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

Cynthia Sims Parr Christopher Lyal



TDWG 2007 Bratislava, Slovakia

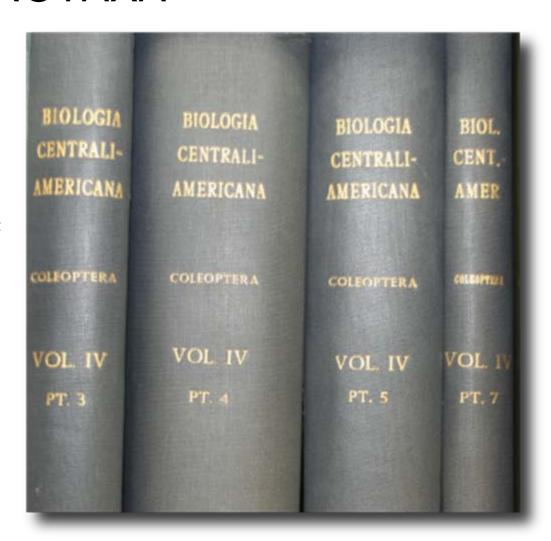


### **Motivation**

- Build systems that meet the needs of users
  - Clarify scope
  - Drawn from <u>community</u> 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 Open TAXAnomic 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

### 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**

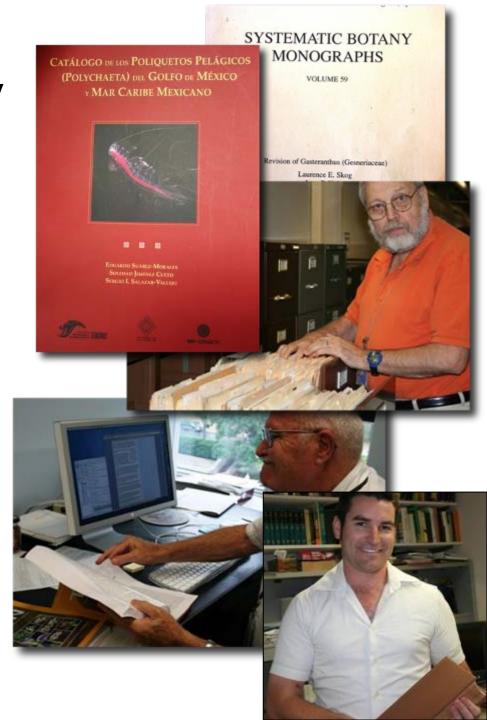
- Student, searching for a group to work on
- Systematist who has never used the system before
- Systematist deciding when and where to collect
- Goals internal to system
- Ease of use critical
- Limit dead-ends

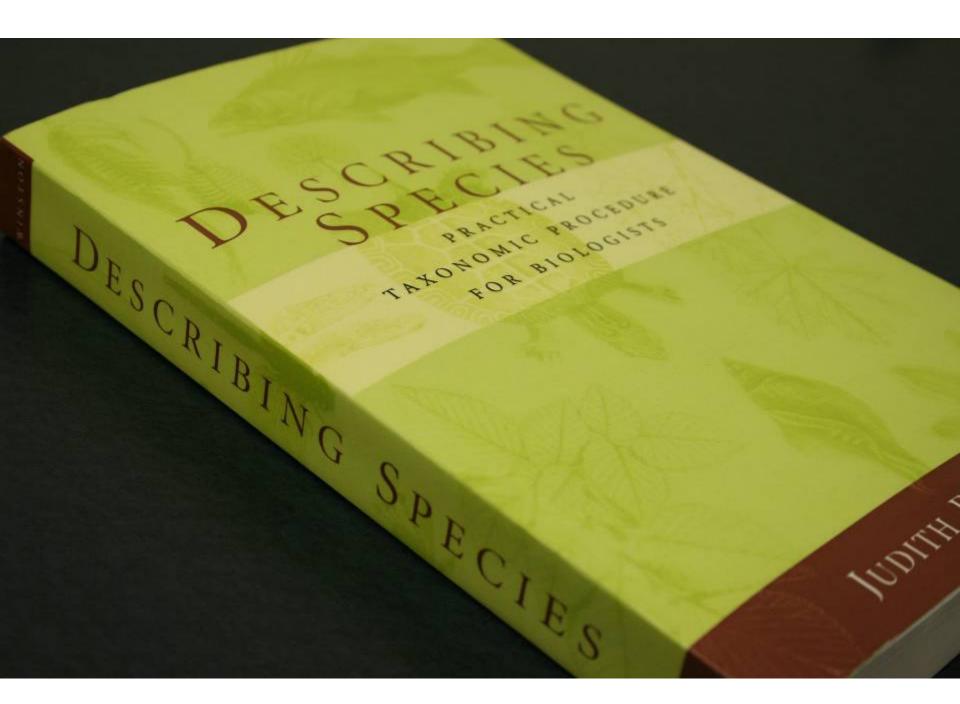
### Classical taxonomy

- 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





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 <u>identify a specimen</u> but it appears to have not yet been described OR Systematist has <u>conducted a phylogenetic analysis</u> 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 finds sub-taxa names in larger taxon 2.Systematist finds specimen data for these taxa 4.Systematist borrows specimens for these taxa 5.Systematist determines distribution from publications and specimens
	Systematist <u>finds synonyms of names</u> in order to get original names of the types  9.Systematist <u>finds authors</u> in this taxonomic group  10.Systematist <u>finds authors</u> in this taxonomic group  Finds institution of authors where their material might be  May need to <u>conduct external search</u> for contact info  Requests material
Extension conditions	Systematist may want to print out the treatments Correct specimen identifications as appropriate

### 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

Dan H. Nicolson and F. Raymond Fosberg

THE FORSTERS AND THE BOTANY OF THE SECOND COOK EXPEDITION (1772–1775)





### Other use cases

Compile character data for monograph descriptions or phylogenetic analysis

Eventually: Publish data associated with descriptions

Come dwarf 1-3 km spacesom on Rio Guanche from beidge and front in the Come and 1971. Menter & Docasier 71163 (holotype: SELL; honype: Lyseconds, only a few men long. Leaves opposite, subscoping has and a first long 3-10 cm wide, ovate to elliptic, bear obtains to acute, aper acute a death of the first acute a special control of the second section of the section of t alex 1-17 cm long, the state of less regime, adutably dark green, pilose, abattaby lighter acquire series. in 13.3 cm keep priore inforcementers open cythes of 2-5 flowers, priore; priore Con 8.5.2 cm long, poderit 8.5.3 cm long. Calya green, outside pilose to village, next, like for nearly to have, subequal, such of the lobes 1.3-17 mes long, 4-7 of latest inter-directed apward. Corollo 3.5-4.5 cm long, broadly furnishing to as long straight distance from base to throat, widered from 5-7 mm at base to 20-24 per year 8-12 men long, to 8 men wide, upon blant, directed downward, threat neur link white, lobes 8-15 mm long, 8-12 mm wide, subequal to unand the second little largest, puberulent, Filaments 13-15 mm long, advances South for ca 5 new, glabrous, authors ca. 2 mm long, 1.5 mm wide. Naviersel ca. I see high, glabraus. Overy ca. 4 mm long, 2 mm in diameter, spanish myle ca. 25 mm long, glabrous; stigma stornaromorphic. Capsule ca. 7 mm wife substitution, bivalient seeds not seen. Figs. 3D, 131. er, Collected in Sower in August, October, and November, (3) Passers (Coking shaded streamsides and banks in lowland set the Discrete, Parsons, Charles Rio Guardie, Astronia 4777 (US), Discribe 4. MILE THE STATE AND CASE MANY OF ALL SAME CASES, Strong or all. 47/00 (MIC), USA, Science

Cabacres del Rio Negro, between El Ampreo and La State, Rio Maryo, roles (2014), Sore de Santander: Eastern Cordillera, Region of Sature, Rio Maryo, roles Cabacres del Rio Negro, between El Ampreo and La State, Rio Maryo, roles New 1941, Caustreonnes (2866) dichotype: COL, increpen. 1

the state of the s

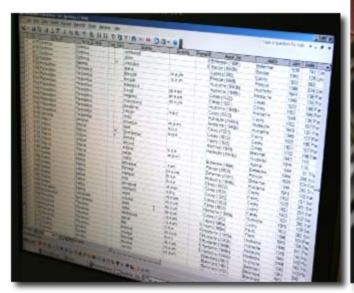
com of 1-3 flowers, glabbroconts, gaardadae-pilose or vilinas; probastics between put 25-5 mt long. Callys green, outside glabbrocons, pobersises or phose perceivate and ment armed longistudinal keel, inside sparsely pubeaces to glabrous, lobb below he was and ment armed longistudinal keel, inside sparsely pubeaces to glabrous, lobb below he was a second sparsely, entire, devisal, ventral, and lateral lobes all directed forward. On the sparsely pubeaces to all directed forward. On the sparsely flow he area, 1.7–2.5 cm long, widered from 3-5 ms; at har to 6-6 ms a walest point, yellow, outside glabrous, inside gladeling-petersism in starvise pobersism of outside pobersism of the sparsely pubeaces, the sparsely longistic pobersism of the size of the sparsely pubeaces. However, the sparsely pubeaces, the sparsely longistic pobersism of the sparsely glabrous. Filaments 12-21 ms long, admit to lose of crouls tole for a plateous above attachment to corolla table; suchers at 2 ms long, 2 ms in digitions above attachment to corolla table; suchers at 2 ms long, 2 ms in digitions, style 12-17 mm long, glabrous. Overy cs. 4 ms long, 2 ms in digitions, style 12-17 mm long, glabrous. Over yes, 4 ms long, 2 ms in digitions, style 12-17 mm long, globous.

Pand, popular. Figs. 2A, 11D, E, 25.
Panology. Collected in flower and fruit only during November.
Distribution (Fig. 15). Colombia (None de Santander), on riverbusia in we

ANTONIA INCOME EXAMENSI Calendia Nort in Santonia Saur ope, (lab

### 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 literaturebased 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

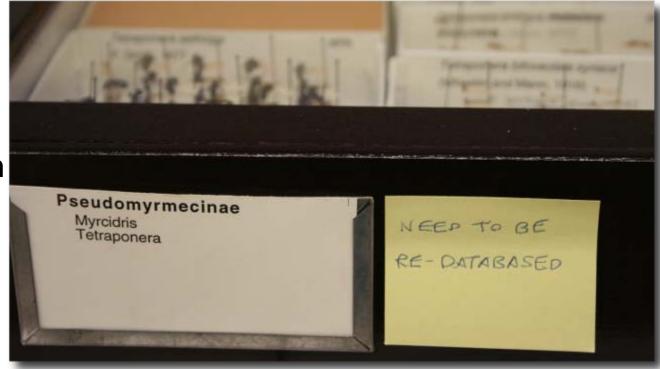
#### Internal to system

- Proofread XML
- Add new material
- Annotations

#### External to system

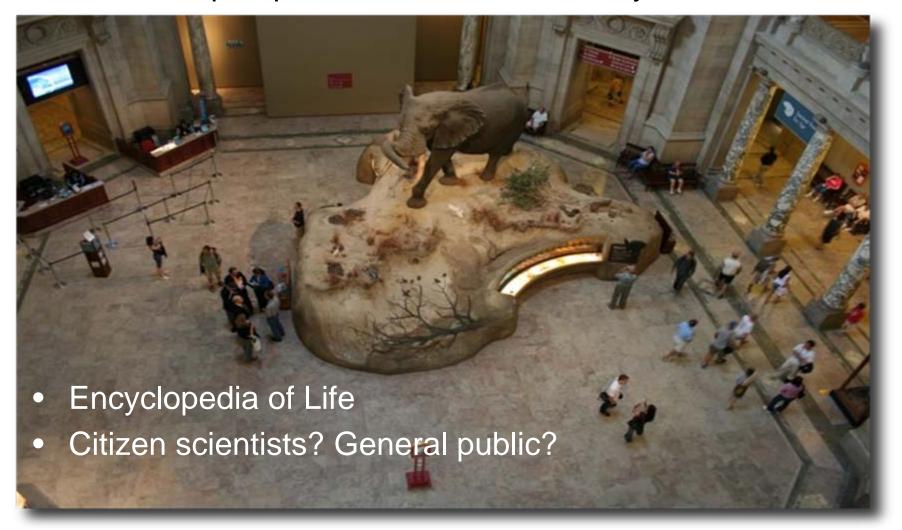
- Proofread data in catalogs
- Verify type status in collections

First category where critical to have login system



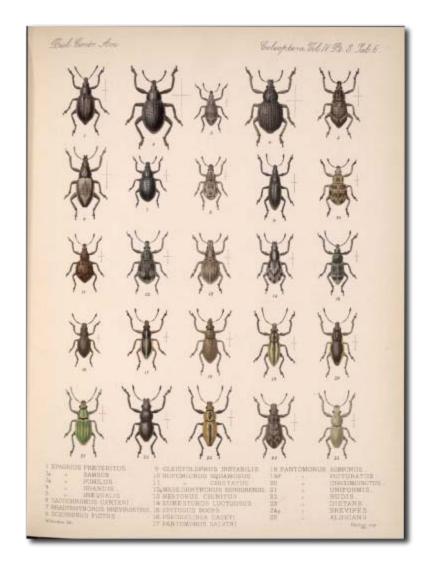
### Web services

Use case perspective: users are other systems



### 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