2015 Pre-TDWG Training Workshop

21-25 September 2015

Multi-Media University
Nairobi, Kenya
(unless indicated otherwise below)

The JRS Biodiversity Foundation awarded a grant to Tulane University to support 25-30 African participants in the 2015 Biodiversity Information Standards (TDWG) conference in Nairobi, Kenya from 28 September to 1 October 2015, and a pre-TDWG biodiversity informatics training workshop. The aim of the pre-conference training activity is to increase the capacity of African biodiversity specialists to mobilize biodiversity data from their countries and to engage with TDWG. The aim of involving them in the TDWG conference is to catalyze collaborations among African participants and TDWG members that will help to sustain African engagement with TDWG for many years to come.

The following are agendas for all of the types of training that will be offered. We ask participants in the workshop to indicate to us which training you are most interested in receiving. Please email Hank Bart (hbartjr@tulane.edu).

JRS Biodiversity Foundation

PaleoCore Workshop - Nairobi, Kenya
September 2015

Presenter: **Denne N. Reed, University of Texas at Austin (reedd@austin.utexas.edu)**

Day 1 - Data Management Concepts and Fundamentals

I. PaleoCore Overview
1. The goals of digital data management in paleoanthropology
2. PaleoCore's key aims and objectives
   a. Data standards
   b. Data collections tools
   c. Data repository
3. Aims and objectives for the workshop

II. Introduction to spatial data management
1. What is spatial data?
2. Overview of databases and spatial databases
3. Open source software and spatial data management systems
4. Data standards for paleoanthropology

III. PaleoCore basic ontology and terms
1. PaleoCore foundation ontology
2. PaleoCore implementation of Darwin Core terms
3. Comparison with other data standards: Dublin Core, ABCD

IV. PaleoCore user accounts
1. Creating user accounts
2. Setting permissions
3. Using PaleoCore online and offline

V. PaleoCore project initialization
1. PaleoCore target audience, key features and limitations
2. PaleoCore project initialization process
3. Scheduling PaleoCore project initialization

VI. Mapping project schemas
1. Project schema documentation
2. Entering terms and linking to existing terms

VII. Downloading Data from PaleoCore
1. Project permissions and access rights
2. Downloading publicly available data

3. Data download and exchange formats

VIII. PaleoCore and QGIS
1. Introduction to GIS and open-source GIS
2. Connecting to PaleoCore data repositories from QGIS
3. GIS data and exchange formats

IX. Discussion
1. PaleoCore and its place in the digital data management landscape
2. Digital data and access rights
3. Digitizing existing collections vs. incoming collections

Day 2 - Field Data Collection using Mobile Devices

I. Introduction to Mobile Data Collection Hardware (iOS)
1. Overview of different mobile computing platforms: iOS, Android, Windows
2. The menagerie of mobile devices: smart phones, tablets, Total Stations
3. Overview of Global Navigation Satellite Systems (GNSS)

II. Configuring GISPro and related apps
1. Creating projects in GIS Pro
2. Configuring settings
3. Installing and customizing the PaleoCore feature class (data collection forms)

III. Developing and fine tuning a data collection workflow
1. Overview of data collection methodologies
2. Data collection dos, don’ts and gotchas
3. Developing data collection workflows
4. Backup...backup and backup again

IV. Caching map tiles for use in the field
1. Overview of publicly available satellite imagery and map tiles
2. Acquiring digital imagery and maps
3. Caching tiles on mobile devices for use in the field

V. Exporting data from GIS Pro and related apps
1. Exporting data through iTunes
2. Export and exchange formats
3. Creating backups

VI. Uploading data to PaleoCore
1. Importing data to PaleoCore
2. Editing data online

2-Day Data Carpentry Workshop – Nairobi Kenya, September 2015

Data Carpentry's aim is to teach researchers basic concepts, skills, and tools for working with data so that they can get more done in less time, and with less pain.
Our curriculum includes:

- Day 1 morning: Data organization in spreadsheets and data cleaning with OpenRefine
- Day 1 afternoon: Intro to and Data management in SQL
- Day 2 morning: Introduction to R or Python based on attendees' preference
- Day 2 afternoon: We will select a topic based on our attendees' preference

---

The concepts, skills, and tools we teach are domain-independent, but example problem cases and datasets will be taken from organismal and evolutionary biology, biodiversity science, ecology, and environmental science. Data Carpentry's teaching is hands-on, so participants are required to bring their own laptops. (We will provide instructions on setting up the required software several days in advance) There are no pre-requisites, and we will assume no prior knowledge about the tools.

Data Carpentry Agenda Day 1

8:30	Welcome and Introduction
9:00	Data organization in spreadsheets (hands-on)
10:15	tea / coffee break
10:45	Using Open Refine to clean data (hands-on)
12:00	lunch
1:00	Intro to Databases and Getting started with SQL
2:30	tea / coffee break
3:00	Manipulating data in SQL
4:00	Review and Wrap up

Data Carpentry Agenda Day 2

8:30	Introducing Day 2 (Framework)
8:45	Intro to R and R Studio, Starting with data, Data frames (or Intro to Python and iPython Notebook)
10:15	tea / coffee break
10:45	Manipulating data in R (or Python)
12:00	lunch
1:00	Choose: • cleaning data with R • using web APIs • best practices for organizing your project to facilitate Reproducible Research • intro to relevant data standards for biodiversity data
2:30	tea / coffee break
3:00	continue from 1 pm choice
4:00	Review and Wrap up

Funding: iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (Cooperative Agreement EF-1115210). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Data Carpentry has been supported by a contribution from the Moore Foundation and by the Gordon and Betty Moore Foundation's Data-Driven Discovery Initiative through grant GBMF4563 to Ethan White.

Instructors.

- Matthew Collins, Senior Systems Programmer at iDigBio. Matt assisted at the first Data Carpentry workshop and has since also instructed at three Data Carpentry and Software Carpentry workshops as well as iDigBio-specific workshops and with the UF Data Science and Informatics student organization. He has taught Intro to SQL, Intro to Python, Intro to Shell, and Intro to web APIs with R.
- Deborah Paul, Technology Specialist at iDigBio. Deb has assisted at two Data Carpentry workshops, wrote and presented the Open Refine material, and has taught biodiversity data standards and best practices for digitization at several iDigBio workshops. Deb has also facilitated many iDigBio workshops and collaboratively put together the current iDigBio Biodiversity Informatics Skills workshop series. She has experience with SQL in addition to covering the better-spreadsheet and Open Refine lessons.

Assistants.

- Libby Ellwood, iDigBio post-doc. Libby has been to a Data Carpentry workshop and has experience with R and is an experienced teacher. She is a researcher studying phenology and is a key member of the team at iDigBio working on ways to engage citizen scientists in collections digitization.
- Kevin Love, iDigBio Technology Specialist, has assisted at many Data Carpentry and iDigBio workshops and also provides support for web conferencing and recording during workshops.

Before the Workshop

Setup and Software Installation

To participate in this Data Carpentry workshop, you will need working copies of the software described at the setup page. Please make sure to install everything (or at least to download the installers) before the start of your workshop.

Instructors and helpers will be available starting at **8:30am** both days to help with any installation issues. Please bring your own laptop and power cord, and plan to come early if you do not have all the required software installed.

Setup Instructions [we will add this link soon]

Workshop surveys

It's important for the instructors to know who the audience is for each workshop. To give us information about the participants for the workshop, please fill out a brief pre-workshop survey before the workshop.

Pre-workshop survey [we will add this link soon]

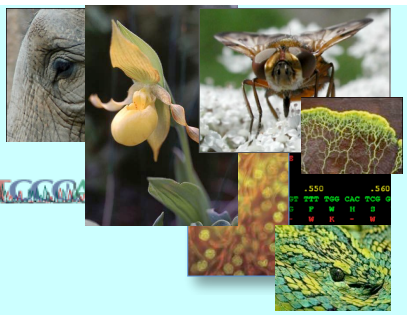
You will also receive a post-workshop survey after the workshop so you can provide feedback and help us gauge the effectiveness of the materials.

Note: The BIS-TDWG 2015 Conference includes a workshop session on the background of the Data Carpentry model and how any group can start using this model for workforce training.

Questions? Please contact either Deb Paul (dpaul@fsu.edu) or Matt Collins (mcollins@acis.ufl.edu) for more information.

Funding: iDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (Cooperative Agreement EF-1115210). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Data Carpentry has been supported by a contribution from the Moore Foundation and by the Gordon and Betty Moore Foundation's Data-Driven Discovery Initiative through grant GBMF4563 to Ethan White.

SPECIFY 6



TRAINING

Session 1: Installation/Implementation/Setup

Pre-installation decision-making process

Conversion:

- Existing Specify 5 users
- New users – conversion, wizard, WorkBench
- Data cleanup, parsing etc.
- Full Specify vs. Specify EZDB
- Mobile Specify – stand alone WorkBench

Installation

- MySQL 5.6
  - GUI Tool interface – database and user management
- Java 7
- Specify
  - Specify
  - Specify1G
  - SpWizard
  - SpBackupRestore
  - SpiReports
  - ImportFileSplitter
  - DataExporter
  - SchemaExporter

Data model and hierarchy information

- Institution, Discipline, Division, Collection, User Group, User

Wizard database creation

- MySQL root username/password and database name
- Master U/P
- Security
- CM username and password
- Institutional information
- Accessions
- Accession Number setup (global only)
- Field formatting editor
- Storage tree definition setup

- Division information
- Discipline type and name
- Taxon tree definition setup
- Pre-loading of taxon tree data
- Geography tree definition setup
- Collection information
- Catalog number setup
- Accession Number setup (collection level)
- Summary and build

Post-installation decision-making process

- Setting up additional disciplines, collections
- Trees and tree definitions
- Form customization
- Reports
- Data conversion
- Consortia

Session 2: Program interface, layout and navigation

Specify login

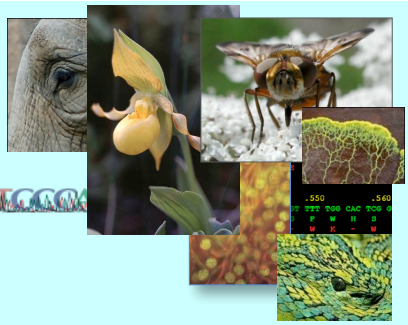
- Username and password (from Wizard or provided)
- Database name (from Wizard)
- Port
- Generate key process

Interface, layout and navigation

- Main menu
- Task bar
- Simple search
- Side bar
- Main pane

- Tabs
- Drag and drop
- Record sets

SPECIFY 6



TRAINING

Session 3: Data entry and editing

Entering data

- Tables and sidebar configuration
- Field types
  - Text
  - Number
  - Formatted number
  - Date
  - Formatted/partial date
  - System pick list
  - User customizable pick list – user defined, field, table
  - Query Combo Box
  - Required fields
- Sub form types
- Carry forward
- Save and New
- Auto numbering
- Locality features
  - Lat Long preference pane
  - GeoLocate
  - Google Earth
  - WorldWind
  - Clone tool
- Series data entry **- limit 500 records**

Editing data

- Edit and View modes
- Batch identify

Session 4: Working with Trees

- Taxonomy
  - Expanding
  - Find
  - Navigating tree
  - Split screen
  - Add node
  - Edit

    - Move
    - Synonym
    - Associated Collection Objects and numbering
    - Merge
- Geography
- Storage
- Paleo
- Tree definitions
- Locking and unlocking trees

Session 5: Querying data

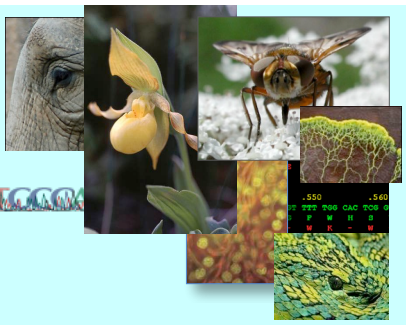
Searching for data using the Query builder

- Tables and Sidebar configuration
- Adding fields
- Adding aggregated or formatted tables
- Operators, Criteria, Sort, Show, Prompt
- Changing order of fields
- Removing fields
- Search Synonyms, Distinct, Count
- Wildcards (*)
- Higher level tree rank searches
- Result bar options – 20,000 row limit
    - Record set, form view, print grid, export, Reports

Searching for data using the Simple search

- All vs. Distinct table
- Primary vs. related searches
- Wildcards (*, ‘, “)
- Configuration
- Result bar options
    - Record set, form view, print grid, export, Reports
- Global search **– coming in future release**

SPECIFY 6



TRAINING

Session 6: WorkBench

- Import data 4000 row limit
- Import file splitter
- Data mapping
- Find – Ctl-F
- Add, remove rows
- Edit cells
- Map – satellite and Google Earth
- GEOLocate
- Convert Geocoordinate formats
- Export data set
- Images
- Carry forward
- Form view and customization
- Validation
- Create record set from data set **– coming in future release**
- Uploading
    - Validate
    - Upload
    - Undo
    - Data integrity (future release)
- Reports

Session 7: Interactions (Loans/Gifts/Borrows etc.)

- Tables and sidebar configuration
- Accession and repository agreement
- Loan
    - Adding preparations – manual vs. record set
    - Returning loans
- Gift
- Borrow
- Exchange in and out
- Information request

Session 8: Attachments and images

Viewing attachments

- Attachment browser
- Query

Attaching attachments

- Import attachments
- Import attachments mapping file
- Drag and drop

Session 9: Preferences and Security

Editing Preferences

- Formatting
- System
- Trees
- Email
- Taskbar
- Google Earth
- MySQL
- Login Dialog

Main menu

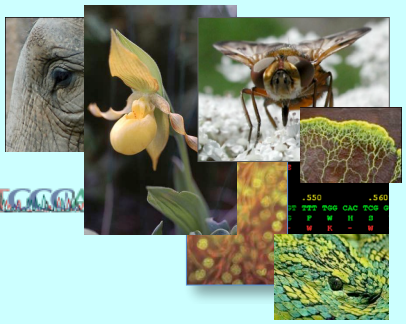
Working with Security

- Security levels – Manager, Full access, Limited Access, Guest
- Multiple disciplines
- Creating new users – new and existing
- Group permissions
- Tables – View, Add, Modify, Delete
- Tools
- Preferences

Security Wizard

- Used to check security preferences – master U/P

SPECIFY 6



TRAINING

Session 10: Additional/Advanced topics

Georeferencing and visualizing data with Plugins

- Google Earth
- GEOLocate

Producing reports and labels

- SpecifyiReports
- Construct query
- Link to SpecifyiReports
- Adding fields
- Specify services

Manipulating the Schema config

- Captions
- Tables
  - Caption
  - Hiding tables
  - Usage notes
  - Table display format
  - Table aggregation
  - Web links
- Fields
  - Caption
  - Hiding field
  - Usage notes
  - Is Required
  - Field formatting
- Localization – different languages
- WB schema config

Containers and Relationships

- Relate collection objects in same and different collections

Customizing forms

- XML files
- User, User type, Collection, Discipline, Institution
- [discipline].views.xml (config/[discipline] directory), common.views.xml (config/common directory), global.views.xml (config/backstop directory)
- Eclipse (or other XML editor)
- Columns
- File structure – views and viewdefs
- Finding correct table
- Specify reload forms
- Specify debug forms

Publishing Specify data with GBIF IPT

- IPT installation
    - Apache Tomcat
    - Memory allocation
    - IPT WAR file
    - Geoserver installation and configuration
- Specify configuration
    - Darwin Core schema selection
    - Mapping of fields
    - Schema export application of tab delimited data
- IPT and Specify integration
    - Source file import and upload
    - Viewing of data
    - Publishing data

SGR

Lifemapper

Specify web search client

- Schema mapper
- Data export
- Apache and Solr setup
- Download Specify web portal files
- Edit Solr files
- Customizations

Future directions

- Specify 7 Thin client

Training on Data cleaning, Data quality and Data publishing through the GBIF using the IPT.
Nairobi, September 2015

Organizing committee : melecoq@gbif.fr, pamerlon@gbif.fr, hbartjr@tulane.edu

Training schedule

Day 1

09:00 General Introduction: Short GBIF France presentation + training agenda

INTRODUCTION AND DATA QUALITY

09:20 Introduction about Data Quality and Fitness for Use

10:50 Coffee Break

11:00 Methods and tools to increase the quality of biodiversity data

12:30 Lunch Break

13:30 Data standard summary and introduction to the Darwin Core Archive

15:00 Coffee Break

GBIF : DATA PUBLISHING AND DATA USE

15:30 How to publish occurrence data and datasets to the GBIF (part 1)

17:00 End of session

Day 2

09:00 How to publish occurrence data and datasets to the GBIF (part 2)

10:30 Coffee Break

10:45 How to discover and extract data through GBIF.org

11:45 Introduction to the Data Paper

12:45 End of training

Presenters: Nelson Rios and Hank Bart (2 days)

Day 1

9:00-9:15	Participant Introductions
9:15-10:15	Introduction to georeferencing • What is georeferencing • Basic geographic concepts • Polymorphic representations • Paper maps • Extracting coordinates • Coordinate conversions • Helpful online resources, Google Earth etc.
10:15-10:30	Break
10:30-11:30	GEOLocate overview
11:30-12:00	Validation
12:00-1:00	Lunch
1:00-2:00	Using GEOLocate Web Client(s)
2:00-2:45	Georeferencing exercises
2:45-3:00	Break
3:00-4:30	Georeferencing on your own

Day 2

9:00-10:15	The Collaborative Georeferencing Framework • Setting up user accounts • Introducing the data management portal • Using the Collaborative georeferencing web client
10:15-10:30	Break
9:00-10:15	The Collaborative Georeferencing Framework (continued)
10:30-12:00	Georeferencing on your own, questions etc.
12:00-1:00	Lunch
1:00-2:00	Interoperability, web services, advanced topics.
2:00-2:30	Wrap up, questions etc.

BHL Africa Workshop

National Museum of Kenya

Presenters

Anne-Lise Fourie, SANBI / BHL
Martin Kalfatovic, Smithsonian Libraries / BHL
Carolyn Sheffield, Smithsonian Libraries / BHL
Grace Costantino, Smithsonian Libraries / BHL
Julia Blase, Smithsonian Libraries / BHL

Agenda

- **Overview of BHL**
  - Structure of BHL and BHL Global
  - BHL Africa, Affiliate status, and relation to consortium structure
  - Communication channels and strategies
  - Discussion of 4-5 year goals for BHL Africa consortium participants
- **BHL website**
  - Overview, contributors' and users' perspectives
  - Titles and Items, Value-added services
  - How related to Internet Archive
- **Collection Development**
  - Scope: What do we mean by biodiversity
  - Deduplication, checking BHL before uploading
  - Copyrights and Permissions
- **Digitization Workflow**
  - Standards of Digitization
  - Gemini workflow
  - Metadata - including what needs to be in document so it is recognized as a BHL Africa collection
  - Macaw
- **Social Media and Marketing**
  - Strategies to promote
  - How we can collaborate
- **Wrap-up**
  - Revisit 5 year goals – any changes based on information covered in training?

iPlant Collaborative - Cyberinfrastructure for life sciences

Presenter: Ramona Walls, The iPlant Collaborative

Do you need to share your data, images, and analyses with collaborators at multiple institutions? Do you work with big data? Have you developed a new algorithm that you want to make available to anyone to use, regardless of whether or not they have command line experience? The iPlant Collaborative (http://www.iplantcollaborative.org/) provides free cyberinfrastructure to all biologists and bioinformaticians to address these very challenges. iPlant is an NSF-funded initiative with a mission to facilitate the transformation of life sciences research and education through computational infrastructure and expertise. Despite the name, iPlant's scope includes any life sciences research, be it genomic or ecological, in plants, animals, or microbes, from single-researcher investigations to community-wide collaborations. This introductory presentation will provide an overview of the tools and services available through iPlant, with an emphasis on their utility to biodiversity researchers. These include: data storage, sharing, and metadata mark-up; cloud-based computing through Atmosphere; web-based access to dozens of applications through the Discovery Environment; data publishing through the iPlant Data Commons; Application Programming Interfaces (APIs); image management and analysis with Bisque; and Educational, Outreach, and Training (EOT) resources. A hands-on workshop will be help during the regularly scheduled TDWG meeting, for those who want to learn more. For the complete workshop agenda, please see http://bit.ly/1euy02s.