

BIODIVERSITY INFORMATICS TRAINING CURRICULUM, VERSION 1.2

A. TOWNSEND PETERSON AND KATE INGENLOFF

Biodiversity Institute, University of Kansas, Lawrence, Kansas 66045 USA

Abstract.—Biodiversity informatics as a field lacks a synthetic, comprehensive summary, guide, and training resource, such as a textbook, a gap that impedes advance in the field. The *Biodiversity Informatics Training Curriculum* (BITC) represents the compilation of digital videos and accessory materials from 3 years of courses held across Africa, on topics covering the full thematic breadth of the field of biodiversity informatics. Instructors include experts on each topic, such that the instruction is at a high level. The BITC is presented as an open-access, fully digital, video-based teaching resource—in essence a first textbook for the field. The curriculum is constantly being updated and improved, but this version 1.2 represents a first version that covers all major subfields in biodiversity informatics. Links to videos hosted on YouTube are provided herein, but versions that do not require Internet access are available upon request.

Key words.—biodiversity, informatics, teaching, digital video

Biodiversity informatics is, simply put, a new field (Schalk 1998; OECD Megascience Forum 1999; Soberón 1999; Bisby 2000). Informatics more generally is the science of creating, organizing, and analyzing digital data. Although biodiversity scientists (systematics, taxonomists, ecologists, biogeographers) have used informatics processes in studies of biodiversity for centuries, the digital dimensions of the challenge are relatively new. Indeed, it is only in the last decade or so that large-scale biodiversity databases have become broadly available and accessible. The major steps of innovation and advancement of the field—in the view of many—will derive from novel but meaningful linkages between disparate realms of information about biodiversity (Peterson et al. 2010).

Given the newness of the field, biodiversity informatics remains largely without any formal synthesis, guide, textbook, or summary. That is, no textbook entitled something like “*Biodiversity Informatics*” exists to guide users through the steps involved in biodiversity informatics applications. Further, no graduate programs exist that provide a comprehensive curriculum for the field. Clearly, books and programs exist treating portions of the field (e.g., ecological niche modeling, Peterson et al. 2011), but no overarching summary has been developed.

The Biodiversity Informatics Training Curriculum (BITC) represents an effort to remedy this problem and fill the gap in availability of knowledge and experience in the field. With the generous support of the JRS Biodiversity

Foundation, a series of detailed training courses has been held at sites across Africa, each treating a different topic within the broader area of biodiversity informatics, with experts brought in from around the world, and trainees from across Africa. Full summaries of the courses, and the experts and trainees who attended, are available on the BITC website¹.

This document is designed to provide a summary and index to the first nearly-complete version of the full Curriculum. Each topic in the long list in the Appendix is a hyperlink to a YouTube video or playlist. The topics are more or less in a logical order, although some topics really belong in multiple categories: for example, georeferencing could go under Data Cleaning or Data Capture, or as introductory material in Ecological Niche Modeling. Above all, we note that this is a first more or less complete version, so we invite suggestions, criticisms, ideas, and revisions.

IF YOUTUBE DOES NOT WORK FOR YOU...

Very early in the history of the BITC initiative, the decision was made to house the video component of each course on YouTube. This decision was made largely in light of the near-global availability and massive usership of that site. However, we are quite cognizant that YouTube is not a universal solution to providing access to video content, as access to the site is prohibited by some institutions—and even some

¹ <http://biodiversity-informatics-training.org/>.

countries. Furthermore, additional limitations may be imposed by infrastructural limitations (e.g., insufficient bandwidth) that may limit or impede long views of YouTube-based movies. As such, we are seeking actively to provide multiple access solutions to BITC materials.

As a more failsafe solution to the access challenge, we have ‘published’ the entirety of the BITC video set and ancillary materials on 32GB USB keys. Of course, the challenge now becomes one of availability of this physical copy. For the moment, however, we offer to send the USB key wherever feasible by mail or courier; requests can be addressed to biodivtraining@gmail.com.

HELP US IMPROVE THE BITC

Subtitling

Many users of BITC materials have expressed frustration in the fact that, apart from a few courses that have been replicated in Spanish and Portuguese, the entire body of materials and resources is in English. In Africa, where the BITC in-person courses were held, this English focus was a significant complication for participation of trainees from Francophone countries. This more general failing in the design and content of the Curriculum simply reflects the failings of the two authors of this summary in not having significant language abilities in French, Arabic, Chinese, Russian, etc.

As a step toward a solution to this shortcoming, we are exploring an automated process of adding subtitles to Curriculum materials. We have laid out a protocol that involves the use of voice-recognition software to provide an initial template of English subtitles, which (of course) is quality-controlled by native English-speaking users. The English subtitles will then be placed on a crowdsourcing website (e.g., [dotsub²](http://dotsub.com)), where users worldwide can help facilitate the development and improvement of subtitling into other languages. This effort is as-yet in its infancy, but has proven effective in preliminary tests involving translation into Arabic, Chinese, Spanish, French, and Portuguese. Opportunities to assist in the development of these online resources in multiple languages will be announced via the BITC Facebook group³.

² <http://dotsub.com>.

³ <https://www.facebook.com/groups/BiodiversityInformatics/>.

Problems

Assembling the online curriculum has been a long and difficult process, involving experimentation with multiple video recording and processing technologies, and considerable evolution in the management of video data. As such, it is entirely possible—indeed probable—that problems and errors have crept into this outcome. We request kindly that any problems that users note be reported to us at biodivtraining@gmail.com. We will make every effort to attend to each and every problem that is reported, at least to the extent that is possible.

Future plans

The BITC content is being updated and expanded by a variety of means, particularly via a monthly online seminar series in biodiversity informatics⁴ (e.g., a series on the history of biodiversity informatics currently in process). One important next step that we are exploring is that of certifying the use and mastery of BITC materials by students and young trainees around the world. We are exploring the idea of a masters degree program based on BITC online materials, which would allow trainees to study the material online, develop a project or prepare for an examination, and then be examined by a panel of experts. The degrees would be granted by host universities around the developing world. This model is as-yet under exploration and development, and we expect a fully functioning prototype to be in place by late 2016.

SUMMARY

The BITC represents years of collective work by dozens of persons, including the two of us as co-directors, but perhaps more significantly by our expert course instructors, and more than 134 student trainees representing 23 countries across Africa. The digital teaching resources that have been assembled are unique, in that no other comprehensive teaching resource exists for the field. Needless to say, we are eager to see BITC usership continue to grow and expand.

ACKNOWLEDGMENTS

The BITC exists thanks to generous contributions of many instructors, including Tanya Abrahamse,

⁴ <http://biodiversity-informatics-training.org/webinar-series/>.

Arturo Ariño, David Blackburn, Kyle Braak, Rafe Brown, Bilal Butt, Lindsay Campbell, Jacob Cooper, Vanderlei Perez Canhos, Firkirte Gebresenbet Erda, Eric Fokam, Lee Hannah, Leonard Krishtalka, Rafael Loyola, Enrique Martínez-Meyer, Adolfo Navarro-Sigüenza, Monica Papeş, Richard Pearson, Thiago Rangel, Mark Robbins, Laura Russell, Moses Nsanyi Sainge, Jorge Soberón, Javier Otegui Tellechea, Melissa Tulig, Kumara Wakjira, Kimberly Watson, Christiane Weirauch, John Wiczorek, and Selwyn Willoughby. Funding was generously provided by the JRS Biodiversity Foundation

LITERATURE CITED

- Bisby, F. A. 2000. The quiet revolution: Biodiversity informatics and the Internet. *Science* 289:2309-2312.
- OECD Megascience Forum. 1999. Final Report of the Working Group on Biological Informatics. Organisation for Economic Co-operation and Development, Paris.
- Peterson, A. T., S. Knapp, R. Guralnick, J. Soberón, and M. T. Holder. 2010. The big questions for biodiversity informatics. *Systematics and Biodiversity* 8:159-168.
- Peterson, A. T., J. Soberón, R. G. Pearson, R. P. Anderson, E. Martínez-Meyer, M. Nakamura, and M. B. Araújo. 2011. *Ecological Niches and Geographic Distributions*. Princeton University Press, Princeton.
- Schalk, P. H. 1998. Management of marine natural resources through by biodiversity informatics. *Marine Policy* 22:269-280.
- Soberón, J. 1999. Linking biodiversity information sources. *Trends in Ecology and Evolution* 14:291.

Appendix

Biodiversity Informatics Training Curriculum, version 1.2

Introductory Material

- Biodiversity Informatics
 - [Introduction to the BITC](#)
 - [Introduction to Biodiversity Informatics](#)
 - [What is Biodiversity Data?](#)
 - [Building the Biodiversity Knowledge Graph](#)
- Writing Scientific Papers
 - [Introduction](#)
 - [Choosing a Journal](#)
 - [Technology Assists](#)
 - [Words to Avoid](#)
 - [Figures](#)
 - [Using Color in Science](#)
 - [Tables](#)
 - [Proofing and Editing](#)
 - [Literature Cited](#)
 - [Cover Letter and Reviewers](#)
 - [Authorship](#)
 - [Response to Reviews](#)
 - [Correcting Proofs](#)
 - [Copyright and Open Access](#)
- [Writing Proposals](#)

Biodiversity Data Capture and Initial Enrichment

- Biodiversity Data Capture
 - [Welcome to Ghana](#)
 - [Introduction to BITC](#)
 - [Introduction to Data Capture Course](#)
 - [What is Biodiversity Data?](#)
 - [Biological Data & Collections: Entomology Perspective](#)
 - [Data Sharing Standards: Darwin Core](#)
 - [Data Acquisition: GBIF Data](#)
 - [Exercise: Darwin Core Mapping](#)
 - [Data Models](#)
 - [Authority Files](#)
 - [Persistent Identifiers](#)
 - [Exercise: Data Models](#)
 - [Introduction to Database Management Software Systems](#)
 - [Software Highlight: Arthropod Easy Capture](#)
 - [Software Highlight: Brahms](#)
 - [Software Highlight: Symbiota](#)
 - [Software Highlight: Specify](#)
 - [Data Capture: Strategy](#)
 - [Data Capture from Images](#)
 - [Data Capture: Demonstration](#)
 - [University of Ghana Herbarium Tour](#)
 - [Data Capture: Workflow Analysis](#)

- [Workflows Large 2D](#)
- [Workflows Large 3D](#)
- [Workflows Small 3D](#)
- [Q&A](#)
- [Specimen Imaging: Image Capture](#)
- [Specimen Imaging: Insects](#)
- [Specimen Imaging: Image Processing](#)
- Georeferencing from BITC
 - [Georeferencing: Introduction](#)
 - [Georeferencing: Geographical Concepts](#)
 - [Georeferencing: Map Concepts](#)
 - [Writing Locality Descriptions](#)
 - [Locality Types](#)
 - [Software Example: Georeferencing Calculator](#)
 - [Georeferencing Exercise: Results](#)
 - [Georeferencing Collaboration & Automation Case Studies](#)
- Georeferencing from iDigBio
 - [Collaboration to Automation](#)
 - [Geographic Concepts: Coordinate Systems](#)
 - [Point-Radius Method & Best Practices](#)
 - [Using Paper Maps: Named Places](#)
 - [GEOLocate: the Basics](#)
 - [Introduction to GEOLocate](#)
 - [Using GEOLocate](#)
 - [GEOLocate: Batch Processing](#)
 - [GEOLocate: Collaborative Georeferencing](#)
 - [Collaborative Georeferencing Demonstration I](#)
 - [Collaborative Georeferencing Demonstration Part II](#)
 - [Georeferencing using Specify 6.0](#)

Biodiversity Inventories

- [Welcome to Cameroon](#)
- [Introduction to BITC](#)
- [Introduction to Biodiversity Inventories](#)
- [Inventories vs. Sampling](#)
- [Inventories Example: Mexico & Landscape Change](#)
- [Why Conduct Biodiversity Inventories?](#)
- [Long-term Biodiversity Studies](#)
- [Assessing Inventory Completeness Part I](#)
- [Species Accumulation Curves](#)
- [Richness Estimators: Parametric vs. Non-Parametric Probabilities](#)
- [Assessing Inventory Completeness Part II](#)
- [Results-Based Sampling](#)
- [Conducting Inventories Example: Avifaunal](#)
- [Specimen Preparation: Ornithological](#)
- [Conducting Inventories Example: Herpetofaunal I](#)
- [Conducting Inventories Example: Herpetofaunal II](#)
- [Specimen Preparation: Herpetological](#)
- [Conducting Inventories Example: Botanical](#)
- [Inventory Field Methods: Botanical](#)
- [Inventory Data: Ornithological](#)
- [Observational Data: Ornithological](#)
- [Inventory Data: Herpetological](#)

- [Daily Lists](#)
- [Additional Data & Field Notes](#)
- [GPS Data](#)
- [Data Security](#)
- [Course Wrap-up](#)

Biodiversity Data Cleaning and Data Publishing

- Basics
 - [Introduction to Biodiversity Data Standards](#)
 - [Darwin Core: Introduction](#)
 - [Darwin Core: Detail](#)
 - [Darwin Core: Archives](#)
 - [Darwin Core: Mapping Demonstration](#)
 - [Darwin Core: Archives Demonstration](#)
- Cleaning
 - [Introduction to Biodiversity Data Cleaning](#)
 - [Data Content Considerations](#)
 - [OpenRefine: Exploring Real World Data](#)
 - [Assessing Biodiversity Data Quality](#)
 - [Primary Biodiversity Data: Data Precision](#)
 - [Biodiversity Data Cleaning Tools: Info XY](#)
 - [Why 'Clean' Biodiversity Data?](#)
 - [Error Flagging](#)
 - [Data Cleaning Demonstration: Notepad](#)
 - [Taxonomic Assessment](#)
 - [Data Cleaning Demonstration: Refine and Taxonomic Names Recognition](#)
 - [Software Demonstration: GBIF Names Parser](#)
 - [Software Demonstration: iPlant](#)
 - [Software Demonstration: Refine and Biodiversity Datasets](#)
 - [Software Demonstration: Refine and Taxonomic Names](#)
 - [Repatriating Improvements](#)
- Publishing
 - [Introduction to Data Publishing](#)
 - [Landscape of Biodiversity Data Publishing Tools](#)
 - [Biodiversity Data Networks](#)
 - [Extracting Source Data](#)
 - [IPT: Importing Biodiversity Data](#)
 - [IPT: Map to Darwin Core](#)
 - [IPT: Metadata](#)
 - [IPT: Metadata Comments](#)
 - [IPT: Darwin Core Archives](#)
 - [Registering Data with GBIF](#)
 - [Publishing Sensitive Data](#)
 - [The Realities of Sharing Biodiversity Data](#)
 - [Data Licensing](#)
 - [General Q&A](#)
- Trainee Presentations
 - [African Conservation Centre](#)
 - [Annet Nambuusi](#)
 - [Yvette Umurungi](#)
 - [Tamene Yohannes](#)

Ecological Niche Modeling

- [Introduction to ENM Course](#)
- [Concepts: Distributional Ecology of Species](#)
- [Concepts: Ecological Niche Modeling](#)
- [Concepts: Species Occurrence Data](#)
- [Primary Biodiversity Data: Considerations](#)
- [Primary Biodiversity Data: Quality Checks](#)
- [Environmental Data: Considerations](#)
- [Bias in Occurrence Data and Model Calibration](#)
- [ENM: Model Selection Considerations](#)
 - [Model Calibration: BIOCLIM and Distance Algorithms](#)
 - [Introduction to Complex Algorithms](#)
 - [GARP](#)
- [Model Thresholding](#)
- [No Silver Bullets](#)
- [Model Evaluation: Concepts & Threshold-Dependent Approaches](#)
 - [Model Validation with Small Samples of Species Observation Records](#)
 - [Model Evaluation: ROC](#)
- [M Hypotheses](#)
- [Maxent Outputs Details](#)
- [General Q&A](#)
- Applications
 - [Niche Conservatism](#)
 - [Rare Species](#)
 - [Conservation Planning](#)
 - [Diseases](#)
 - [Invasive Species](#)
 - [Climate Change](#)
 - [Historical Biogeography](#)

Biodiversity Data Analysis

- [Introduction to Data Analysis Course](#)
- Assessing Biodiversity Survey Completeness
 - [Introduction to Sampling Biodiversity](#)
 - [Sampling Theory](#)
 - [Examples: Survey Completeness & Species Accumulation Curves](#)
 - [Comparative Analyses](#)
 - [Software Highlight: EstimateS](#)
 - [Exercise: Abundance & Inventory Completeness](#)
 - [Survey Site Prioritization: Single Species](#)
 - [Inventory Completeness: Survey Gap Analysis](#)
 - [Survey Gap Analysis Example: Kenya](#)
 - [Biodiversity Data Initiatives & Data Quality: GBIF](#)
 - [Exercises: Data Analysis](#)
- [Biodiversity Summaries](#)
- [Macroecological Analyses](#)
- Conservation Prioritization
 - [Introduction and History](#)
 - [Prioritization Example Analyses](#)
 - [Software Highlight: Zonation](#)
 - [Software Exercise: Zonation](#)

Species Descriptions

- [Introduction to Species Descriptions Course](#)
- [New Species Discovery Case Study: Cameroon](#)
- [Nomenclatural Codes](#)
- [Species Descriptions: Nuts & Bolts Part I](#)
- [Systematics Principles](#)
- [Naming Species Example: *Platymantis*](#)
- [Naming Species Example: *Bufo*](#)
- [Naming Species Example: *Caprimulgidae*](#)
- [Naming Conventions & Name Selection](#)
- [Species Descriptions: Nuts & Bolts Part II](#)
- [Species Description Example: Frog 1](#)
- [Species Description Example: Bird](#)
- [Species Description Example: Frog 2](#)

Public Health Applications of Biodiversity Data

- [Introduction to Public Health Applications Course](#)
- [Mapping Diseases Transmission Risk](#)
- [Introduction to Disease Ecology](#)
- [Current PHA Toolkit](#)
- [Q&A: Part I](#)
- [Distributional Ecology Primer](#)
- [How Disease Systems Are Different](#)
- [Ecology and Biogeography of Diseases](#)
- [Q&A: Part II](#)
- [Niche Modeling Overview](#)
- [Example: MPX in Congo Basin](#)
- [Disease Particulars](#)
- [Beyond Suitability](#)
- [Evaluating Risk Maps](#)
- [Example: Buruli Ulcer](#)
- [Example: White Nose Syndrome](#)
- [Example: Mycetoma](#)
- [Example: H5N1 Flu](#)
- [Q&A: Part III](#)
- [Future Constraints](#)
- [Opportunities and Ideas](#)
- [Public Health and Biodiversity](#)

Building Biodiversity Informatics Institutions

- [Welcome to SANBI](#)
- [Introduction to Building Biodiversity Informatics Institutions Course](#)
- [Introduction from SANBI CEO](#)
- [Information is Power](#)
- [Why Start a Biodiversity Institution?](#)
- [Need for Biodiversity Knowledge](#)
- [Biodiversity Policy](#)
- [Biodiversity Information to Policy](#)
- [BI Institution Example: CONABIO Part I](#)
- [The Biodiversity Informatics Initiative Game](#)
- [The Universe of Biodiversity Informatics Players](#)
- [Transitioning the Scope of BI Products: Local to Global](#)
- [Discussion: Part I](#)

- [Policy & Biodiversity Science Example: U.S.](#)
- [Ways Forward in Building BI Institutions](#)
- [BI Demonstration: Data Quality](#)
- [Discussion: Part II](#)
- [GBIF: Part I](#)
- [BI Data Quality Example: Survey Gap Analysis using Kenya GBIF Data](#)
- [GBIF: Part II](#)
- [Biodiversity Informatics Partnerships](#)
- [BI Institution Example: CONABIO Part II](#)
- [BI Partnerships: Centro de Referência em Informação Ambiental \(CRIA\)](#)
- [BI Institution Example: Centro de Referência em Informação Ambiental \(CRIA\)](#)
- [BI Institution Example: University of Kansas Biodiversity Institute](#)

National Biodiversity Diagnoses

- [Introduction to National Biodiversity Diagnoses Course](#)
- [Publication Plans](#)
- [Data Cleaning](#)
- [Data Quality Checks: Taxonomy](#)
- [Data Quality Checks: Geography](#)
- [Data Quality Checks: Temporal](#)
- [Basic Patterns](#)
- [Assessing Inventory Completeness](#)
- [Gaps: Knowledge vs. Data](#)
- [Environmental Gaps in Biodiversity Inventories](#)
- [Environmental Variation Across Study Regions](#)
- [Environmental Variation Demonstration](#)
- [Richness and Endemism](#)
- [Beta Diversity](#)
- [Protocol for National Biodiversity Diagnoses](#)

Biodiversity Conservation Implementation

- Introductory material
 - [Course introduction](#)
 - [Why areas are protected](#)
- Overview presentations
 - [Digital Accessible Knowledge about biodiversity and choosing sites for conservation action](#)
 - [Conservation biology and conservation science](#)
 - [People and protected areas: past present and future](#)
 - [Climate change and wine](#)
 - [Conservation in Ethiopia](#)
- Case studies
 - [Maasai Mara, Kenya](#)
 - [Conservation lessons](#)
 - [Madre de Dios region, Peru](#)
 - [Tehuacan-Cuicatlan Biosphere Reserve, Mexico](#)
 - [Lions in Nech Sar National Park, Ethiopia](#)
- Thematic presentations
 - [History of PA conservation strategies](#)
 - [Design Strategy](#)
 - [Balancing needs of local communities and conservation priorities](#)
 - [Ethiopian protected areas](#)
 - [Dynamic conservation planning](#)
 - [Climate change and South African reserve example](#)
 - [Indicator species](#)

- [Maintaining conservation as climate changes](#)
- [Climate change primer](#)
- [Integrity of protected areas](#)
- Participant case studies:
 - [Egypt](#)
 - [Cameroon](#)
 - [Liberia](#)