Free (or low cost) and open source database options for collections digitization

Gil Nelson
Integrated Digitized Biocollections
Florida State University

OBFS Biological Research Station Collections Digitization Workshop
OBFS Annual Meeting
Rocky Mountain Biological Laboratory
16 September 2015

This material is based upon work supported by the National Science Foundation under 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.

Choices…..

Excel (or other spreadsheet)

Microsoft Access

Specify

Symbiota

Advantages

- Familiar
- Easy to use
- Generally available
- Good as an intermediate step for analysis, cleaning, data prep
- Easily modified

Disadvantages

- Not normalized
- Not relational
- Lacks field constraints
- Difficulties converting to other formats
- Doesn’t use key fields, indexes
- Not dependable for managing globally unique identifiers

Advantages

- Supports data normalization
- Relational (multiple related tables)
- Supports keys, indices, and GUIDs
- Generally available
- Datasheet view resembles Excel
- Searchable via Structure Query Language (SQL) syntax
- Advanced export features
- Easy way to process data before exporting to more sophisticated, online systems

Disadvantages

- Requires echnical expertise to utilize fully
- Potentially steep learning curve
- Can require significant development time and interface design
- Not accessible online (single-use database)

Specify 6

Advantages

- Built on MySQL (relational)
- Long history
- Help desk (Theresa Miller)
- Free
- Powerful tools
- Customizable interface and schema
- Allows multiple users
- Has a web interface for serving data

Disadvantages

- Learning curve can be steep
- Requires a server for multiple users
- Not web based for data management
- Can be difficult to install

Symbiota Promoting Bio-Collaboration

Advantages

- Built on MySQL (relational)
- Long history
- Open source
- No installation
- Web based: managing and serving data
- Supports images
- Free
- Powerful data management tools
- Direct data mobilization to iDigBio and GBIF
- Easy to use
- Intuitive
- Built in OCR and georeferencing
- Allows multiple users from remote locations

iDigBio Integrated Digitized Biocollections

iDigBio

Integrated Digitized Biocollections

North American Network of Small Herbaria

iDigBio
Integrated Digitized Biocollections