

Development of biodiversity informatics cookbook and regional training workshop for Asia in 2016

Mao-Ning Tuanmu¹ & Tsuyoshi Hosoya²

¹Taiwan Biodiversity Information Facility (TaiBIF)

²Japan Node of Global Biodiversity Information Facility (JBIF)

2017 GBIF Asia Regional Meeting

2017-06-13

Hanoi, Vietnam



The Team



Dr. Takeshi Osawa



Dr. Tsuyoshi Hosoya



Dr. Yu-Huang Wang



Dr. Kwang-Tsao Shao



Dr. Mao-Ning Tuanmu



Dr. Utsugi Jinbo

Goal

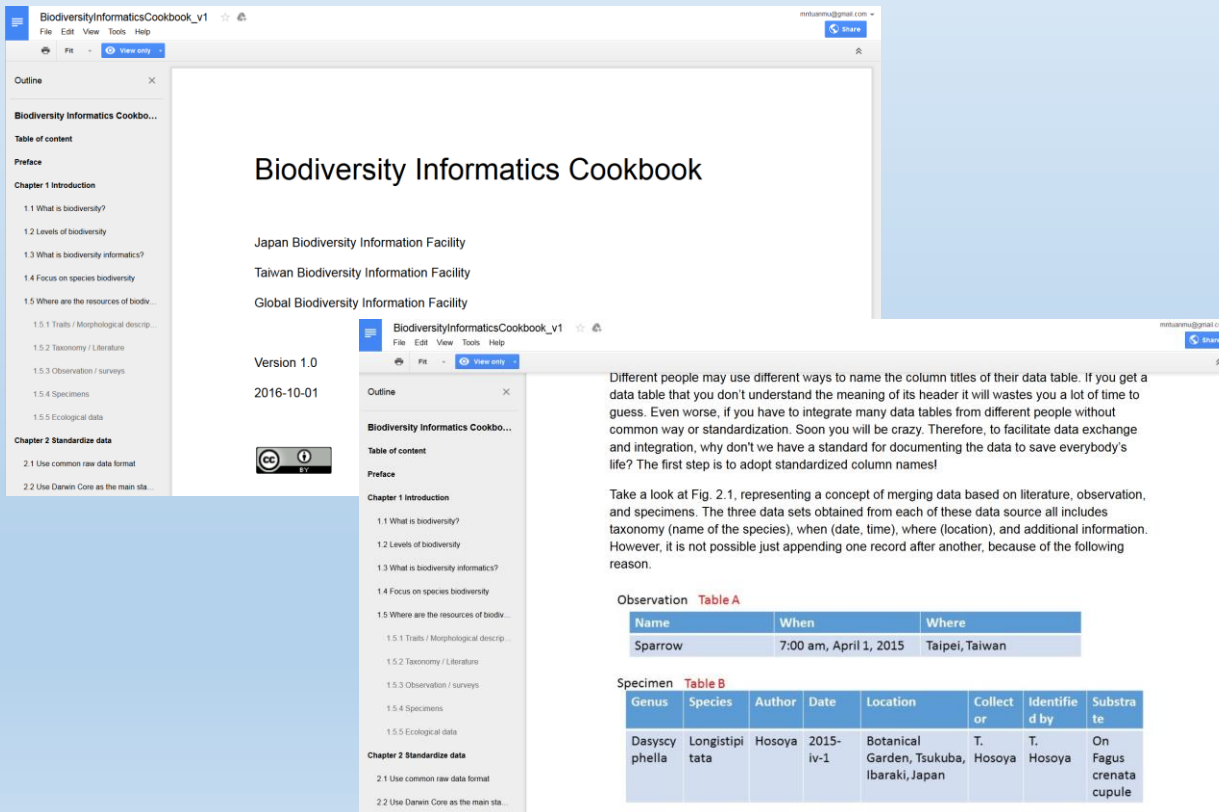
To foster the application and development of **biodiversity informatics** and expand the **engagement of Asian countries in GBIF**

Objectives

- To develop curriculum materials on biodiversity informatics
- To establish a “training seed trainers” model as long-term support for the development of biodiversity informatics
- To enhance the capacity of nodes for providing biodiversity informatics services

Major Achievements - BI Cookbook

- Version 1.0 was released as a Google Document on Oct. 1, 2016



- Chapter 1. Introduction
- Chapter 2. Standardize data
- Chapter 3. Prepare datasets for publishing
- Chapter 4. Improve data quality
- Chapter 5. Adapt open data license
- Chapter 6. Publish biodiversity data
- Chapter 7. Reuse biodiversity data

Workshop on niche modeling and biodiversity informatics

- Dr. Townsend Peterson
- Regional and national diagnoses for identifying the priorities, gaps, concentrations of biodiversity information
- Tutorial materials at the Biodiversity Informatics Training Curriculum web site



BIODIVERSITY INFORMATICS TRAINING CURRICULUM

HOME **BI CURRICULUM** BI WEBINAR SERIES BITC TRAINING COURSES SOFTWARE & DATA RESOURCES

INTRODUCTION

Introduction : BITC Introductory Materials

| TOPIC | LECTURE TITLE | VIDEO LINK(S) | LECTURE MATERIALS | LECTURER(S) | SPONSOR | DATE | LANGUAGE | SUBTITLES |
|--|---|-------------------------|----------------------------------|------------------------------------|--------------|---------------|----------|----------------|
| Introduction to Biodiversity Informatics | | | | | | | | |
| | Introduction to Biodiversity Informatics | Lecture | | Town Peterson | BITC | February 2013 | English | In Development |
| | History of Biodiversity Informatics - part I | Lecture | Presentation PDF | Dr. Arthur Chapman | BITC | January 2016 | English | In Development |
| | History of Biodiversity Informatics - part II | Lecture | | Jorge Sobrado | BITC | February 2016 | English | In Development |
| | History of Biodiversity Informatics - part III | Lecture | | John Wliczonek | BITC | March 2016 | English | In Development |
| | History of Global Biodiversity Information Facility | Lecture | | Donald Hobem | BITC Webinar | June 2016 | English | |
| | History of Visualizations in Biodiversity Informatics | Lecture | | Javier Ortega | BITC Webinar | August 2016 | English | |
| | Subtitling BITC Videos Initiative | | | | | October 2016 | English | |
| The "other" bioinformatics | | | | | | | | |
| Primary data vs. secondary data | | | | | | | | |
| Research-grade data vs. "for show" data | | | | | | | | |
| Organization, visualization, analysis, integration | | | | | | | | |
| | Building the Biodiversity Knowledge Graph | Lecture | | Rod Page | BITC | June 2014 | English | |
| Funding Proposals | | | | | | | | |
| | Writing effective proposals for funding | Lecture | | Town Peterson | | March 2013 | English | In Development |
| Writing scientific papers for publication | | | | | | | | |
| | Introduction | Lecture | | Town Peterson | | December 2012 | English | In Development |

GBIF Asia Biodiversity Informatics Workshop

- Sept. 7-10, 2016
- Academia Sinica, Taipei, Taiwan
- 41 participants (potential trainers) from 11 Asian countries (including 8 from non-GBIF member countries and 9 from AP BON members)



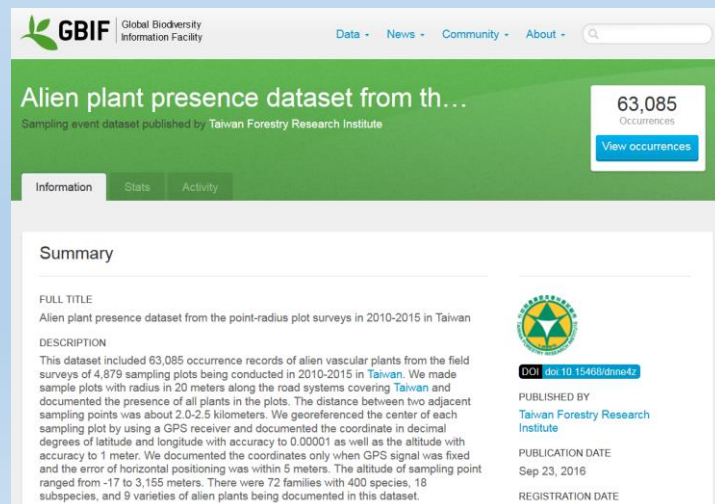
GBIF Asia Biodiversity Informatics Workshop

- Standardizing and cleaning data
- Publishing occurrence data
- Publishing sample-event data
- Preparing data papers
- Obtaining data from GBIF
- Reusing data
- Lectures and hands-on exercises

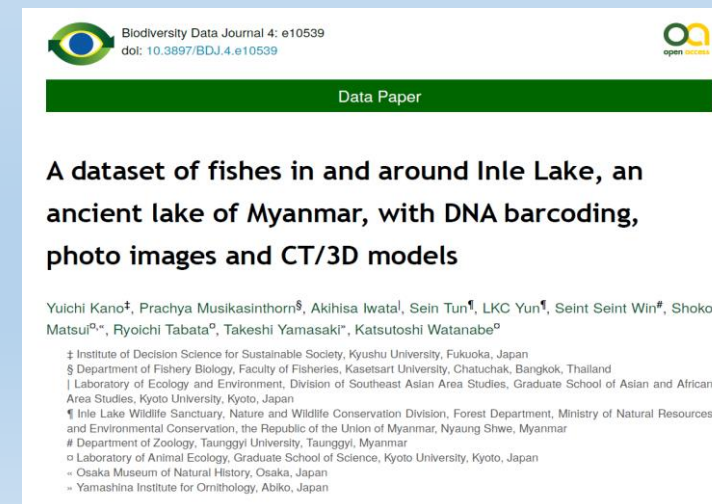


Data Publishing & Data Papers

- Alien plant presence dataset
- 63,085 occurrence records at 4,879 sampling plots
- Taiwan Forestry Research Institute
- Kano, Yuichi, et al. "A dataset of fishes in and around Inle Lake, an ancient lake of Myanmar, with DNA barcoding, photo images and CT/3D models." *Biodiversity Data Journal* 4 (2016).



The screenshot shows the GBIF (Global Biodiversity Information Facility) website interface. At the top, the GBIF logo and navigation links (Data, News, Community, About) are visible. The main heading is "Alien plant presence dataset from th..." with a sub-heading "Sampling event dataset published by Taiwan Forestry Research Institute". A prominent green box displays "63,085 Occurrences" and a "View occurrences" button. Below this, there are tabs for "Information", "Stats", and "Activity". The "Summary" section is expanded, showing the "FULL TITLE" and "DESCRIPTION". The description states: "This dataset included 63,085 occurrence records of alien vascular plants from the field surveys of 4,879 sampling plots being conducted in 2010-2015 in Taiwan. We made sample plots with radius in 20 meters along the road systems covering Taiwan and documented the presence of all plants in the plots. The distance between two adjacent sampling points was about 2.0-2.5 kilometers. We georeferenced the center of each sampling plot by using a GPS receiver and documented the coordinate in decimal degrees of latitude and longitude with accuracy to 0.00001 as well as the altitude with accuracy to 1 meter. We documented the coordinates only when GPS signal was fixed and the error of horizontal positioning was within 5 meters. The altitude of sampling point ranged from -17 to 3,155 meters. There were 72 families with 400 species, 18 subspecies, and 9 varieties of alien plants being documented in this dataset." A logo for the Taiwan Forestry Research Institute is also present.



The screenshot shows the Biodiversity Data Journal 4: e10539 page. The header includes the journal title, DOI (10.3897/BDJ.4.e10539), and the Open Access logo. A green bar indicates "Data Paper". The main title is "A dataset of fishes in and around Inle Lake, an ancient lake of Myanmar, with DNA barcoding, photo images and CT/3D models". The authors listed are Yuichi Kano[‡], Prachya Musikasinthorn[§], Akihisa Iwata^{||}, Sein Tun[¶], LKC Yun[¶], Seint Seint Win[¶], Shoko Matsui^{¶,•}, Ryoichi Tabata[¶], Takeshi Yamasaki[¶], and Katsutoshi Watanabe[¶]. The page lists affiliations for each author: [‡] Institute of Decision Science for Sustainable Society, Kyushu University, Fukuoka, Japan; [§] Department of Fishery Biology, Faculty of Fisheries, Kasetsart University, Chatuchak, Bangkok, Thailand; ^{||} Laboratory of Ecology and Environment, Division of Southeast Asian Area Studies, Graduate School of Asian and African Area Studies, Kyoto University, Kyoto, Japan; [¶] Inle Lake Wildlife Sanctuary, Nature and Wildlife Conservation Division, Forest Department, Ministry of Natural Resources and Environmental Conservation, the Republic of the Union of Myanmar, Nyaung Shwe, Myanmar; [•] Department of Zoology, Taunggyi University, Taunggyi, Myanmar; [•] Laboratory of Animal Ecology, Graduate School of Science, Kyoto University, Kyoto, Japan; [•] Osaka Museum of Natural History, Osaka, Japan; [•] Yamashina Institute for Ornithology, Abiko, Japan.

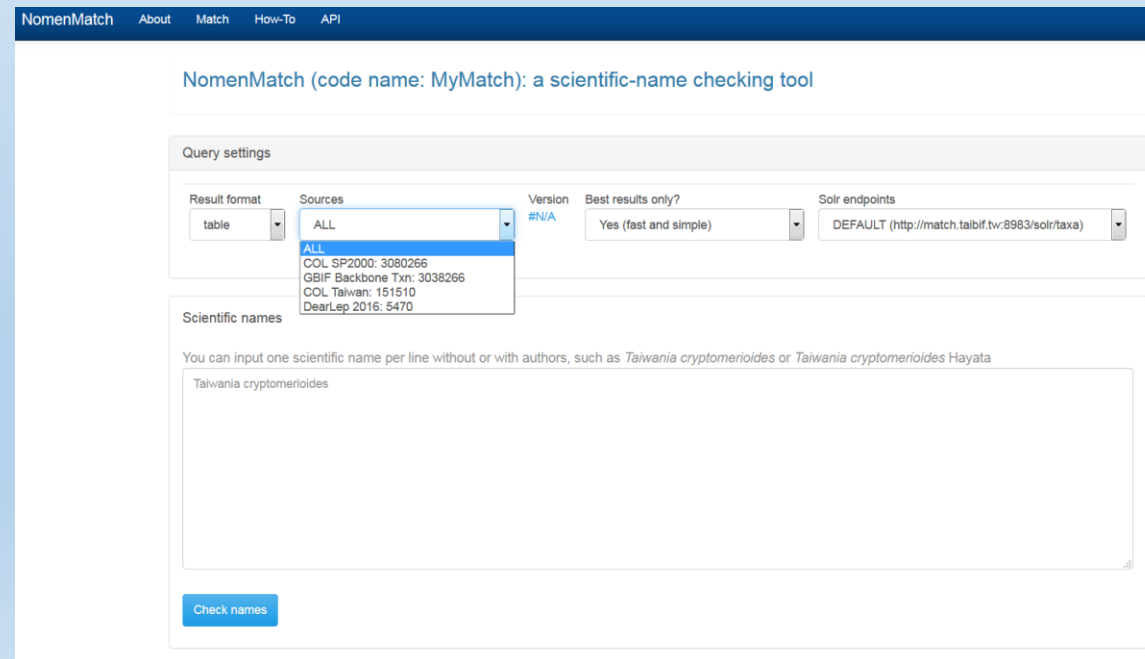
National BI Workshops

- Jan. 17-19, 2017
 - Academia Sinica, Taipei, Taiwan
 - 29 participates
 - Graduate and undergraduate students
- Aug. 28-31, 2017
 - Endemic Species Research Institute, Nantou, Taiwan
 - Emphasizing more on the reuse of open biodiversity data



NomenMatch – Scientific name matching service

- Match bulk scientific names (> 10,000) in datasets against the defined names sources (COL SP2000; GBIF Backbone; COL Taiwan)
- GitHub source repository of TaiBIF



NomenMatch About Match How-To API

NomenMatch (code name: MyMatch): a scientific-name checking tool

Query settings

Result format: table

Sources: ALL (dropdown menu showing: ALL, COL SP2000: 3080266, GBIF Backbone Txn: 3038266, COL Taiwan: 151510, DearLep 2016: 6470)

Version: #N/A

Best results only?: Yes (fast and simple)

Solr endpoints: DEFAULT (http://match.taibif.tw:8983/solr/taxa)

Scientific names

You can input one scientific name per line without or with authors, such as *Taiwania cryptomerioides* or *Taiwania cryptomerioides* Hayata

Taiwania cryptomerioides

Check names

Available resources resulted for the future

- [BI cookbook](#) (version 1.0)
- [GBIF Asia BI workshop materials](#) (videos, lecture slides, exercise datasets, software, tools)
- [Domestic BI workshop materials](#)
- [Scientific name matching tool](#)

Lessons learned

- Integration of available resources in biodiversity informatics training
 - BITC web site
- Local needs
 - BI curriculum suitable for local use
 - Training in local languages
- Resource limitations
 - Limited Internet connection
 - Excel templates
- Incentive to open up data
 - Data papers
 - Demonstration of data reuse
 - Facilitation tools for data cleaning, standardization and publishing
- Connection with AP BON
 - Data standards, metadata
 - Bidirectional flows of data