

Evaluating and Selecting a Database

Greg Schneider
UMMZ Collections Manager
Division of Reptiles and Amphibians



- History: Computing in Herpetology at UMMZ
- Databases Mean Different Things to Different People
- Considerations / Recommendations

History of Databases in Reptiles and Amphibians

- Taxir (1974)
- Foxpro2.6 (1994)
- FilemakerPro 7, 11 (2004), FMP14 server (2015)
- KeEMU (considered briefly in 2014)
- Specify (next migration decided on in 2015)

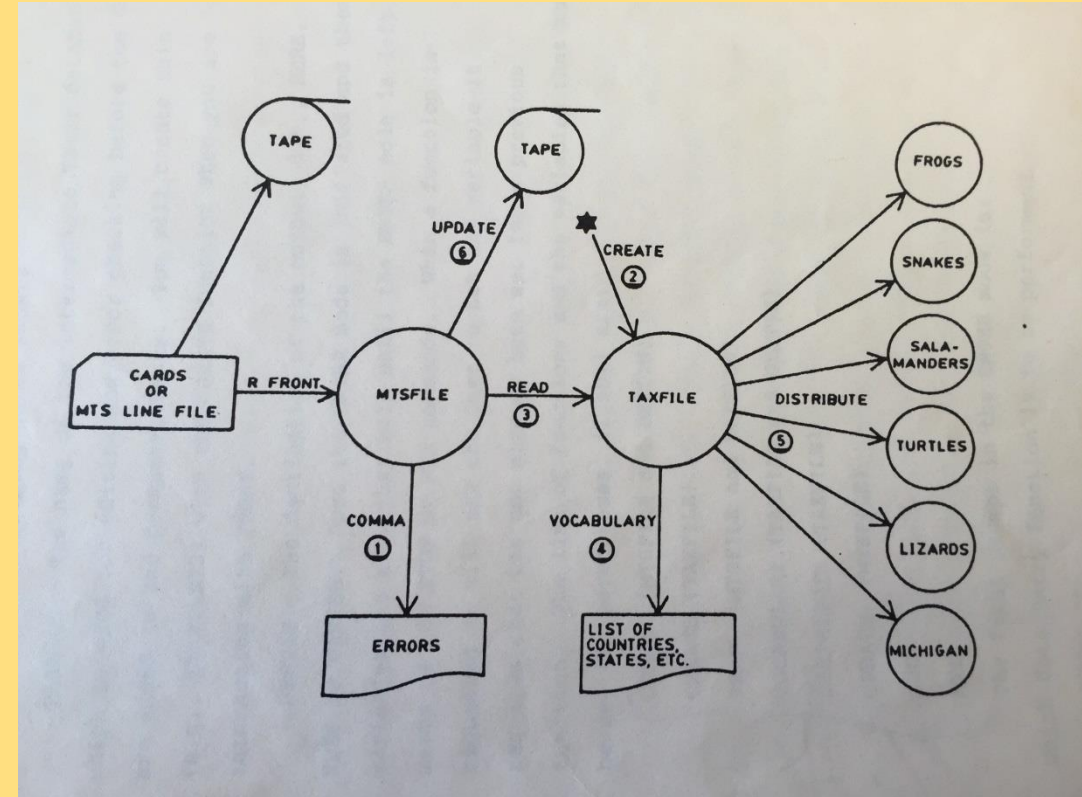
History of Databases in Reptiles and Amphibians

- Taxir (Taximetrics Information Retrieval system)
 - NSF funding to computerize catalogue in mid 1970's – “EDP”
 - Developed by **George Estabrook** and Bob Brill at UC Boulder –
 - Both came to University of Michigan in early 1970's
 - Taxir evolved independently at U of M and UC Berkeley (and Washington State)
 - Data entry on punch cards on the earliest computers



Taxir on MTS

- Speed and Space were big considerations
- Time was money
- Comparisons of descriptors (fields) based on Boolean expressions
- Not every descriptor was registered (indexed) – we had 12 “queriable descriptors” – catalogue number, taxonomy, and higher geography
- Interaction with magnetic tapes for storage of complex text fields

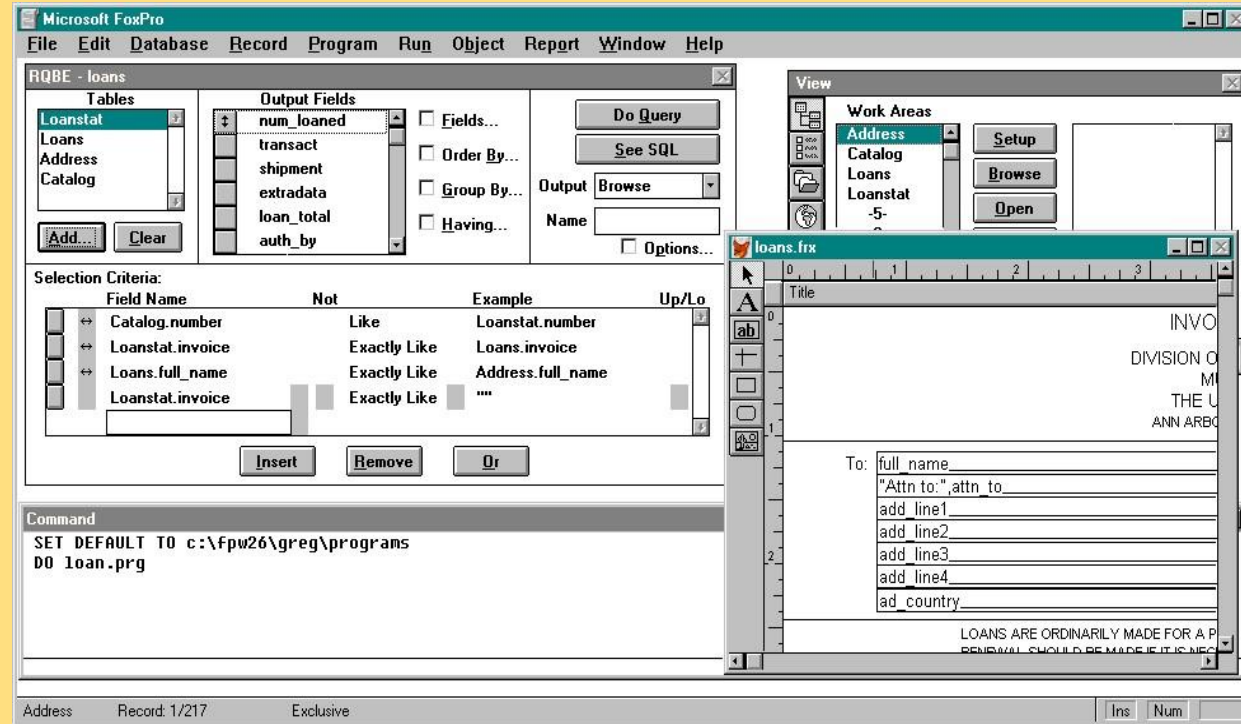
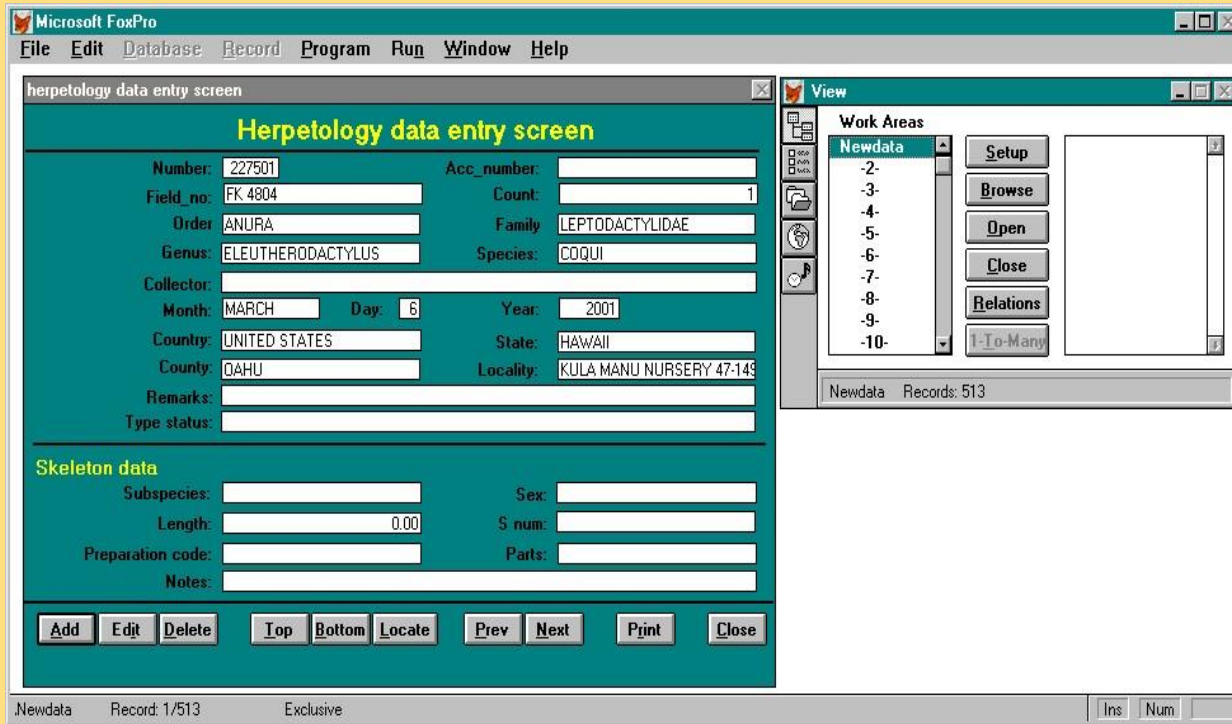


EDP Workshop 1977

- Ours was a pilot project funded by NSF
- Discussion topics included: “community needs” and does EDP fulfill them, “interest”, “standardization”, “legal and ethical problems”, what EDP projects would be most beneficial to the community”
- Two major conclusions:
 - **Primary role of Museum Collections includes making the information associated with them available to users.**
 - **EDP is an economical and quick way to make these data available**



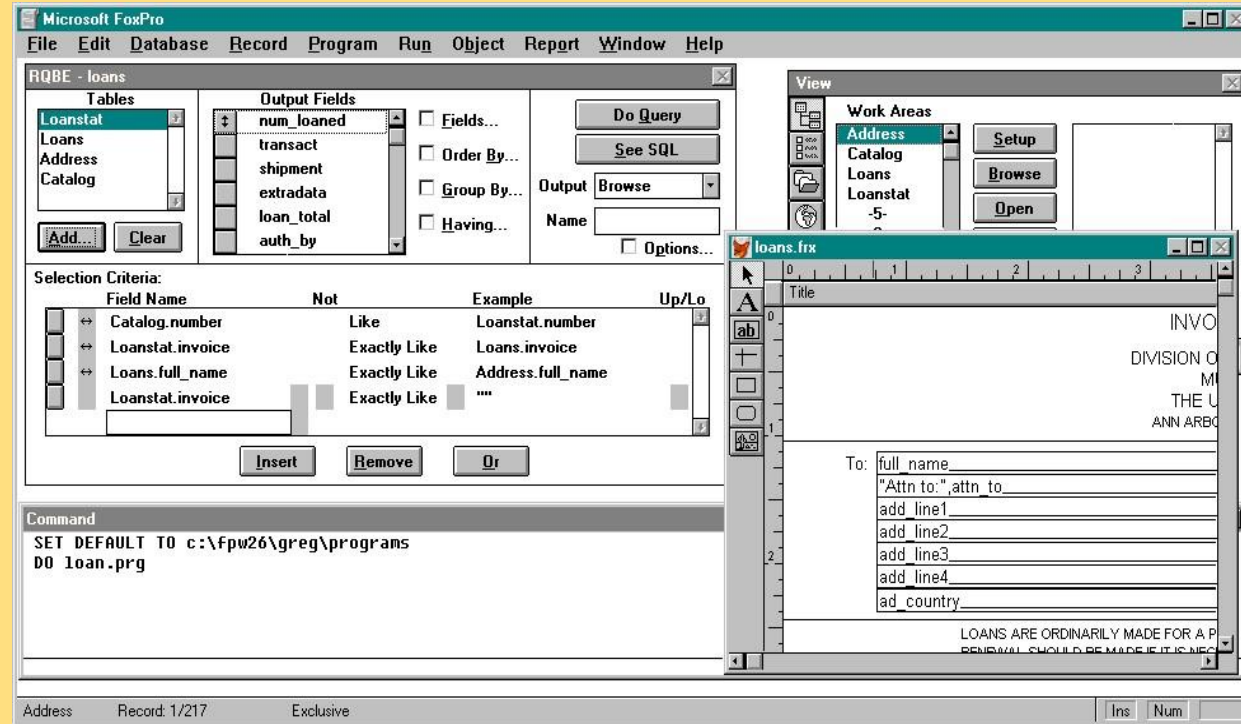
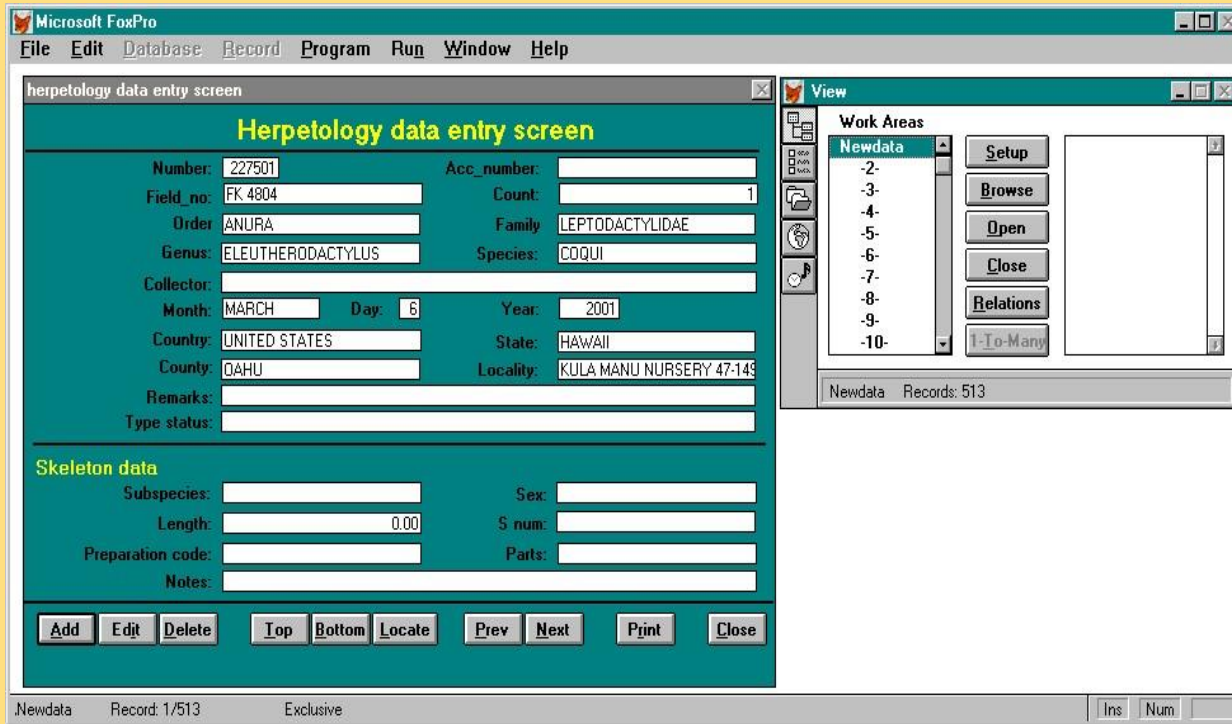
Foxpro2.6 (1994)



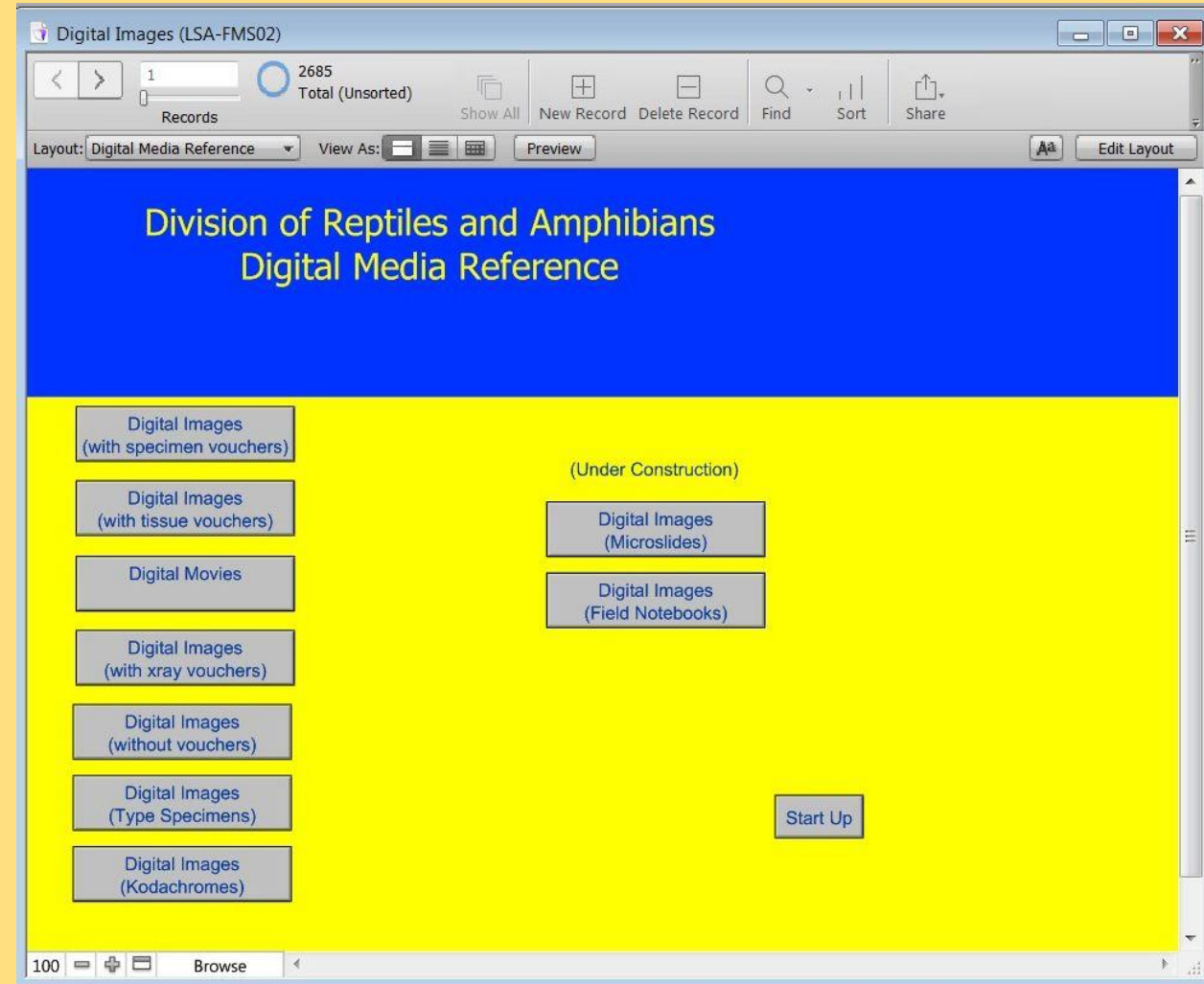
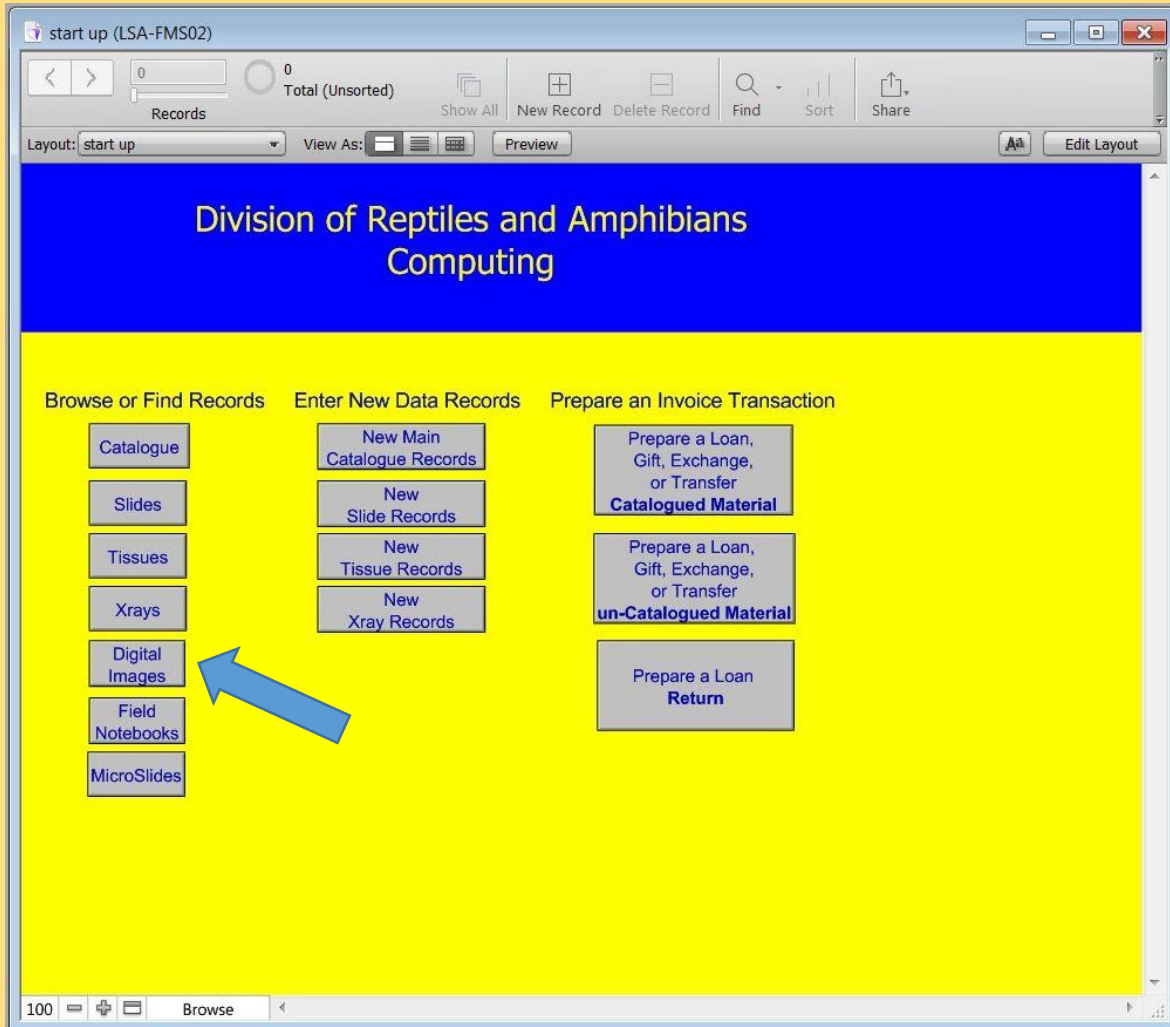
Why Foxpro

- Relational
- Robust -1 billion records / 252 Tables
- Stable – programming in dbase – business applications
- Inexpensive - \$99
- Multiplatform – PC and Mac
- Versatile – especially import/export formats and reports
- Programming ability – run external programs (especially dbase)
- Not Reliant on others (viewed as a positive)

Foxpro2.6 (1994)



FilemakerPro – started transitioning 2004



FilemakerPro : 2005 - present

- Evolving computers and operating systems necessitated another change
 - Lack of software support by Microsoft
 - Incompatibility with newer Windows operating systems
 - Change from 32 bit machines to incompatible 64 bit ←
- Similar reasoning in decision making (as Foxpro) ie. Relational, Robust, Stable, Inexpensive, Versatile, etc.
- But, also
 - Networkable
 - Container fields for images
 - Ability to serve published version on internet

E.O. WILSON

"Exploring the Complexity of Life"

Life Sciences Grand Opening

Convocation

May 14, 2004

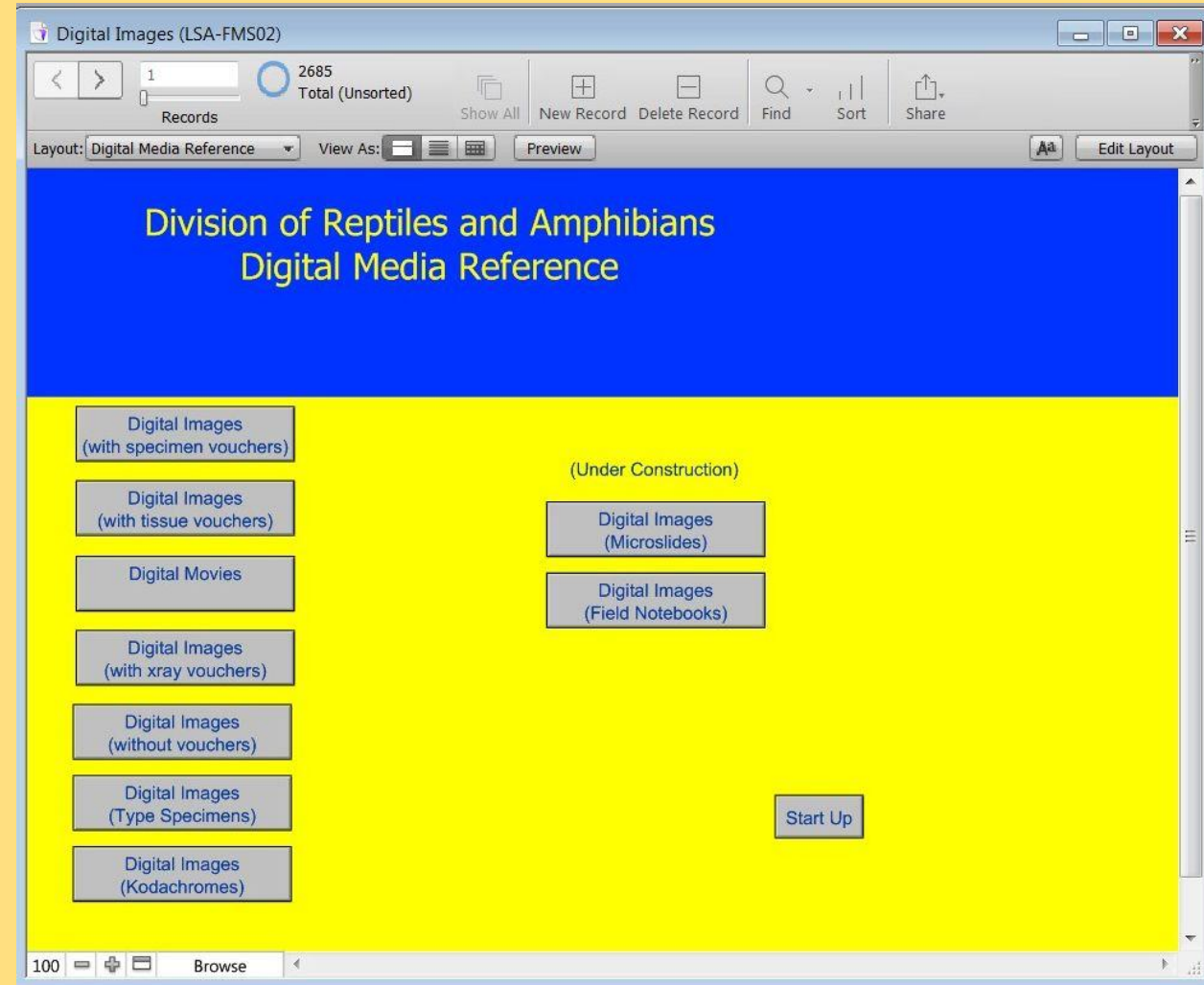
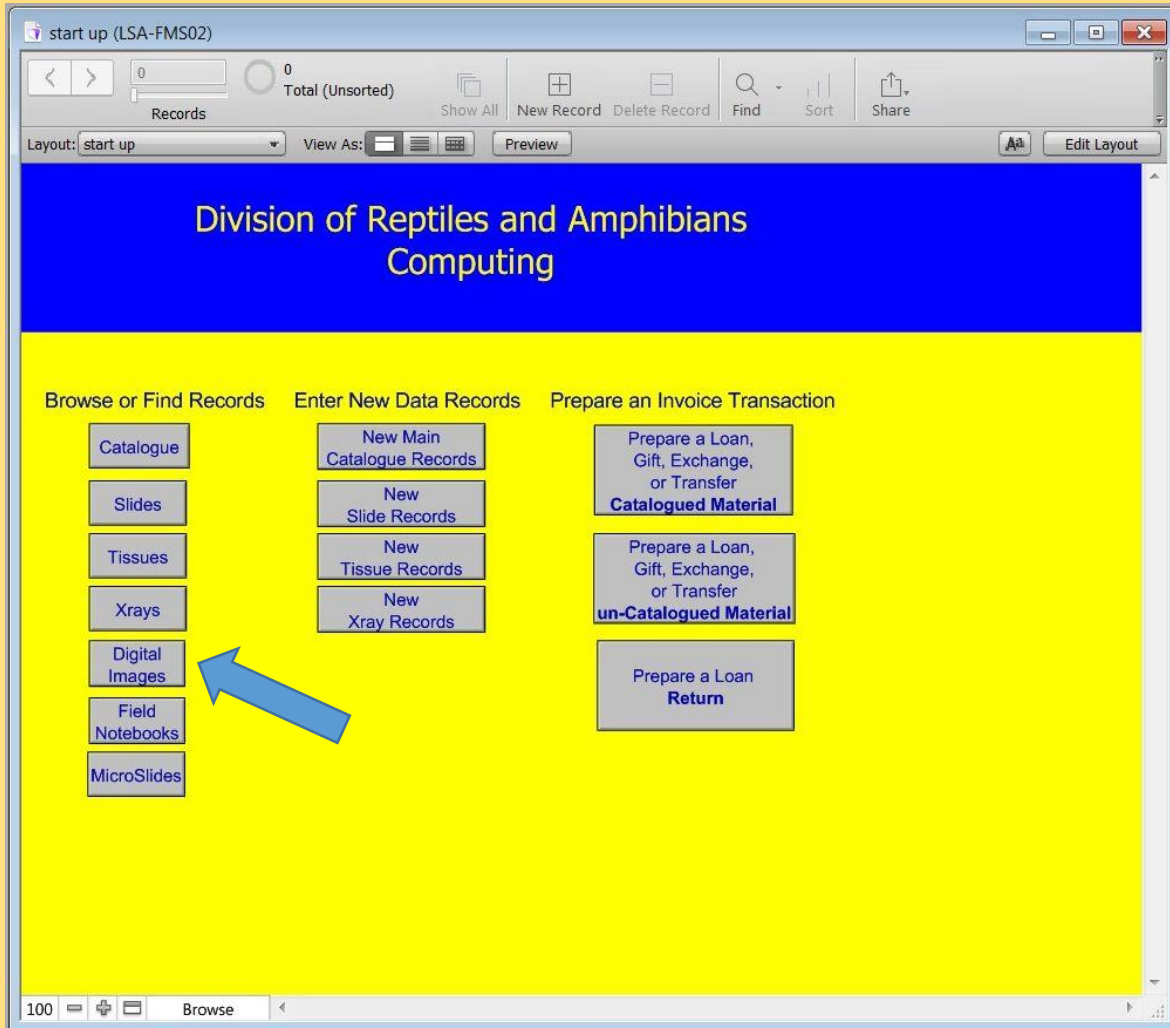


I am among the biologists now proposing that it would be an enormous benefit to science and humanity to get on with the exploration of biodiversity. The prospect of the initiative has now been accelerated greatly by new technology, in particular high resolution digital photography, Internet publication and rapid DNA sequencing.

With these new methods, we expect, and we held a conference 2 years ago at Harvard to examine all these methods and to estimate what needs to be done to complete the exploration of life on earth. And we agreed, leaders on biodiversity studies on a continental and global basis, that we could do this within 25 years.

Upon this effort, then can be built an **Encyclopedia of Life**, where on each electronic page for which species known, from virus to whale, is available anywhere by single access, on command. The page would contain its scientific, and if available common name of the species, a pictorial, a genomic presentation, the primary type specimen on which the name was based so that we will have stability in nomenclature and reference systems, and a summary of the species diagnostic traits. **The page will open out directly or by links to other databases to a summary of everything known about the specie's biology and its perceived practical importance to the environment and humanity.**

FilemakerPro



Institutional Decision – 1 database for all

- 2013 – 2014 College wide decision for all LS&A Museums to use a common database solution KeEMU
- Project Goals
 - **Cross Collections searching**
 - **Archival Security**
 - **Appropriate interaction / integration with UMDL (Digital Library)**
 - **Promotion of data to other institutions/peer organizations**
 - **Promotion of College Assets/Resources and [Institutional Prowess \(Go Blue\)](#)**
 - **Expectations at the organizational level for the future development and support needs of these collections**

KeEMU – Pros and Cons

- Subject to and reliant on LSA & Emu changes in staff expertise, direction and leadership
- The software is not user friendly
- The learning curve is steep
- Not necessarily a collection management tool
- Data entry is cumbersome
- What about quality control of data input?
- No team of experts or “Uber” users – reliant on a few (ie. collection managers)
- How will new students, staff, faculty learn Emu?
- We pay for a fixed number of licenses (users) per year without knowing how many we need
- Does it run on a Mac? Or an Ipad in the Collection space and / or on WiFi
- Is it a tool for doing work? Or rather a product that will require massive amounts of person/hours of data entry (including data gathering) to become utilitarian.
- What about reports? It is designed to work with “Crystal Reports” but we are not buying into that.

Institutional Decision – 1 database for all

- 2013 – 2014 College wide decision for all Biology Museums to use a common database solution
- November, 2014: Demos of Arctos and Specify
- By February 2015, The Decision was made to migrate the Biology Museums to Specify
- Should be completed by the end of the year
 - Birds and Mollusks were already using it
 - Mammals and Fish have migrated
 - Insects are in the process

Evaluating and Selecting a Database

- History: Computing in Herpetology at UMMZ
- **Databases Mean Different Things to Different People**
- Considerations / Recommendations

Databases Mean Different Things to Different People

- Collections Staff – Tools for doing work
- Data Consolidators – Data Sharing
- Deans / Directors / Administrators – “Promotion” or [Institutional Prowess](#)

Collections Staff – Tool for doing Collection Management work

- Data entry
- Queries
- Invoice Transactions
- Reports
- Administrative Records
- Data Import / Export
- Accommodating Multimedia
- Accommodating new projects / enhancements / solutions

New caecilian xray imaging project

FileMaker Pro

File Edit View Insert Format Records Scripts Window Help

ummz_herp_xrays (LSA-FMS02)

1950 2369 / 19252 Found (Sorted)

Records Show All New Record Delete Record Find Sort Share

Layout: maps & plates View As: Preview

XRAY Number 2684
Institution UMMZ (show UMMZ catalog record)
catalog number 182927
Field Number RAN 25936

map thumbnail (click to enlarge) plate thumbnail (click to enlarge)

map X-2684-M plate no. X-2684-P

map image path \\lsa.m.storage.umich.edu\lsa- file extension .jpg
plate image path \\lsa.m.storage.umich.edu\lsa- file extension .jpg

ORDER GYMNOPHIC
FAMILY CAECILIIDAE
GENUS HYPOGEOPI
SPECIES ROSTRATUS
COUNTRY SEYCHELLE
STATE MAHE
COUNTY
LOCALITY MARE AUX C
MONTH JUNE
DAY 8
YEAR 1986
Vertebral count
Type Status

Top Next Find Mode
Bottom Previous All Records Start up

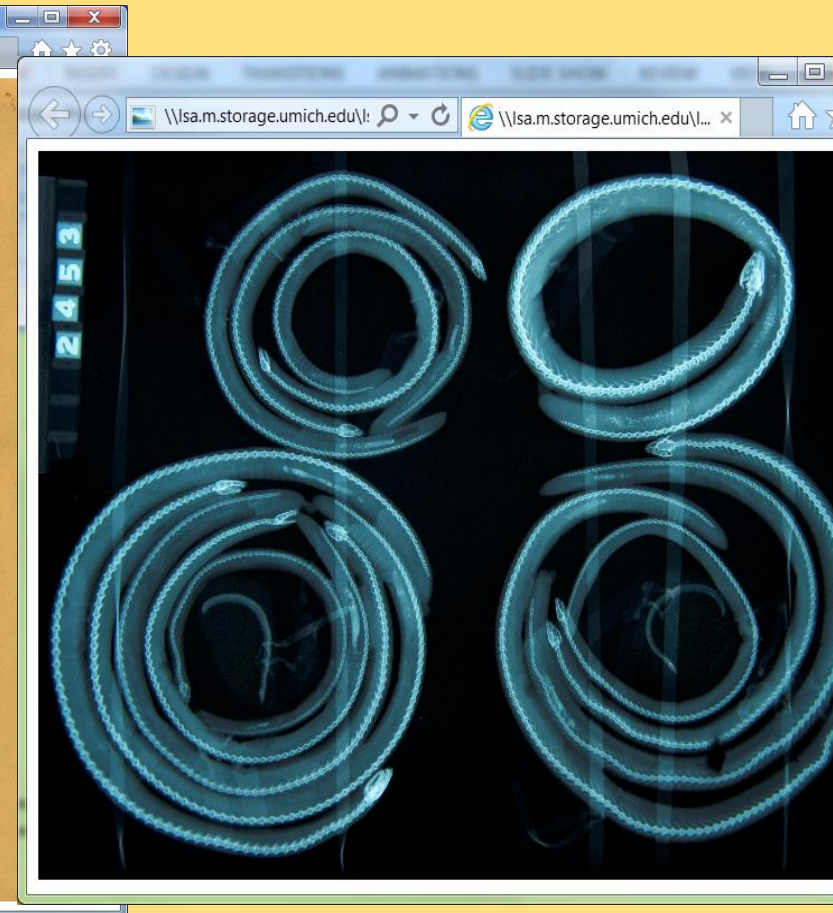
X-2684

No. 2684
Name _____
Date _____
Dr. _____
Remarks All UMMZ
Hypogeophis - Male 7
+ female - 182927 + 183066

182926
182927
183021
182923

182926
182927
182927
182927
182927
182927
183024

CONE INSTRUMENTS
REORDER #410006



Possible dietary studies



Data Consolidators – Data Sharing

- Different Goals from Collection Management (although some in common – specifically data sharing)
- Different clients / end users (although some in common)
 - Public – citizen science
 - Government agencies: local - international
- Different Issues requiring different solutions - **Programming**
 - Standardization – including Darwin Core & Integrated Publishing Toolkit
 - Security – copyright, privileges
 - Redundancy – aggregators
 - Updating data and repatriation of data to source

Data Consolidators – Programming

- Taxonomy – synonyms
- Geography – standard names and variants

- These solutions are particularly challenging because there is not agreement and they are not static.

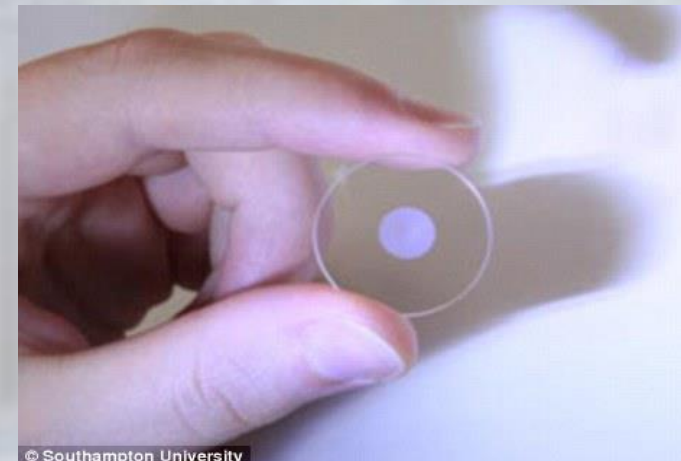
Deans / Directors / Administrators – “Promotion” or Institutional Prowess

- Goals are different (although data sharing is still a common goal)
 - Cross Collection Searching
 - Impact Factor – metrics on usage and citations
- Advertising and Promotion
 - Participation toward a larger effort
 - Networking and funding opportunities
 - Promotion to prospective students and faculty
- These are different considerations for evaluating a database.

Conclusions & Recommendations / Considerations

- When evaluating / selecting a database, all should consider
 - Versatility & flexibility
 - Accommodation of others while accomplishing immediate goals
 - Import / Export capabilities and compatibility
 - Longevity and obsolescence (COTS)
 - Networkability including operating systems and compatibility
 - Reliance on others and operating systems
 - Storage and backup systems and compatibility – **Data Management Plan**
 - Plan for huge growth - Speed and Space may become issues
 - Accommodation of new technology and Innovation

Superman 'memory crystals' to become a reality as scientists store computer data on powerful glass hard drive



**Institutional Ecology, 'Translations'
and Boundary Objects: Amateurs and
Professionals in Berkeley's Museum of
Vertebrate Zoology, 1907-39**
Susan Leigh Star and James R. Griesemer

Social Studies of Science (SAGE, London, Newbury Park and New Delhi),
Vol. 19 (1989), 387-420

The central analytical question raised by this study is: **how do heterogeneity and cooperation coexist**, and with what consequences for managing information? The museum is in a sense a model of information processing. In the strategies used by its participants are several sophisticated answers to problems of complexity, preservation and coordination. Our future work will examine these answers in different domains, including the history of evolutionary theory and the design of complex computer systems.