

Biodiversity Information Fund for Asia(BIFA) Development of biodiversity informatics cook and regional training workshop for Asia in 2016 FINAL ACTIVITY REPORT

Contents

1. Executive summary	1
2. Contact information.....	2
2.1. Main contact point	2
2.2. Participants involved in the project.....	2
3. Project summary	3
3.1. Activities completed	3
3.2. Ongoing and post-project activities	5
4. Project objectives.....	5
5. Project deliverables	6
6. Project communications.....	7
7. Evaluation: findings and conclusions.....	7
8. Recommendations and lessons learned.....	7
9. Future plans	8

1. Executive summary

With the Funding supports from BIFA, the Ministry of Science and Technology and the Biodiversity Research Centre of Academia Sinica, Taiwan, and the Biodiversity Centre, Ministry of Environment Japan, the Biodiversity Information Facilities of Japan (JBIF) and Taiwan (TaiBIF) have finished the following tasks to promote the development of biodiversity informatics and to expand the engagement in GBIF in Asia:

1) [2016 GBIF Asia Biodiversity Informatics Workshop](#), 7-10 September 2016

JBIF and TaiBIF co-organized the [2016 GBIF Asia Biodiversity Informatics Workshop](#) at the Biodiversity Research Centre, Academia Sinica, Taiwan. This workshop focused on open data licenses, data cleaning and preparation, publishing biodiversity data (emphasized on sampling-event data), and re-using GBIF-mediated data for integrated analysis using open source tools. There were 30 participants from 10 Asian countries, and 11 domestic participants (PhD students, professors, researchers) from the universities and insitutes in Taiwan.

2) [8th AP BON workshop](#), 12-13 September 2016

TaiBIF and the Biodiversity Centre, Ministry of Environment Japan co-organized the [8th AP BON workshop](#) at the same venue as back-to-back activities to promote the collaboration between GBIF and AP BON. During the workshop, participants have discussed the issues on managing and publishing data from AP BON. GBIF Asian nodes can support managing and publishing biodiversity observation data for AP BON using the IPTs of JBIF and ACB.

4) Scientific name matching service

To facilitate the process of checking scientific names, TaiBIF developed a scientific name matching service, [NomenMatch](#). [NomenMatch](#) can accept bulk names (>10,000) to match the scientific names against the defined name sources in a batch. This will facilitate the name matching task. The source code is released on the [GitHub source repository of TaiBIF](#) with GPLv3 license.

5) Publishing sampling-event dataset and data paper

Through the [2016 GBIF Asia Biodiversity Informatics Workshop](#), the following sampling-event dataset and occurrence data paper have been published:

- Taiwan Forestry Research Institute registered the dataset “[Alien plant presence dataset from the point-radius plot surveys in 2010-2015 in Taiwan](#)” with CC0 license to GBIF. This dataset included 63,085 occurrence records from 4,879 sampling plots.
- Dr. Yuichi Kano, published a data paper “[A dataset of fishes in and around Inle Lake, an ancient lake of Myanmar, with DNA barcoding, photo images and CT/3D models](#)” on the Biodiversity Data Journal. It is a ready and prominent outcome of the collaboration between GBIF and AP BON.

6) [Biodiversity informatics cookbook](#)

JBIF and TaiBIF collaborated on producing a biodiversity informatics cookbook and released its first version as a Google Document with CC BY license.

Through this project, we hope participants will learn how to reuse existing resources to help the development of biodiversity informatics and promote biodiversity open data in Asia. The node managers can collaborate with the seed trainers to organize domestic training activities to strengthen the capabilities of universities and institutions in managing biodiversity relevant data and in integrated analysis using open source tools and open data. We expect the outcome of this project will fulfil the GBIF Implementation Plan 2017-2021 at national and regional scales.

2. Contact information

2.1. Main contact point

Name of the contact person	Yu-Huang Wang
Institutional affiliation	Taiwan Biodiversity Information Facility, Biodiversity Research Center, Academia Sinica, Taiwan
Mailing address	Room C301, Building of Interdisciplinary Technology, Biodiversity Research Center, Academia Sinica, 128 Academia Rd., Sec. 2, Nankang, Taipei 11529, Taiwan
City and country	Taipei, Taiwan
E-mail	yuhuangwang@gmail.com
Telephone	886-2-27872221
Role(s) in this project	Project coordinator, co-author of BI cookbook, organizer & trainer of regional and domestic workshops

2.2. Participants involved in the project

GBIF	Name and role of the representative	Role of the Participant in the	Confirmed
------	-------------------------------------	--------------------------------	-----------

Participant represented		project	participation?
JBIF	Tsuyoshi Hosoya	1; 2; 3; 4; 5	Yes
JBIF	Takeshi Osawa	1; 2; 3; 4; 5	Yes
JBIF	Utsugi Jinbo	1; 2; 3; 4; 5	Yes
TaiBIF	Mao-Ning Tuanmu	2; 3; 4; 5	Yes
TaiBIF	Guan-Shuo Mai	2; 3; 4; 5	Yes
TaiBIF	Kwang-Tsao Shao	2; 3	Yes

1: co-author of BI cookbook; 2: regional workshop co-organizer; 3: regional workshop trainer; 4: domestic workshop organizer; 5: domestic workshop trainer

3. Project summary

To promote the development of biodiversity informatics and to expand the engagement in GBIF in Asia, the original plan of this project was using the BIFA fund and the matching funds from Taiwan to produce an online open access biodiversity informatics cookbook and to support the travel costs of participants from Asia (including, GBIF and non-GBIF members as well as AP BON members) to attend the 2016 GBIF Asia Biodiversity Informatics Workshop. Through this workshop, we hope to train the seed trainers to disseminate the knowledge and technique of biodiversity information and promote publishing biodiversity open data (especially the sampling-event data) in Asia. We also hoped that 1-2 data papers produced by the workshop participants could be published after this workshop. To establish the collaboration between GBIF Asian nodes and AP BON, it was suggested to have the 8th AP BON workshop just after the biodiversity informatics workshop in Taiwan. Since some participants of the biodiversity informatics workshop are AP BON members, we think it is reasonable to make these two workshops back-to-back for further discussing the collaboration on publishing biodiversity observation data from AP BON to GBIF.

3.1. Activities completed

1) Meetings for compiling Biodiversity Informatics Cookbook and planning workshops

For compiling the biodiversity informatics cookbook as well as preparing the regional training workshop on biodiversity informatics and the 8th AP BON workshop in Taiwan, four meetings have taken place:

- [11 May 2016, Tokyo, Japan](#)

TaiBIF staffs (Yu-Huang Wang and Guan-Shuo Mai) had a meeting with the AP BON chair and the staffs of Biodiversity Centre, Ministry of Environment Japan at the Tokyo Office of the Kyushu University to discuss the nomination of AP BON members to attend the biodiversity informatics workshop and the tentative agenda of the 8th AP BON workshop. BIFA fund was used to support the travel cost for the meeting in Japan.

- [12-13 May 2016, Tsukuba, Japan](#)

TaiBIF staffs had a meeting with JBIF members (Tsuyoshi Hosoya, Utsugi Jinbo, and Takeshi Osawa) to discuss how to compile the biodiversity informatics cookbook and to practice on using R scripts to convert the huge raw data tables into the structure of DarwinCore sampling-event data. During this meeting, we also discussed the tentative agenda for the 7th Asia regional node meeting in Philippines and the biodiversity informatics workshop in Taiwan. BIFA fund was used to support the travel cost for the meeting in Japan.

- [24 May 2016, Taipei, Taiwan](#)

National Taiwan Normal University and TaiBIF invited Townsend Peterson to Taiwan to have a niche modelling and biodiversity informatics workshop for the international graduate students in Taiwan. Townsend also had a meeting with TaiBIF staffs to discuss the regional and national diagnoses for identifying the priorities, gaps, concentrations of biodiversity information. Townsend provided TaiBIF a 32GB USB drive with all the tutorial materials (presentation slides and videos) of the BITC web site. He encouraged TaiBIF to distribute all the training materials to participants of the GBIF Asia Biodiversity Informatics Workshop and anyone who want learn biodiversity informatics. BIFA fund was used to support the airfare for Townsend Peterson's travel between US and Taiwan.

- 21-22 July 2016, Taipei, Taiwan

TaiBIF and JBIF had the meeting in Academia Sinica, Taiwan to further discuss compiling the biodiversity informatics cookbook and preparing the biodiversity informatics workshop. BIFA fund was used to support the travel cost for the meeting in Taiwan.

2) [2016 GBIF Asia Biodiversity Informatics Workshop](#), 7-10 September 2016

With the support from the BIFA fund as well as the co-funds from the Ministry of Science and Technology and the Biodiversity Research Centre of Academia Sinica, Taiwan, The Biodiversity Information Facility of Japan and Taiwan co-organized the 2016 GBIF Asia Biodiversity Informatics workshop. This workshop focused on open data licenses, data cleaning and preparation, publishing biodiversity data (emphasized on sampling-event data), and re-using GBIF-mediated data for integrated analysis using open source tools. There were 30 participants from 10 Asian countries, and 11 domestic participants (one PhD student, seven professors, three researchers) from the universities and museum/institutes in Taiwan. Among the 30 foreign participants, eight of them were from non-GBIF member countries (one from Bhutan, two from Malaysia, two from Myanmar, three from Singapore), nine were from AP BON members (7 out of 9 were also from GBIF members). All the presentation slides, video and tutorial datasets are released with CC BY license and can be accessed the [workshop programme](#) of a Google site.

3) [8th AP BON workshop](#), 12-13 September 2016

TaiBIF and the Biodiversity Centre, Ministry of Environment Japan co-organized the [8th AP BON workshop](#) at the same venue as back-to-back activities to promote the collaboration between GBIF and AP BON. There were seven participants of the biodiversity informatics workshop also joined the AP BON workshop. Thus, during the 8th AP BON workshop, participants also have discussed the issues on managing and publishing data from AP BON. GBIF Asian nodes can support managing and publishing biodiversity observation data for AP BON using existing information infrastructure of GBIF Asia, the IPTs of JBIF and ACB. The Biodiversity Centre, Ministry of Environment Japan supported the cost for the workshop.

4) Scientific name matching service

It is a tedious and time-consuming task to correct errors of scientific names in publishing data sets; however, it is essential to make reliable and high-quality dataset. To increase the efficiency in checking scientific names, TaiBIF developed a scientific name matching service, [NomenMatch](#). [NomenMatch](#) can accept bulk names (>10,000) to match the scientific names against the defined name sources in a batch. This will facilitate the name matching task. The source code is released on the [GitHub source repository of TaiBIF](#) with GPLv3 license.

5) Publishing sampling-event data sets and data papers

For demonstrating how to publish sampling-event data during the workshop, one sampling-event data set has been published by Taiwan Forestry Research Institute and been registered to GBIF with CC0 license:

[Alien plant presence dataset from the point-radius plot surveys in 2010-2015 in Taiwan](#). This dataset included 63,085 occurrence records from 4,879 sampling plots.

During the Biodiversity Informatics training workshop, one of the participants from AP BON, Dr. Yuichi Kano, submitted a data paper manuscript "[A dataset of fishes in and around Inle Lake, an ancient lake of Myanmar, with DNA barcoding, photo images and CT/3D models](#)" to the Biodiversity Data Journal. Recently, this paper has been formally published. It is a ready and prominent outcome of the collaboration between GBIF and AP BON.

6) [Biodiversity Informatics Cookbook](#)

JBIF and TaiBIF collaborated on producing a biodiversity informatics cookbook based on the needs from Asia regional nodes activities but also referring to the concurrent developmental works from GBIF global nodes. The first version of this BI cookbook has been released as a Google Document with CC BY license.

3.2. Ongoing and post-project activities

1) Domestic training workshops

Based on the experiences and training materials from the BIFA biodiversity informatics workshop, TaiBIF has planned to have the first domestic biodiversity informatics workshop on 17-19 January 2017 in Taipei. This workshop will invite students, research assistants and researchers to join. The topics will include introducing the importance of open data and the GBIF in global biodiversity community, how to use GBIF data standards to publish biodiversity open data, how to use open source tools to do integrated analysis using data from GBIF and other open data sources. All the training courses will be conducted in Chinese to help domestic participants easily learn more about biodiversity informatics.

2) Publishing biodiversity observation data (with focus on sampling-event data)

TaiBIF works with NGOs to help them manage and publish the data from monitoring projects conducted by volunteers. The [Kurushio Ocean Education Foundation](#) actively promotes the conservation of ocean biodiversity in Taiwan and has documented the occurrences of whale and dolphin along the offshore of Hualien, eastern Taiwan. Recently the [Kurushio Ocean Education Foundation](#) has published the sampling-event dataset "[The observation records from the activities of whale and dolphin watching in the offshore of Hualien, Eastern Taiwan](#)" with CC BY license on TaiBIF IPT. Currently, the dataset includes 5,247 occurrence records in 1998-2015; however, it still needs to be provided more detailed metadata and errors in the raw data must be corrected or removed. Once the dataset is cleaned and provided with English metadata it will be registered to GBIF.

4. Project objectives

1) Continuously develop and update the contents of biodiversity informatics cookbook through community collaborations on GitBook open access platform.

Originally, we plan to use the GitBook plan create and co-edit the cookbook but we all feel it is not easy to edit tables, citations and references on GitBook. Finally, we switched to Google Document to create and share the cookbook. The first version can be downloaded

for reuse and modification. This cookbook will be continuously updated by Asian nodes to support the need of regional training activities.

- 2) Establish “training seed trainers” model in regional nodes for long-term supporting the development of biodiversity informatics in Asia.

We provided every participant a 64GB USB 3.0 pen drive for storing all the presentation slides, tutorial data, software tools, which were used in the workshop, as well as the training materials of [the BITC web site](#) shared by Townsend Peterson. We expect the participants as the seed trainers can share the knowledge and resources they got from the workshop with their people. All the contents in the USB drive are freely distributable for training and self-paced learning. This could be useful for sharing the knowledge and techniques of biodiversity information in Asia.

- 3) Enhance capacity of nodes for providing biodiversity informatics services for universities, research institutions, and governmental agencies to support science-policy needs.

Another dimension to be considered is the diversity of local languages. We encouraged the participants (seed trainers) reuse, remix, or modify the freely available resources and help convert the new contents into local languages to create a biodiversity informatics training curriculum suitable for local use. National and regional nodes will be the helpdesk to provide technical support and some information service to help the development of biodiversity informatics in Asia.

- 4) Advance the engagement in GBIF in Asia by inviting scientists from non-GBIF member countries in South-East Asia and AP BON.

To expand the engagement in GBIF, ASEAN Centre for Biodiversity (ACB) and Asia-Pacific Biodiversity Observation Network (AP BON) endorsed scientists from non-GBIF members or from AP BON to join the biodiversity informatics workshop. There were thirty [participants](#) from ten Asian countries (not including Taiwan) attending this workshop. Eight out of these participants were from non-GBIF member countries: one from Bhutan, two from Malaysia and Myanmar respectively, and three from Singapore.

5. Project deliverables

- 1) Release the first version of Biodiversity Informatics Cookbook

The first version of [Biodiversity Informatics Cookbook](#) has been created using Google Document and been released under CC BY 4.0 license. Anyone can download and modify the content.

- 2) Have a regional training workshop on Biodiversity Informatics

JBIF and TaiBIF co-organized the 2016 GBIF Asia Biodiversity Informatics workshop. This workshop has taken place on 7-10 September 2016 at the Biodiversity Research Centre of Academia Sinica, Taipei, Taiwan. All the workshop resources, including presentation PPT and videos as well as tutorial datasets can be downloaded and reused under CC BY 4.0 license. All the resources can be accessed via the Google Site for the [2016 GBIF BIFA Biodiversity Informatics Workshop](#).

- 3) Publish 1-2 data papers (with emphasis on sampling-event data) by the end of 2016.

- Dr. Yuichi Kano, one of the participants of biodiversity informatics workshop from AP BON, published a data paper “[A dataset of fishes in and around Inle Lake, an ancient lake of Myanmar, with DNA barcoding, photo images and CT/3D models](#)” on the Biodiversity Data Journal.
- Taiwan Forestry Research Institute has registered a sampling-event dataset “[Alien plant presence dataset from the point-radius plot surveys in 2010-2015 in Taiwan](#)” with CC0 license to GBIF. The manuscript for submitting the data paper of this dataset is under preparation. We expect to submit it to Biodiversity Data Journal or Ecological Research by the end of 2016.

4) Have a domestic training workshops on Biodiversity Informatics in Taiwan by the end of January 2017.

To promote publishing sampling-event data and reusing GBIF mediated data in Taiwan, TaiBIF has planned to have the first domestic training workshop on biodiversity informatics. With the three-year project funding support from the Ministry of Science and Technology, TaiBIF will hold biodiversity informatics workshops for domestic students, research assistants and researchers in universities and institutes during summer and winter vacation.

6. Project communications

The information, presentations, and tutorial datasets of the biodiversity informatics workshop as well as the biodiversity informatics cookbook can be accessed by visiting:

- [2016 GBIF Asia Biodiversity Informatics Workshop](#) Google site
- EML-DarwinCore Taxon / Occurrence / Sampling-event [data templates in Excel](#)
- Workshop [tutorial datasets](#)
- [Biodiversity Informatics Cookbook](#) (version 1)

7. Evaluation: findings and conclusions

- 1) Access to stable and high speed internet is very limited in some countries of South-East Asia. This is a big obstacle to disseminating the knowledge of biodiversity informatics and to sharing biodiversity data. Thus, the need for offline data managing / publishing tool is necessary for South-East Asia. Currently, Using the Excel templates for collecting datasets of taxon / occurrence / sampling-event data with EML metadata is an easy alternative. Data providers can send the datasets created by the Excel templates to the node managers via email or USB drive / memory card, with the help from the node to publish datasets.
- 2) Currently, most of the IPTs in Asian are older versions which cannot support publishing sampling-event data. We encourage the Asian nodes to upgrade their IPT to latest version to support publishing sampling-event data and issuing DOI.

8. Recommendations and lessons learned

- 1) Reuse existing online resources

There are comprehensive online high-quality resources on [the BITC web site](#) for self-paced learning biodiversity informatics. There are very abundant tutorial videos to learn open source tools, like DIVAS-GIS, GRASS GIS, gvSIG, QGIS, R, Python, etc., on the YouTube. Many web sites or blogs also have advanced tutorial resources for learning application of

remote sensing, GIS, niche modelling, etc. in R. It will be very useful and fruitful to aggregate and re-organise / remix these existing resources to design the biodiversity informatics curriculum.

2) Provide offline training materials for self-paced learning

For some people in area without high speed internet to access the online learning resources, it will be better to provide a high-volume USB drive with learning resources and software tools to local learning centres, like universities, institutes, and NGOs to share these learning resources.

3) Enhance the training in using open data and open source tools for integrated biodiversity analysis

Open data and open source tools are ever increasing. It is essential to learn how to utilise these resources efficiently to map issues, priorities, trends in biodiversity change for rescuing, conserving and restoring biodiversity. GBIF annual science review and the Mendeley public library are good aggregated sources to learn the real cases and the progress. However, we still need a place to show the transparent processes of integrated analysis for people to learn how to use the resources for local biodiversity conservation. Maybe adopting Jupyter Notebook or R-Shiny server can help make a transparent analysis platform facilitating the learning process.

9. Future plans

1) Upgrading the IPTs to latest version to support publishing sampling-event data

Most of the nodes in Asia still use old versions of IPT, which is not supporting publishing sampling-event data. Since sampling-event data are important to support producing essential biodiversity variables and other applications, and the goal of this project is to promote publishing sampling-event data from Asia, we will encourage all Asian nodes to upgrade their IPT to latest version. We can track the progress via by regular online regional meetings or questionnaire survey with nodes.

2) Domestic summer / winter biodiversity informatics training workshops

TaiBIF has gotten the funding support from the Ministry of Science and Technology to promote the development of biodiversity informatics in Taiwan. Thus, we plan to have domestic summer / winter biodiversity informatics training workshops for students, research assistants and researchers in universities and institutes during summer and winter vacation in 2017-2018. These domestic training activities will strengthen the capabilities of universities and institutions in managing biodiversity relevant data and in integrated analysis using open source tools and open data. We expect the outcome of this plan can fulfil the GBIF Implementation Plan 2017-2021 at national scale.

3) Continuation in updating the content of Biodiversity Informatics Cookbook

New methods and new data sources for biodiversity informatics are ever increasing. Asian nodes will try to update the content and structure in the Biodiversity Informatics Cookbook to aggregate new progresses for easily exploring these online resources.

4) Online webinar series on biodiversity informatics in local languages

Just like [the BITC webinars](#), national nodes also can organize local teams to have series of webinars on biodiversity informatics in local languages. The most efficient way is to reuse the existing resources. Thus, we suggest Asian nodes can help translate the subtitles of the BITC online resources.

Signed on behalf of the project partners

Date

Yu-Huang Wang

2016-12-02
