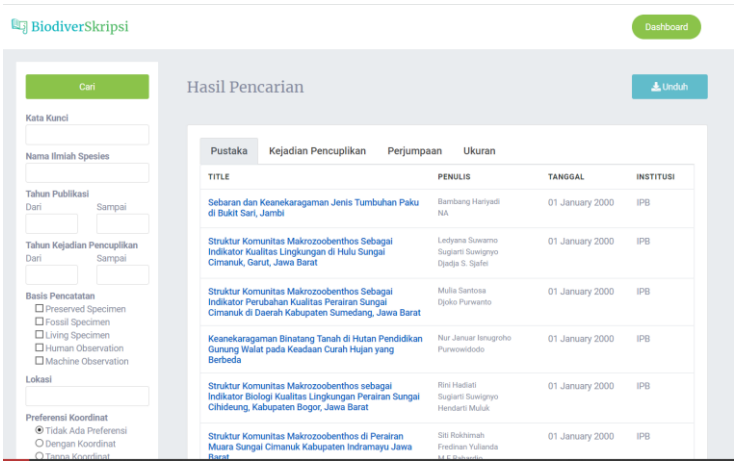
1.1. Kickoff meeting with at least ten universities and other stakeholders

Our project page in GBIF has been updated with information about kickoff meeting. It was conducted with five universities instead of ten. We also made a short report of our workshop [in our website]:



1.2. Catalogue of transcribed theses

The catalogue of transcribed theses was available in our data portal, but this list of theses has not represented the full extent of theses available from the five target universities on the year period 2000-2017. Thus, although it was published, it was not completed yet. The catalogue of transcribed theses was [available in our data portal] and will be constantly updated:



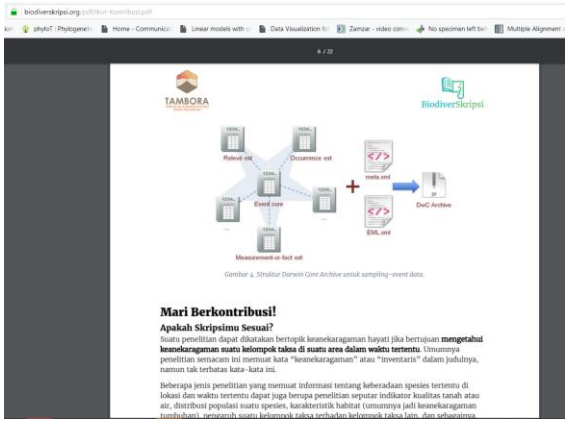
1.3. Meeting with experts to discuss review mechanism

Most of the local experts that we consulted for review mechanisms was unsure on how to best validate our data set as the nature of our data was quite unique. Few were familiar with sampling-event data templates. At the onset, we invited the experts to assist the volunteers on how to best validate a species occurrence record in the theses they transcribed. Review mechanisms has been discussed during data enumeration workshop and conducted informally through personal communications afterwards. The volunteers were instructed to check the validity of species identification based on available identification guide, literature, and any source of identification within the theses.

The willingness of the targeted experts to review the data also depends on their interest towards the project, which we tried to accommodate by developing a MoU between their institution and Tambora Muda Indonesia as the responsible maintainer of Biodiverskripsi. InaBIF has great interest in continuing this project, which we addressed in the sustainability plan.

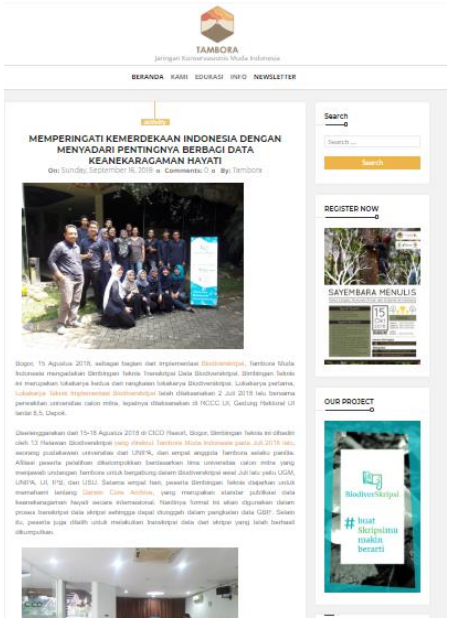
1.4. Develop data enumeration protocol guidebook

The guidebook was available in our data portal: https://biodiverskripsi.org/pdf/Ikut-Kontribusi.pdf



1.5. Workshop on data enumeration

Our project page in GBIF has been updated with information about data enumeration workshop. We also made a short report of our workshop [in our website]():



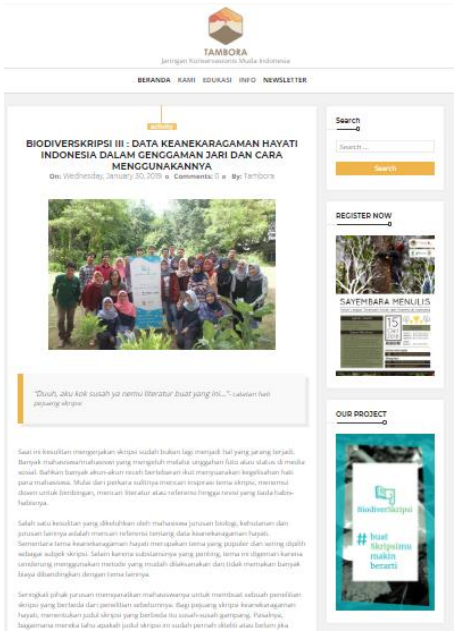
1.6. Data enumeration

The current result of data enumeration could be accessed in our IPT account or a submenu in our project page. The overall raw data that we have been working to improve could be accessed in our Researcher's GitHub.

2. Create sustainable platform for data sharing and publishing

2.1. Workshop for analysing data from data portal

Our project page has been updated with information about our workshop for analysing data from data portal. We have implemented it using GBIF portal instead of Biodiverskripsi portal and got the necessary feedback. We also made a short report of our workshop in our website:

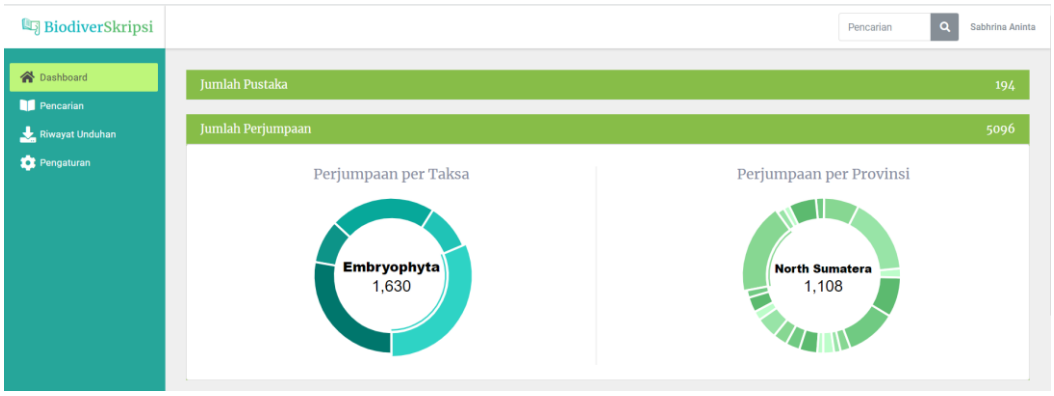


The photos of our workshop could be viewed here.

We also asked them to fill a questionnaire about the workshop to understand how students perceive the design and usability of GBIF data portal. The result of the workshop evaluation could be accessed here (in Indonesian).

2.2. Construction of custom web-based information system

The custom web-based information system has now available online in https://biodiverskripsi.org



2.3. Evaluation of custom web-based information system

We provide the evaluation of user experience using data portal in this following link (in Indonesian). Half of our respondents were not familiar with the toolbars of biodiverskripsi.org and how to download data from this portal web-based data portal. According to their suggestions in the open-ended question, the layout was not intuitive and user-friendly enough. Also, the data information in the front page

were not considered visually engaging. In addition, some links went to the incorrect sites and some downloading efforts were not working properly.

We were short of respondents willing to evaluate our data portal by the end of this project term so we consider this evaluation was not representative to depict the potential data user. Nonetheless, we take every single comment into account to improve the data portal usage.

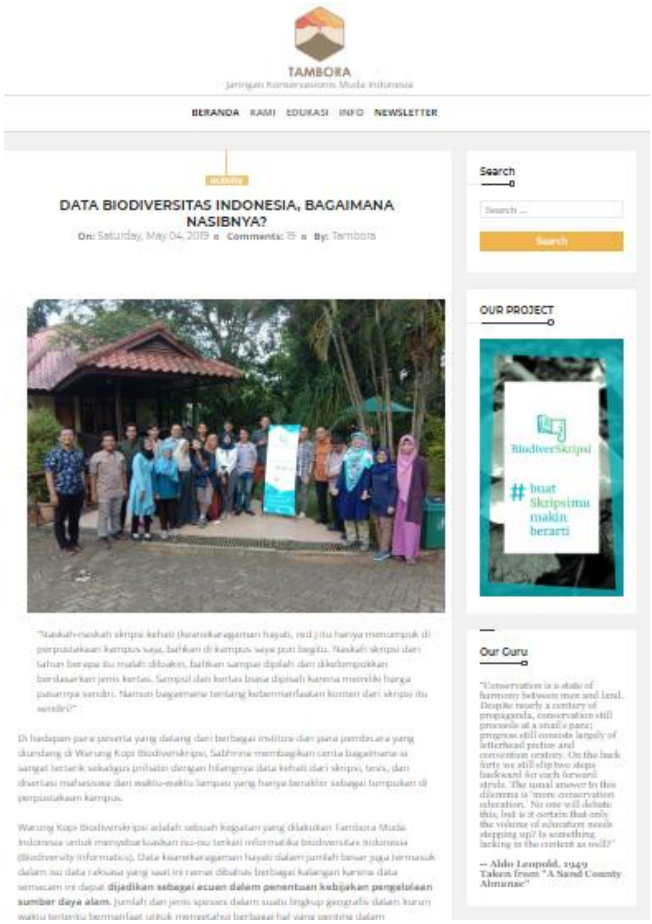
3. The state of student research

3.1. Data paper preparation and submission

We have finished the manuscript for the data paper but still waiting for the technical validation and the following peer-review.

3.2. Closing conference

Our project page has been updated with information about our closing conference. We also made a short report [in our website]:



The photos of our event could be viewed [here]

3.3. Communicate project to national audience for wider awareness

We have not yet any article in national mass media to promote our project, but we constantly update our project activities in tamboramuda.org. There were also additional article from Institut Pertanian Bogor news account that reported our workshop on data analysis using data portal.

Additionally, we boosted our social media posts to get more view, engagement, and awareness from relevant individuals. The boosted posts are here and here in our Facebook page, also here in our Instagram, and here and here in our Twitter account.