

# Community Resources for Digitization Support

BIS (TDWG), GBIF, Data Carpentry



Joanna McCaffrey, iDigBio Biodiversity Informatics Manager

Scripps Institute of Oceanography

Wednesday, 3rd February 2016, La Jolla CA



*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.*

# Biodiversity Information Standards

- BIS replaced in name
  - TDWG** – Taxonomic Databases Working Group
- started in 1985
- International standards body
  - Meets annually
  - Work is accepted by GBIF on a periodic basis

## **BIS Mission**

- Develop, adopt and promote data standards and guidelines for the recording and exchange of data about organisms
- Promote the use of data standards through the most appropriate and effective means
- Act as a forum for discussion via meetings and publications

## BIS Community of Interest

biologists taxonomists  
geneticists  
information  
entomologists scientists  
technologists  
zoologists library ecologists

## Metadata standards

- Darwin Core – an extension of the Dublin Core for biodiversity information
  - Intended to facilitate the sharing of information about biological diversity – label data and beyond
  - Definition of terms - defines a namespace and includes a glossary of terms, and a recommended reserved vocabulary
  - Support for specimen data, observations

## Darwin Core extensions

- Audubon Core –
  - Represents media related to specimen records
- Other TDWG recommendations
  - BioCase
  - ABCD
  - OBIS – extends Darwin Core

## Data Carpentry

- Power tools for biodiversity informatics
  - one comprehensive model for teaching biodiversity informatics data management, research skills, and training instructors

## Student-organized Teaching

Data Science and Informatics



student organization @ University of Florida

- Founded in spring 2015
- Engineering students saw need for Python and data analysis skills for their careers
- Membership from many colleges (over 230 students)
- 4 three hour trainings/semester + seminars



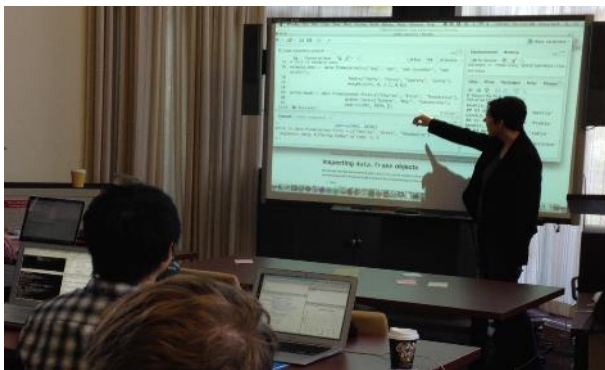
## What led to establishing Data Carpentry?

- COLLAB-IT did a survey about Informal Skills / Needs Assessment
- Came from nascent Software Carpentry movement – software skills for scientists
- BRAEMBL report (Australia, 2013):
  - <http://braembl.org.au/news/braembl-community-survey-report-2013>
- Biggest gap is expertise
- Biggest need is training

## **Data Carpentry: Moore Foundation Grant Focus**

- Expand into new domains
  - Life sciences
  - Physical sciences
  - Library and Information and Social Sciences
- Train thousands of scientists
  - Skills for large scale data
- Scale our instructor pool
- Develop strategies for delivering workshops online

## Typical Data Carpentry Workshop



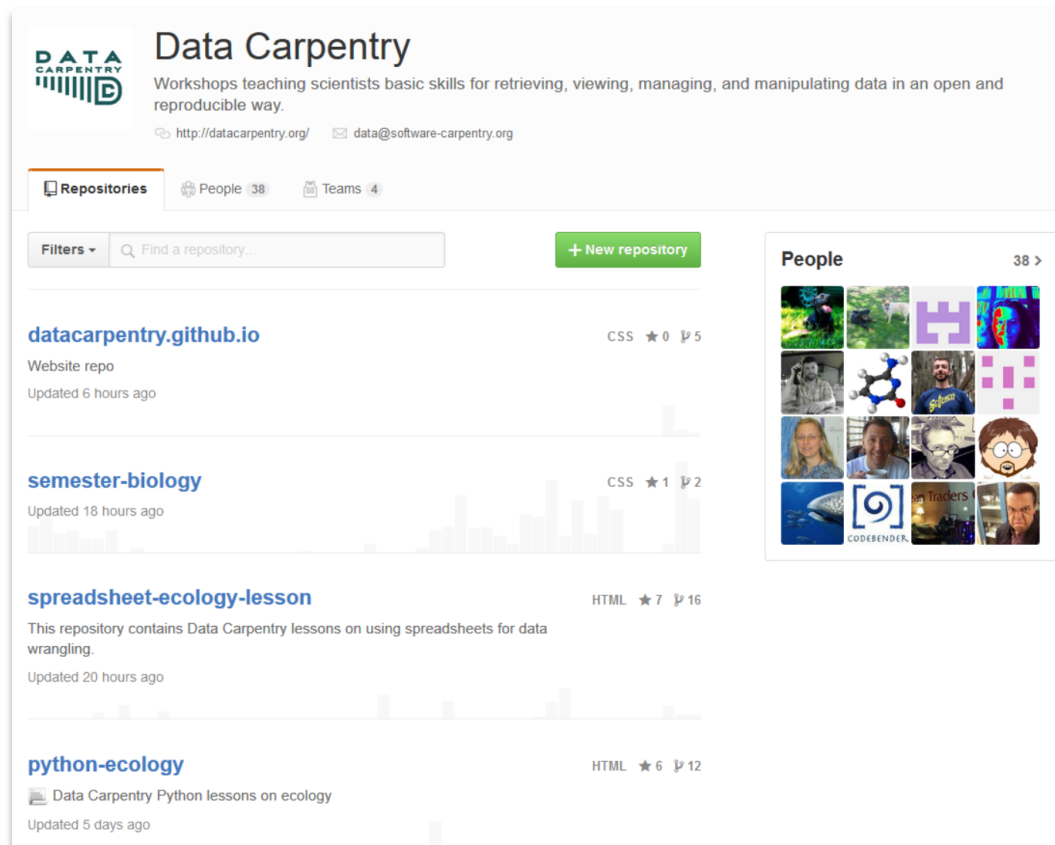
- 2-Days
- Better spreadsheet skills
- Introduction to SQL
- Introduction to R and Python
- Introduction to Analysis and Visualization using R

## Workshop & Community Feedback

- wish I'd known this stuff before graduate school!
- ...“the course is well suited for beginners; I received the courage to venture and learn more about R”
- “I feel I can start using R again”
- “I appreciate all the hands-on work”
- “the class is well organized and easy to follow”
- “really great working through all the examples/challenges, lots of examples = lots of practice = understanding”
- we need to hire a data manager
- *I can now send cleaner, more standardized data to GBIF*

# Lessons online for anyone to use

- Data Carpentry Lessons on GitHub



The screenshot shows the GitHub profile for Data Carpentry. The profile header includes the Data Carpentry logo, the name "Data Carpentry", and a description: "Workshops teaching scientists basic skills for retrieving, viewing, managing, and manipulating data in an open and reproducible way." It also lists the website "http://datacarpentry.org/" and email "data@software-carpentry.org".

Below the header, there are tabs for "Repositories", "People" (38), and "Teams" (4). A search bar with "Find a repository..." and a "+ New repository" button are present. The repository list shows:

- datacarpentry.github.io**: Website repo, updated 6 hours ago, CSS, 0 stars, 5 forks.
- semester-biology**: Updated 18 hours ago, CSS, 1 star, 2 forks.
- spreadsheet-ecology-lesson**: This repository contains Data Carpentry lessons on using spreadsheets for data wrangling. Updated 20 hours ago, HTML, 7 stars, 16 forks.
- python-ecology**: Data Carpentry Python lessons on ecology. Updated 5 days ago, HTML, 6 stars, 12 forks.

On the right side, there is a "People" section with 38 members, displaying a grid of profile pictures.