Use cases for online taxonomic literature from taxonomists, conservationists, and others

Cynthia Sims Parr
Christopher Lyal

INFORMATION INTERNATIONAL ASSOCIATES (IIA)

www.iiaweb.com

TDWG 2007
Bratislava, Slovakia

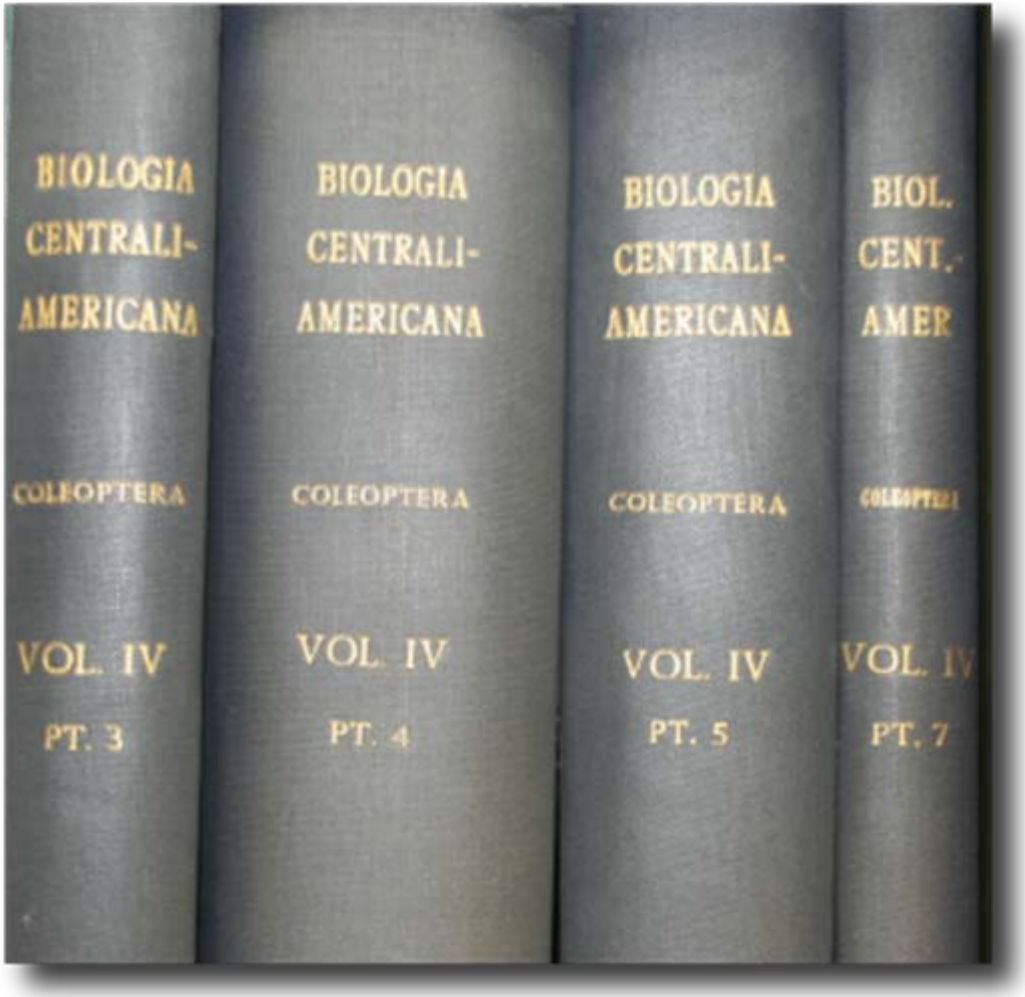
NATURAL HISTORY MUSEUM

Motivation

- Build systems that meet the needs of users
  - Clarify scope
  - Drawn from community work practices
  - Enable intuitive interaction -- easily learned, easily used
  - Support typical workflows
- Assist standards development
  - Data models should reflect community needs
- Assist developers
  - Explicit system goals
- Assist testing
  - Explicit sequences of tasks
- Does anybody really want your system?

INtegrated **O**pen **TAXA**nomic System
INOTAXA

- Both older and newer literature
- TaXMLit
- Integration with other resources – taxonomic databases, gazetteers, etc.



What are use cases?

- Stories
- Scenarios
- Capture behavior of users and the system in pursuit of user goals
- Can be oriented for a particular system, or for general work activities (aiming here at that)
- Usually written in natural language, but structured

Alistair Cockburn. 2000. Writing effective use cases. Addison-Wesley.

Use case descriptions

- Primary actor
- Scope
- External or internal goal
- Priority
- Preconditions and trigger
- Minimal success guarantees
- Success scenario
- Sequences of tasks – actions of both user and system
- Extension conditions
- Alternate technology

Methods

- Interview different categories of users
- Compile use case titles at different levels
- Pull from requirements; personal experience
- Avoid implementation decisions, focus on tasks
- Group summary level use cases
- Flesh out each description
- Organize into sub-use cases
- Review against and modify requirements, review by users again

ITERATE

Interviews

- So far, 13 formal interviews ~ 1 hour each (*in situ* if possible)
- Systematists, USDA Agricultural Research Service, Science Policy, etc.
- Different levels of comfort with technology
- Different taxonomic foci
- International users
- Our own knowledge
- Interactions with colleagues

How do they do things now,
how would they do things if they could

Results
Categories of use cases

- 50+ use cases, some redundancy
- Summary use cases grouped into:
  - Exploration
  - Classical taxonomy
  - Identification
  - Extra-taxonomy research
  - Policy-making
  - Data maintenance
  - Web services

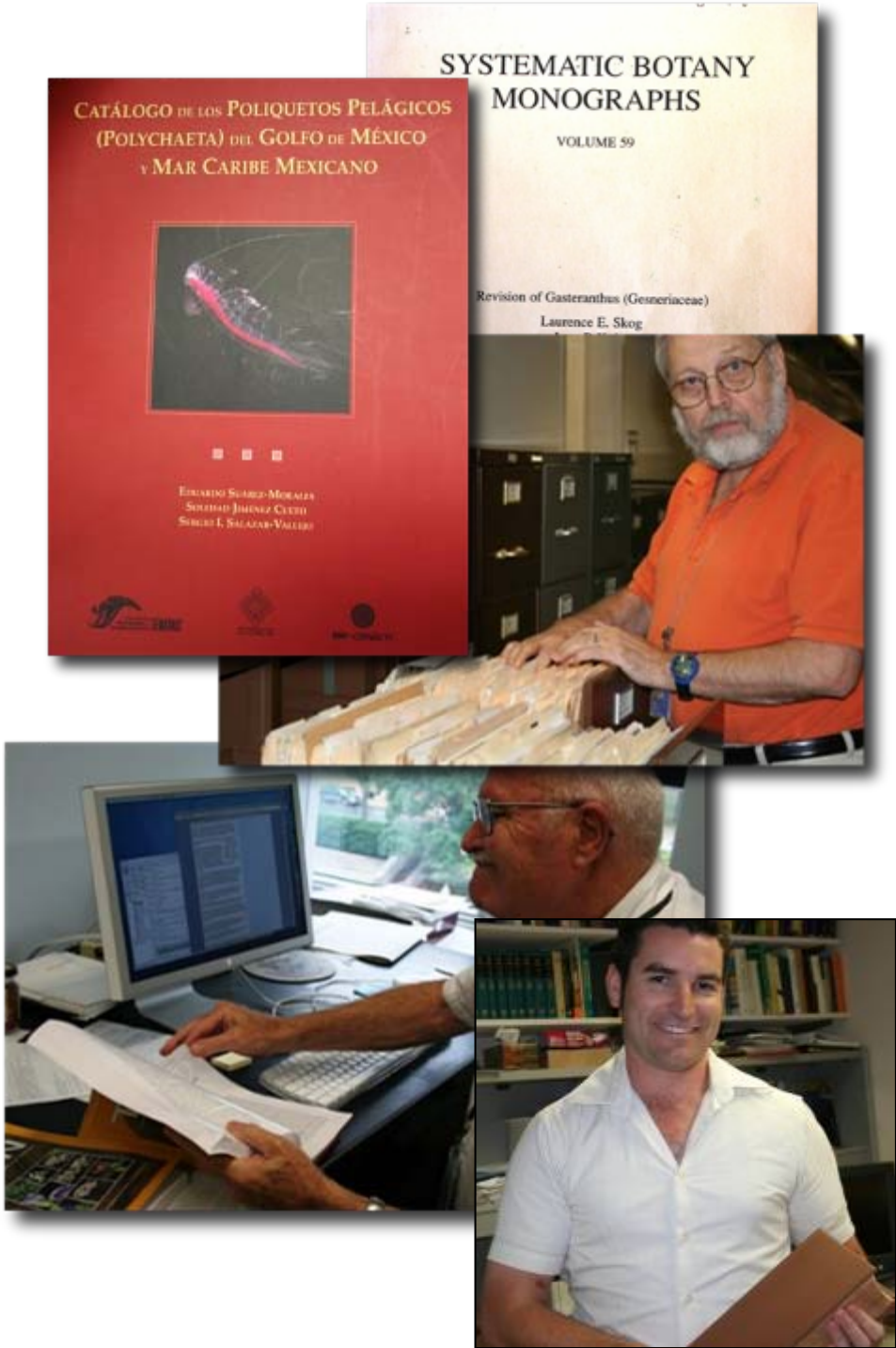
Exploration

- Student, searching for a group to work on
- Systematist who has never used the system before
- Systematist deciding when and where to collect

- Goals internal to system
- Ease of use critical
- Limit dead-ends

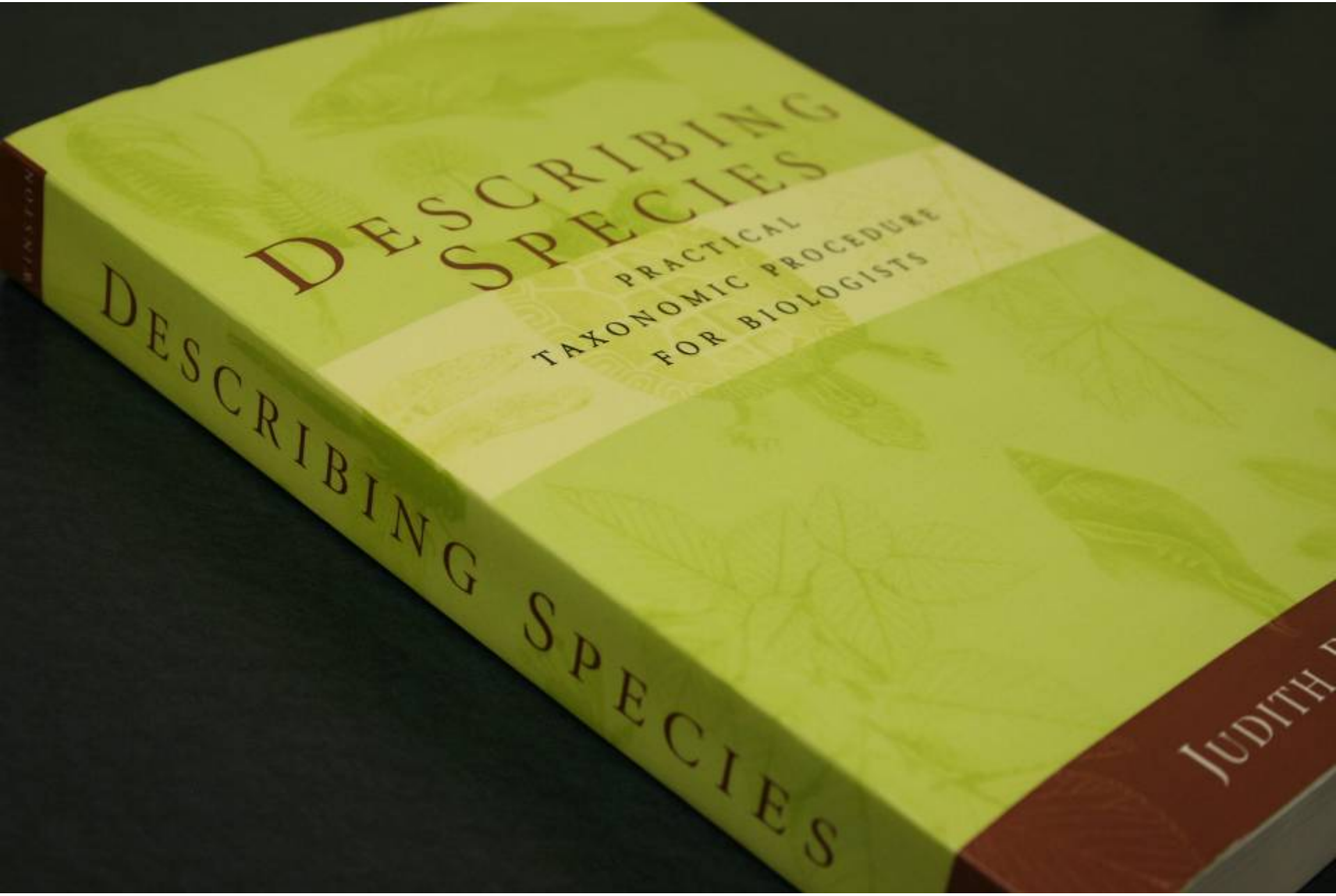
Classical taxonomy

- Systematists:
  - Describe species
  - Prepare revision
  - Prepare checklists, keys, catalogs
- Code committee:
  - Arbitrating nomenclatural acts



Goal often a product that would be outside the scope of the system BUT

•Find and save critical PDFs

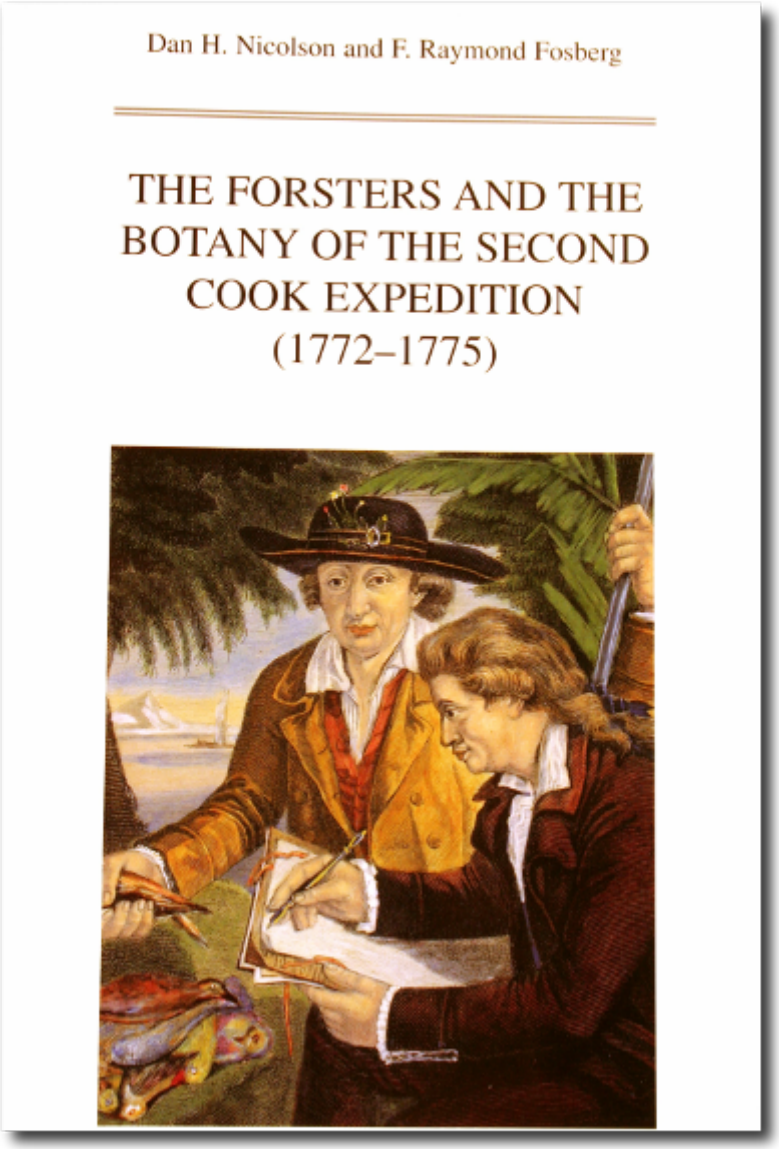

DESCRIBING SPECIES
PRACTICAL TAXONOMIC PROCEDURE FOR BIOLOGISTS
DESCRIBING SPECIES
JUDITH

Use Case Name	**Describe (new) taxon**
Use Case Description	Systematist verifies that species or group is distinct and (if it is new) has not already been named. Compiles information necessary for description.
Scope	INOTAXA and others
Level	Summary goal
Primary actor	Systematist
Priority	High
Precondition	Systematist tries to identify a specimen but it appears to have not yet been described OR Systematist has conducted a phylogenetic analysis suggesting that a group is different than previously described. Systematist knows some larger taxonomic grouping for the undescribed taxon
Main success scenario	1.Systematist finds sub-taxa names in larger taxon 2.Systematist finds specimen data for these taxa 3.Systematist gathers publications (treatments, etc) for these taxa 4.Systematist finds images for these taxa 5.Systematist borrows specimens for these taxa 6.Systematist determines distribution from publications and specimens 7.Systematist 8.(Optional) Sy species) 9.Systematist g 10.Systematist 11.Systematist 12.Systematist ORDER OF T
Extension conditions	Systematist may want to print out the treatments Correct specimen identifications as appropriate

Systematist finds synonyms of names in order to get original names of the types
Finds authors in this taxonomic group
Finds institution of authors where their material might be
May need to conduct external search for contact info
Requests material

Other use cases

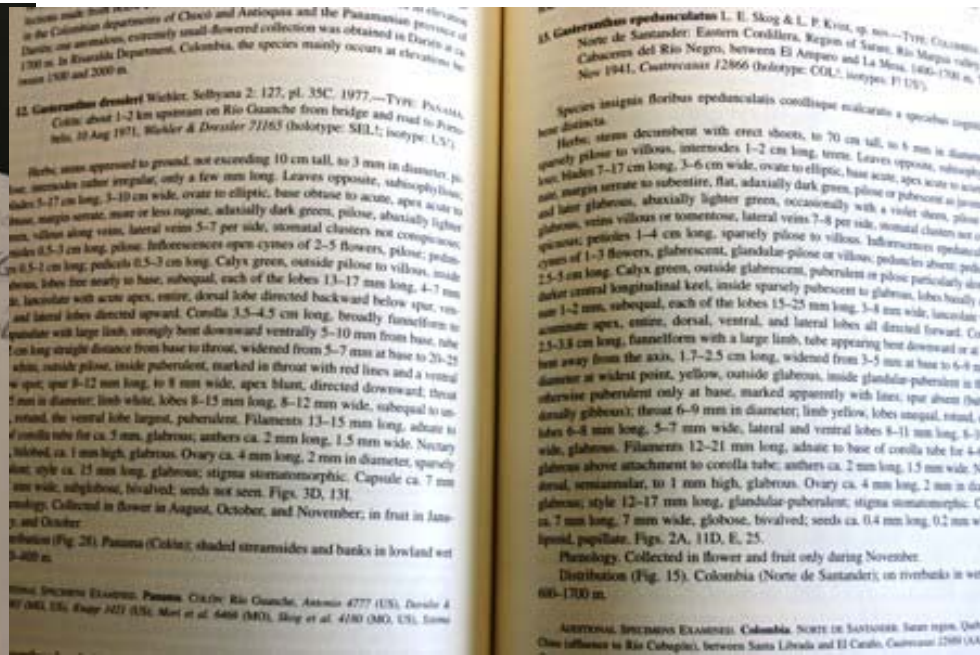
- Find specimens with associations
  - Same expedition or same collector
  - Parts of same plant specimen occurring in multiple collections
  - Host/parasite ecological associations



Other use cases

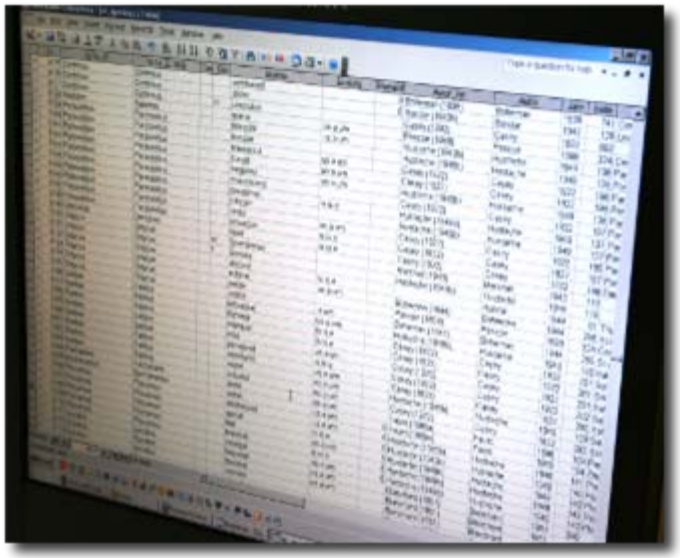
Compile character data for monograph descriptions or phylogenetic analysis

Eventually: Publish data associated with descriptions



Classical taxonomy

- Will probably take advantage of most available features
- Will have other technologies at their disposal
- Access to specimens and building personal networks
- Digital resources particularly important for systematists not at major institutions



Identification

- Find likely options in geographic region or within taxonomic group
- Search for images
- Currently unlikely to search for characteristics
- Some might search for characteristics if available
- DELTA/INTKEY works well, except when it doesn't
- Take advantage of both literature-based and online keys

Extra-taxonomy research

Ecologists, conservation biologists, etc.

Want detailed natural & life history & character information for each taxon

- change over time in measurements or distribution
- tease out geographic patterns of host associations
- species richness and niche modeling (integrating info from descriptions and specimens)
- identify organisms at risk (checklists over time)
- test hypotheses about character evolution
- natural products drug discovery
- characterization of invasive species

Extra-taxonomy research

- Less likely to require explicit details important to taxonomists BUT
- Probably need implicit use of synonymy etc.
- Need guidance on completeness and fitness-of-use
- May be more likely to use browsing rather than advanced search
- More likely to need links to external information
- Desire atomized data and spreadsheet/database downloads
- Less likely to want to read entire documents from original literature

Policy-making

- People advising policy makers
- e.g. evidence to support for decisions such as what area to protect (based on diversity, etc.)
- Especially need to navigate valid and invalid names: legal implications
- Are lumping and splitting of taxa based on objective evidence rather than politics?

- Focused needs for atomized data - need easily extracted information
- Particularly problematic are surveys where taxa not mentioned in abstract

Data maintenance

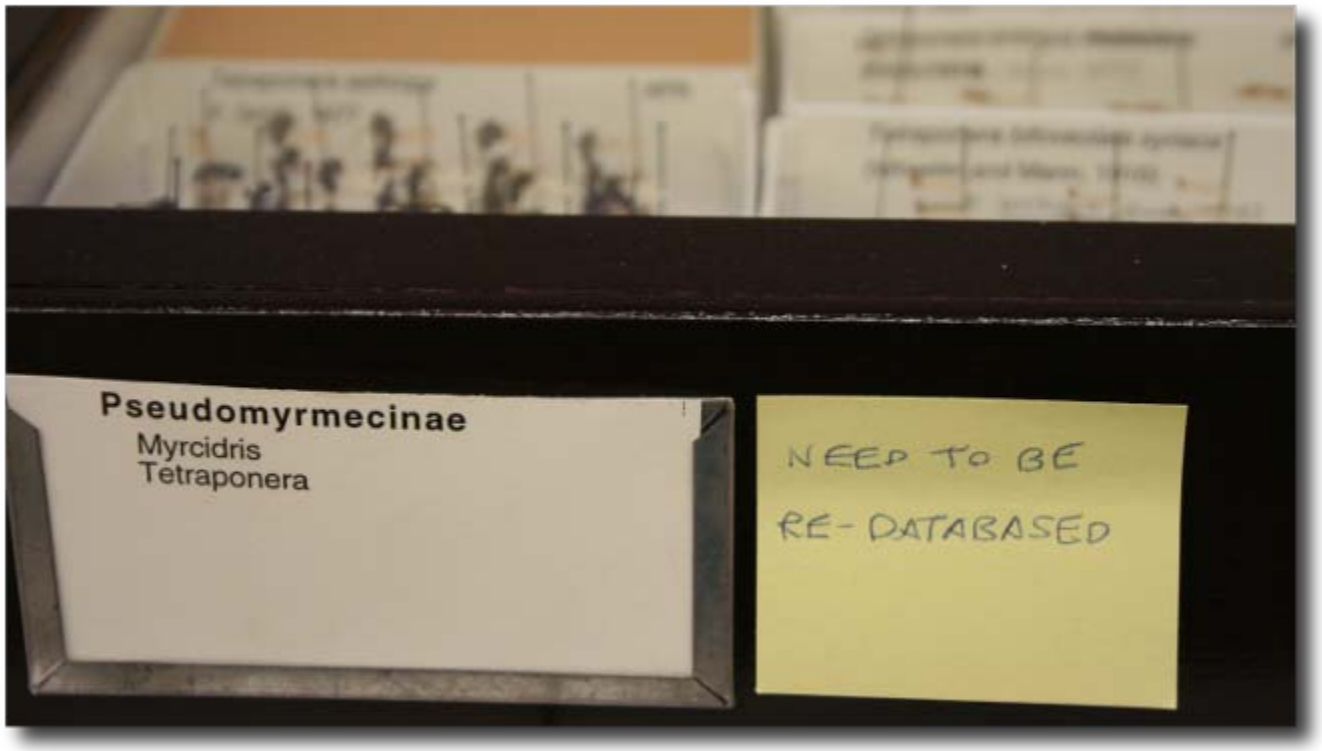
Internal to system

- Proofread XML
- Add new material
- Annotations

External to system

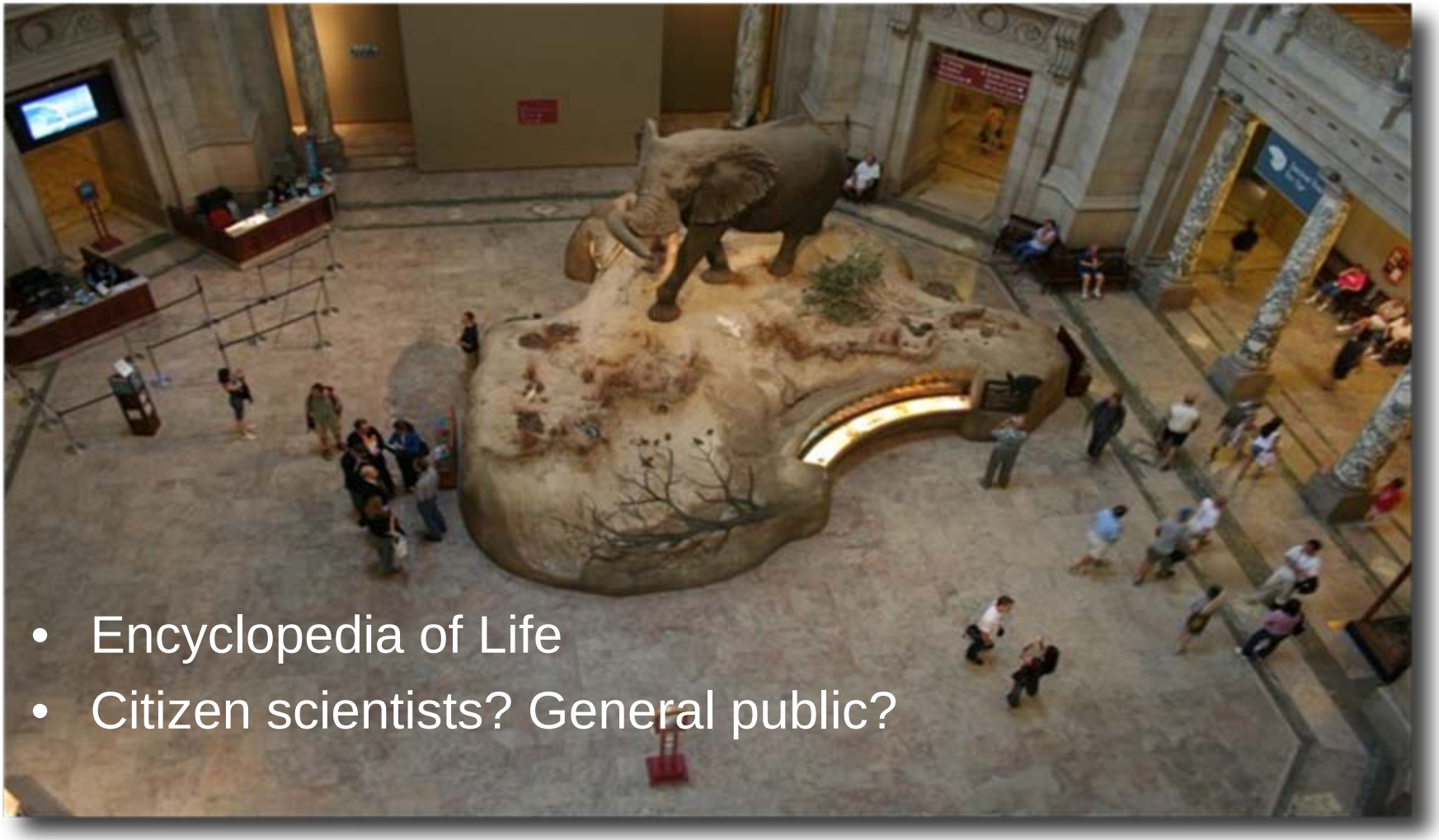
- Proofread data in catalogs
- Verify type status in collections

First category where critical to have login system



Web services

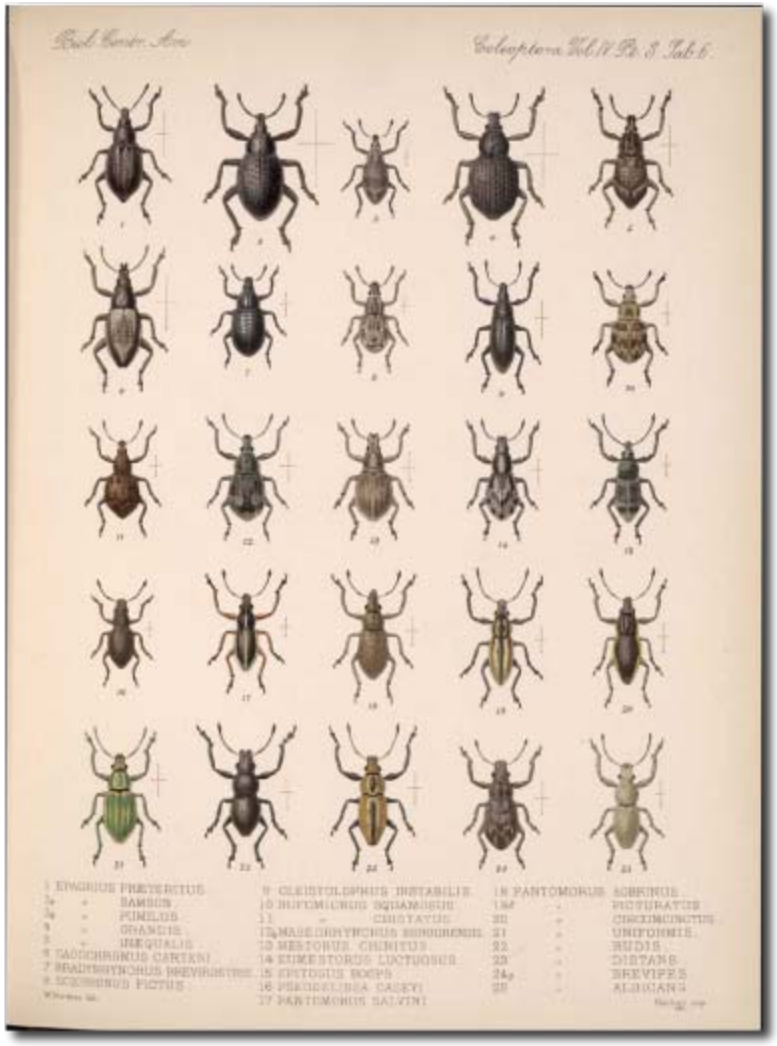
- Use case perspective: users are other systems



- Encyclopedia of Life
- Citizen scientists? General public?

Conclusions

- Diverse user base and use cases for literature
- Multiple points of entry
- Atomized data necessary for important use cases
- Need for both PDF and database report formats
- Information from literature a vital supplement to existing databases
- Integration of literature content with other sources of biodiversity information valuable
- Opens door to flexible publication of literature



Thanks to: Interviewees, Anna Weitzman, and Farial Shahnaz
Funding from Atherton Seidell Fund of the Smithsonian Institution