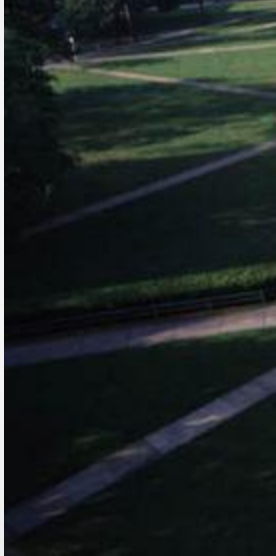




Collection Databases

Greg Riccardi
Florida State University
griccardi@fsu.edu

Florida State University



My Background

- Computer scientist
 - Research and outreach activities in scientific information management
 - Emphasis on technology transfer from IT to domain science
- Information management in experimental physics: 1990-2005
 - Supported collaborative information management at a Dept. of Energy Lab
- Director of the Morphbank project
 - <http://www.morphbank.net>
- Co-PI iDigBio
 - PI of FSU subproject, Director of iDigInfo institute
 - Emphasis at FSU is Digitization support and interaction with digitization projects

Why

- Why have a database?
 - Collection management
 - Information collection, distribution and management
- Why have a session on databases?
 - Learn about issues that will have a major impact on success

Overview

- What to store in a database
- How information is organized
- Some examples of database contents
- Interpreting the meaning of data properties
- Issues of adoption of database technology

What to store in a database

- Collections
 - Characteristics of each collection
 - ala BCI or Darwin Core
- Specimens
 - Not physical specimens
 - A digital proxy for each specimen
 - The properties of each specimen
- Related objects
 - Photographs
 - Videos
 - Sounds
 - Taxa
- Relationships between things
 - Image of
 - Determined as
 - Same as
 - Is a part of

How information is organized

- In a database
 - Tables
 - Rows
 - Foreign keys
- Conceptually
 - Classes, Objects, and properties
 - A table is a set of objects all of the same class
 - A row is the properties of an object
 - A foreign key is a relationship with another object

Some examples of database contents

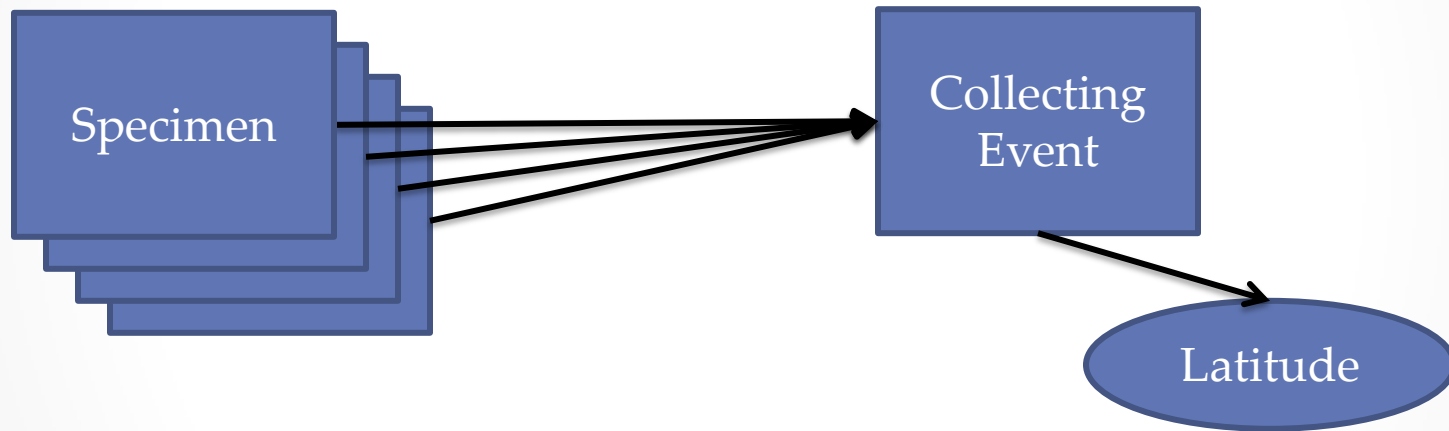
- Different databases store different information in different ways
- Example 1: GBIF specimen
- Example 2: Morphbank specimen
- Example 3: Morphbank image

Exercise 1

- Choose an online biodiversity database
 - Symbiota.org, Arctos.org, NYBG, Berlin Botanical Garden (biocase.org), Flora de Brasil, Morphbank image, etc.
- Search for one or more objects and display the most detailed view of the contents
- List
 - The type of object
 - A Darwin Core property
 - Properties that are unusual
 - The relationships to other objects
- Example from Morphbank specimen
- Discussion follows

Example 4: Specify Tables

- Specify includes a “collecting event” as an object
- A specimen comes from a collecting event
- Many specimen properties are stored in the collecting event table



- In this model, does a specimen have a latitude?
- Yes the choice of table structure does not imply meaning
- Don't be overly concerned with table structure

Interpreting the meaning of data properties

- Example Darwin core properties
 - Scientific name and phylum
 - Scientific name represents determination
 - Phylum represents classification
- Example: geo-location properties
 - Of a preserved specimen
 - Of a live organism
 - Of an image of an organism
 - Of a fossil

Issues of adoption of database technology

- Many groups are eager to create their own information system components
 - It's better to own the system and all of its components
 - We call this NDH (not developed here)
- Specify has been under continuous development for more than 20 years
- Morphbank received \$2.4 million for a 4 year period
- Take advantage of the efforts of many smart dedicated people
 - Find a way to use tools that others are using

Summary

- Databases are good
- Data is hard to organize and maintain
- Software is hard to create and harder to maintain
- Adopt NIMB: Not in my backyard
 - Make some else accept the mess and the responsibility
 - Use community software