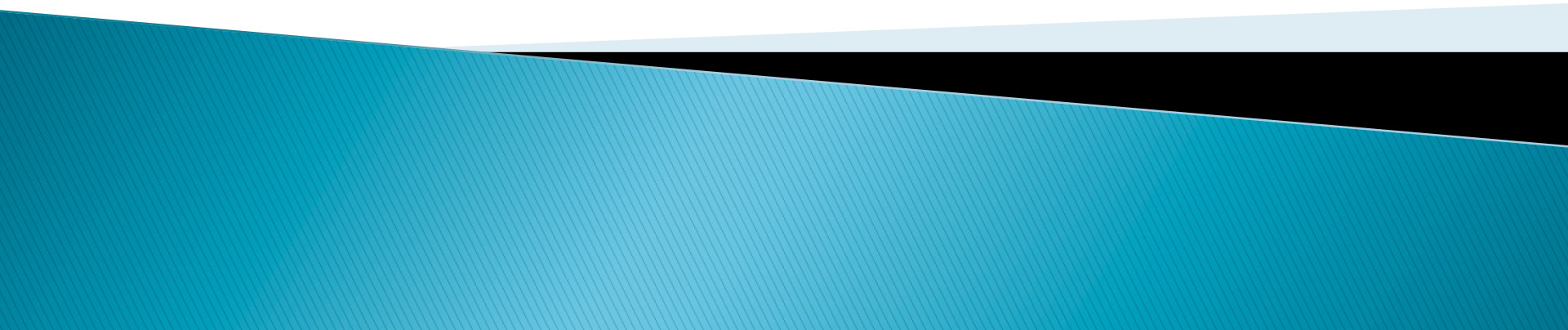


Understanding Data

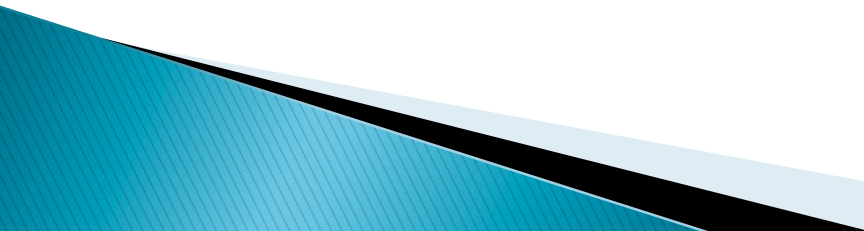
Edward Gilbert



Data

- ▶ Specimen data
- ▶ Determination history
- ▶ Imaging specimens
 - Image archive (TIF, RAW, DNG)
 - Web images (JPG)
 - Skeletal data
- ▶ Other
 - Genetic links, voucher info, references, general annotations, comments, etc.

Data Management

- ▶ Specify, KE-Emu, Brahms
 - ▶ Symbiota, Arctos, CollectionSpace
 - ▶ Best is relative
 - ▶ Needs:
 - Easy to use
 - Affordable
 - Appropriate IT support
 - Adherence to standards
 - Publishing, data extraction
 - Protocols for quality control and cleaning
- 

Specify

- ▶ Locally installed software
- ▶ Full service
- ▶ IT support needed
 - Install and update
 - Server
 - MySQL database server
 - Backups
 - Image support



KE-EMu

- ▶ Comprehensive museum management
- ▶ Costs money
 - Initial data loading, training, license
- ▶ Some IT support needed
 - Server, images, software install and updates
- ▶ Data publication support?

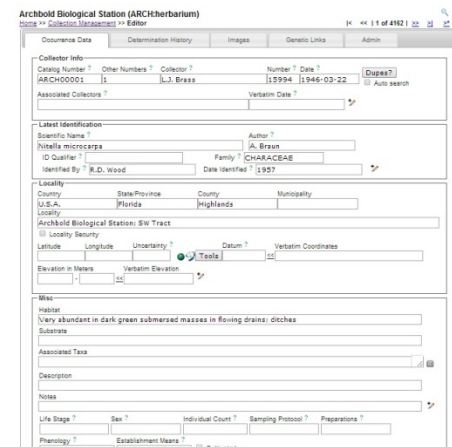


NANSH (Symbiota Portal)

- ▶ Publishing with management options
- ▶ Data stored in cloud
- ▶ Some services not available
 - E.g. custom invoices
- ▶ Live data / Online presences
- ▶ Distributed support
 - Backups automatic
 - Image support
 - No software updates



The screenshot shows the homepage of the North American Network of Small Herbaria. The header features the title "North American Network of Small Herbaria" in a stylized font, accompanied by a photograph of yellow flowers. Below the header, there is a navigation menu with links such as "Homepage", "Search Collections", "Dynamic Checklist", "Dynamic Key", "Image Library", "About Network", "Resource Links", "Symbiota Help Page", "Other Networks", "Great Plains", "Intermountain", "MABA Flora", "SEINet", and "SEINCC". The main content area includes a "Welcome Edward Gilbert" message, a "Plant of the Day" section with a photograph of a green plant and a "What is this plant?" quiz link, and a paragraph of text explaining the network's mission and the importance of biodiversity data.

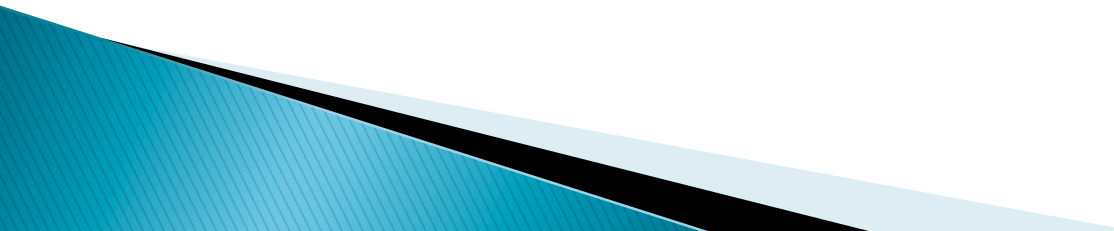


The screenshot displays the data entry form for the Archbold Biological Station (ARCHherbarium). The form is organized into several sections: "Occurrence Data" (including Occurrence, Determination History, Images, Genetic Links, and Admin), "Collector Info" (with fields for Catalog Number, Other Numbers, Collector, Number, Date, and Associated Collectors), "Latest Identification" (with fields for Scientific Name, Author, ID Qualifier, Family, Date Identified, and Locality Security), "Locality" (with fields for Country, State/Province, County, Municipality, and specific locality details), "Elevation in Meters" (with fields for Elevation and Verbatim Elevation), "Misc" (with a text area for description), "Associated Taxa", "Description", "Notes", "Life Stage", "Sex", "Individual Count", "Sampling Protocol", "Preparations", and "Phenology". The form is designed for detailed data collection and management.

Other Management Systems

- ▶ **Arctos**
 - Shared cost
 - Browser based system
- ▶ **Brahms**
 - Software purchase
 - Local install and updates
- ▶ **CollectionSpace**

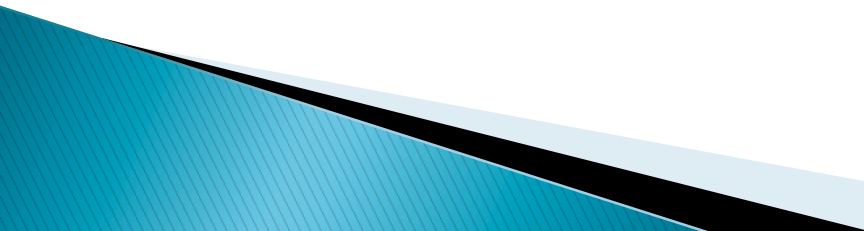
Custom Database

- ▶ Local IT support needed
 - Future?
 - ▶ Image management?
 - ▶ Regular local backup system?
 - ▶ Standard compliance?
 - ▶ Data publishing?
 - ▶ GUID support?
 - ▶ Avoid Excel “databases”
- 

Data Augmentation & Cleaning

- ▶ Authority lists
 - Taxonomic Names
 - TROPICOS, ITIS, Catalog of Life, EOL
 - Geography
 - Collectors
- ▶ Georeferencing
 - GeoLocate
- ▶ Get data online!
 - Use the data
 - Let others contribute
 - Local experts

Data Publishing

- ▶ Darwin Core Archive
 - ▶ IPT – Integrated Publishing Toolkit
 - ▶ Data aggregators
 - iDigBio
 - GBIF
 - BISON
 - Symbiota
 - EOL
 - ▶ Why publish?
 - ▶ Why are there so many options
- 

Data Exchange Standards

- ▶ Darwin Core
 - ▶ Darwin Core Extensions
 - ▶ Darwin Core Archive
 - ▶ Audubon Core
 - ▶ Global Unique Identifiers (GUIDs)
 - ▶ Enables data sharing, integration, and use
 - ▶ Who should care?
- 