

# The Electronic Biologia Centrali-Americana

A vision for electronic access to  
taxonomic resources: the  
information interface between  
libraries and systematic biology



Anna L. Weitzman



Smithsonian  
*National Museum of Natural History*

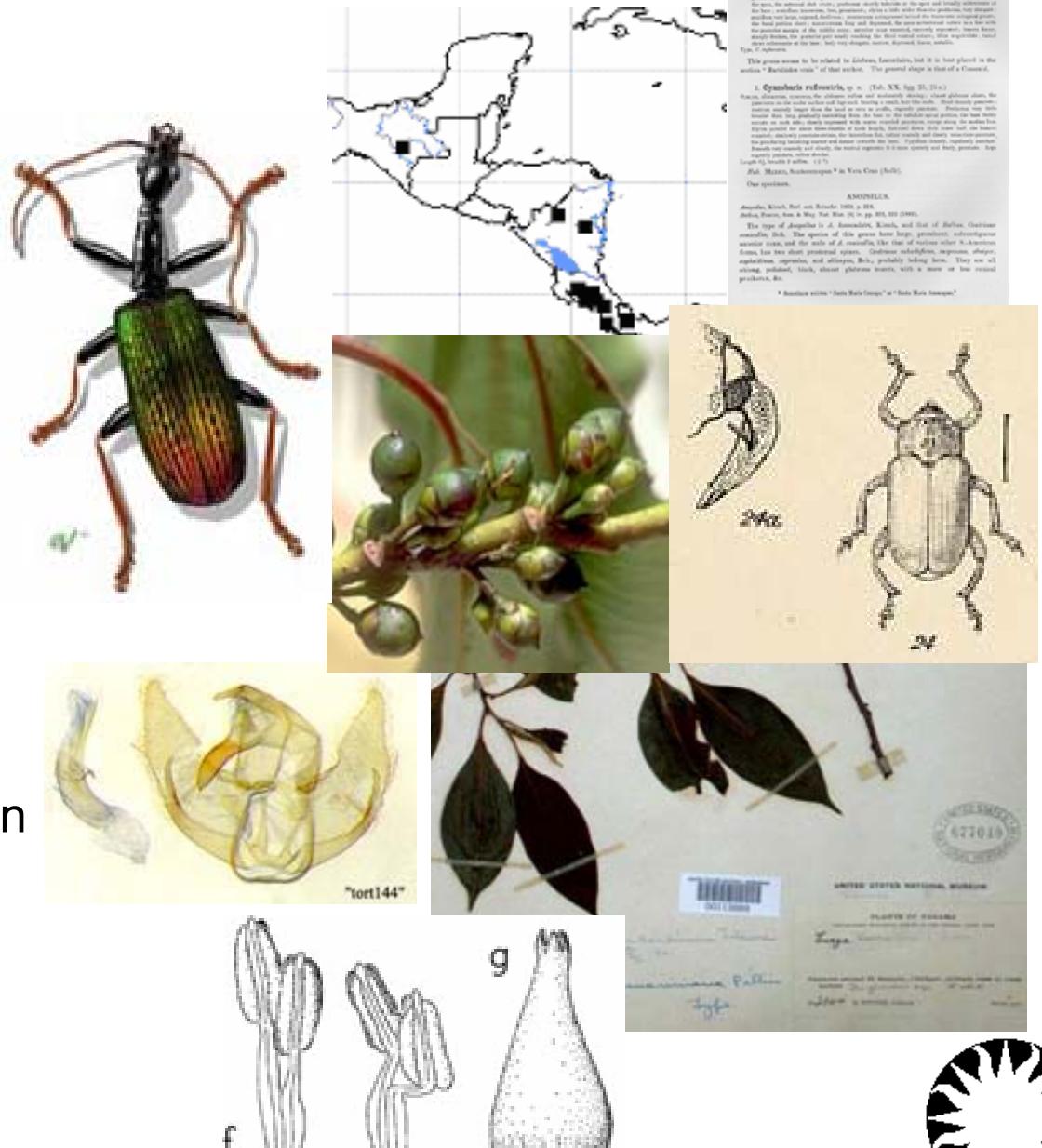
Christopher H. C. Lyal



# The problem of too much data

Data are of many types:

- specimens & associated data
- original descriptions
- images of dissections
- organs
- current treatments
- synonymies
- observations
- identification keys
- geographic information
- images of living specimens



# *The problem of too much data*

Data can be found in many unconnected places:

- Specimen collections
- Databases
- Publications
- Observations
- ‘grey’ literature
- Index cards
- Field notebooks



# ***The problem of too much data***

Many taxonomists and other researchers and ‘users’:

- cannot access all of these data sources
- do not know how to find them
- cannot afford the time or money to access them

Consequently:

Only a limited subset of potential data are used in most analyses, *limiting the adequacy of results*



# *The problem of inadequate data*

The data used in such analyses may be biased:

- subsamples based on ease of access, *not* rational decisions about what data are most important to particular analyses
- likely to be biased in collecting methods, collecting localities, and other institutional biases
- published observational data predating abstracting services likely to be missed
- unpublished data very likely to be missed
- data not catalogued or stored by an expected logic likely to be missed



# *Data are needed from all collections:*

File Edit View Favorites Tools Help Address <http://habanero.nhm.ukans.edu/TSA>ShowDataX.asp?TableID=6&RSID=3&format=Summary> Go

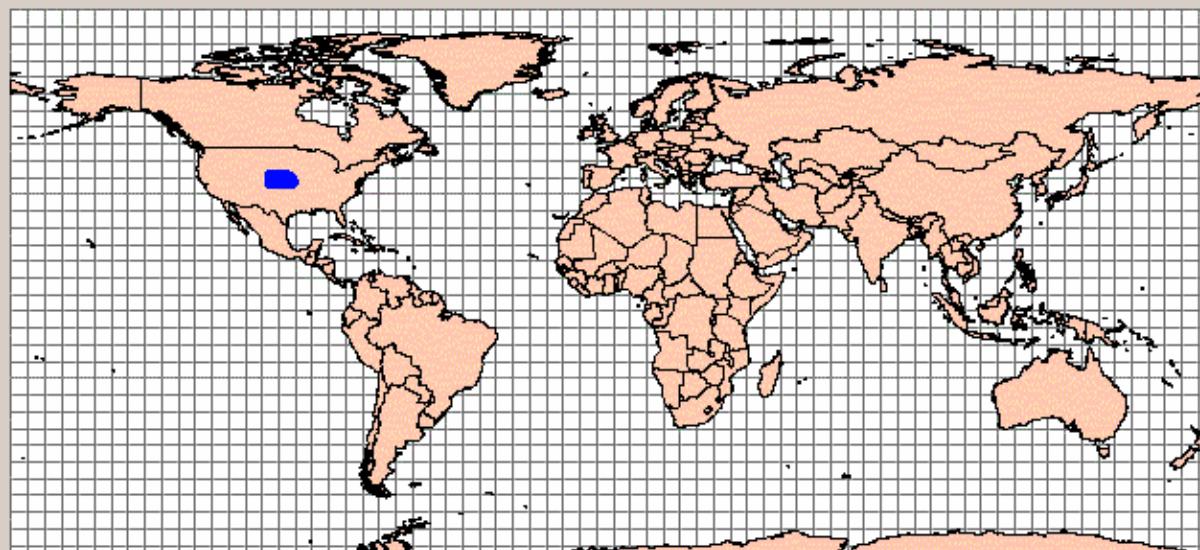
### Summary Information

Summarizes the results from this query by listing the unique scientific names, the number of records (actually the number of records with valid year entries) and the earliest and latest years of collection for each name. A direct link to the ITIS and GenBank (nucleotide or protein) databases is provided. Clicking on those links will open a new window.

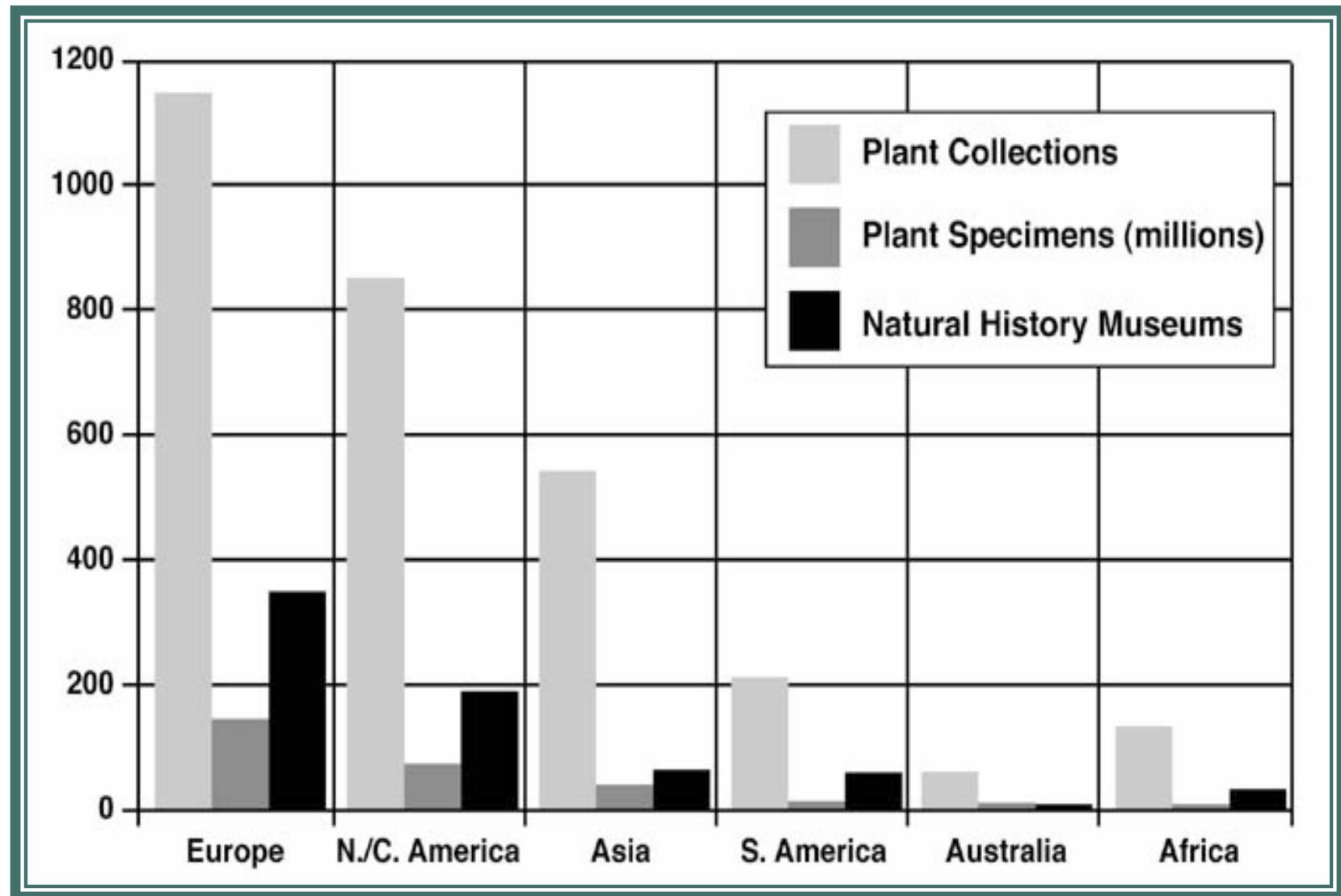
Scientific Name <a href="#">(ITIS Link)</a>	Num. Records	Earliest Year	Latest Year	<a href="#">GENBANK Link</a>	Zoo Record
<a href="#">Opuntia macrorhiza</a>	153	1868	1995	<a href="#">Nucleotide</a> <a href="#">Protein</a>	<a href="#">ZR</a>
<a href="#">Opuntia macrocentra</a>	1	1999	1999	<a href="#">Nucleotide</a> <a href="#">Protein</a>	<a href="#">ZR</a>
<a href="#">Opuntia megarhiza</a>	1	2000	2000	<a href="#">Nucleotide</a> <a href="#">Protein</a>	<a href="#">ZR</a>

### Distribution Map

This distribution map provides an indication of the global distribution of collection sites for the records identified by your query.



# *Poor distribution of systematics infrastructure*



# ***Common questions from megadiverse, taxonomy-poor countries :***

- What biota occur in my country/district/protected area?
- Where can I find descriptions and pictures of the species?
- Where can I find specimens of the species?
- Does this pest species occur in the country this fruit has been imported from?
- Is anyone working on this group?



# Use case: predicting the impact of invasive moth *Cactoblastis cactorum* on an important economic resource in Mexico

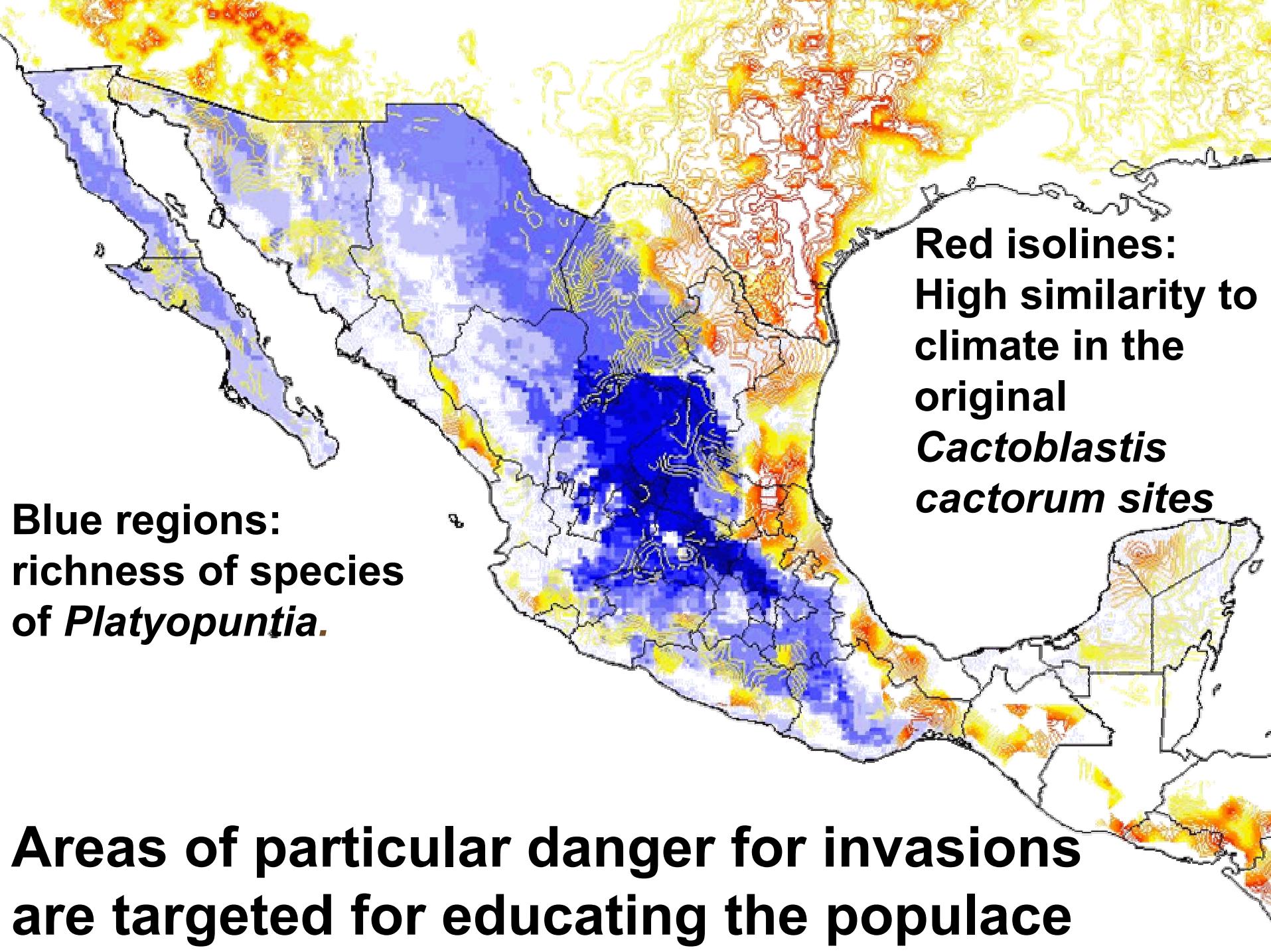
- Originally introduced from Argentina to Australia to control invasive species of *Opuntia*
- Invasive in US; of high risk to prickly pear cacti (*Opuntia* subgenus *Platyopuntia*) in Mexico
- Mexico has 56 native species, 38 endemic
- Many species are important for food (fruit and cladodes) or cattle and goat fodder forming a significant economic resource for the nation (good sources of protein, fibre, and water)



# Methods:

- About 40 collections and databases were queried for *Platyopuntia* localities
- Smithsonian collections were used to obtain localities for *Cactoblastis cactorum*
- Extrapolation algorithms (GARP & FloraMap) were used to obtain:
  - Approximations of the distribution of *Platyopuntia*
  - Regions of high climactic similarity to the original distribution of the moth





# *The Vision: uniting the data and making it accessible*

Ideally, key data should be accessible:

- From any location
- In the appropriate format(s)
- With a single query for each data type
- Using simple links
- Interoperably across data sets
  - ... *digitally*



# *The Vision: uniting the data*

Digitisation of names is underway, with several standards emerging (GBIF, ITIS, Species2000, UBio, Species Dictionary)

Digitisation of specimen data is underway, and standards developing (Darwin Core, ABCD)

Type images are being made available on the web

Numerous databases are on the web, in various forms

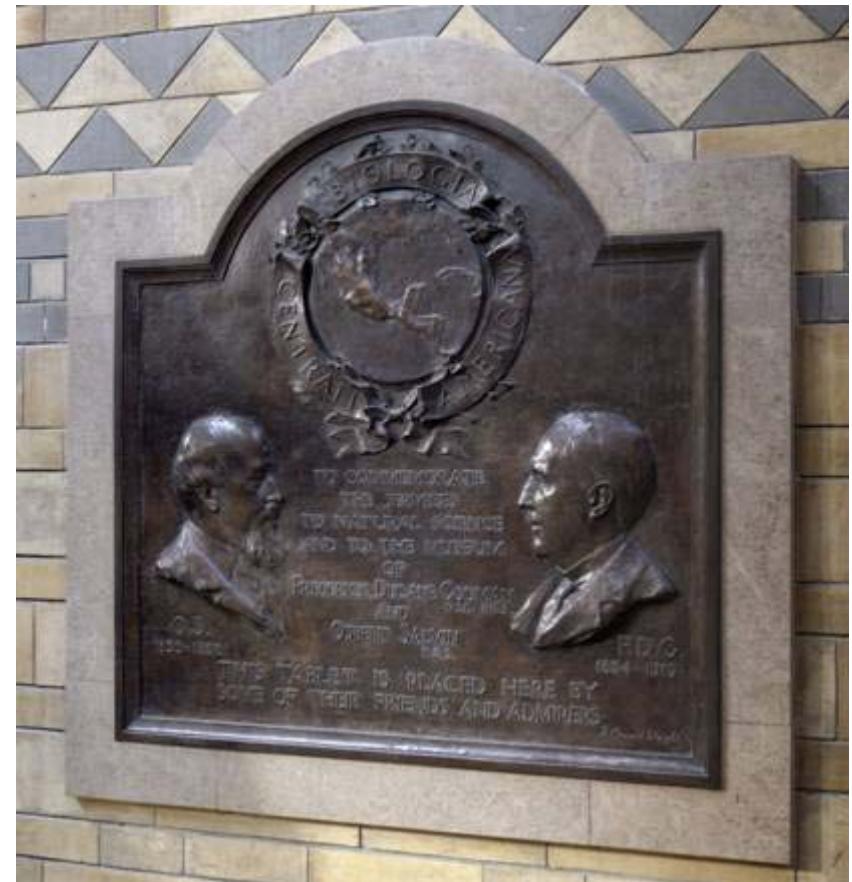
Access to data in the literature remains an issue: images of pages (e.g., jpeg, pdf) are slowly being made available, which provide greater accessibility, but cannot be searched or be made interoperable with other data



# *Creating the information interface between libraries and systematic biology*

## *The Biologia Centrali-Americana*

- a fundamental work for the study of New World biota
- includes most everything known at the time about the region's biological diversity
- privately issued (1879 – 1915) by F. DuCane Godman and Osbert Salvin of The Natural History Museum (London)
- 63 volumes with 1677 plates covering 50,263 species of plants, vertebrates, insects, spiders and related invertebrates, and mollusks



# *The Biologia Centrali-Americana*



- leading biologists of the time provided treatments
- for many groups still the current state of published knowledge
- few select volumes have been republished but never the entire series
- believed that the entire 63 volume BCA is held by only 8 libraries; many other libraries hold individual volumes or partial sets
- some Central American countries lack a complete set and the BCA is not generally accessible to taxonomists working in the region



# *The Biologia Centrali-Americana* Centennial Project



- Conceived at a Mellon-funded meeting to encourage collaboration between several large collections institutions
- Initial concept was to digitise the entire BCA and link it first to *Flora MesoAmericana* (and similar modern works), specimen data, and beyond--as a tool for those working in the region, and as an example of how mobilising collections and research data can serve the world in a number of ways
- Smithsonian Institution Libraries took up the task of funding and implementing the first phase: “The Electronic *Biologia-Centrali Americana*”



# *The Electronic Biologia Centrali-Americana*



- create images in multiple formats of all 40,000 pages of the 58 biological volumes
- work with the taxonomic community to create a DTD (Document Type Description) for taxonomic literature
- code in eXtensible Markup Language (XML) the full text
- provide some facility to link to specimen, taxonomic, and geographic data
- make the entire project freely available on the World Wide Web



# *The Biologia Centrali-Americana* Centennial Project



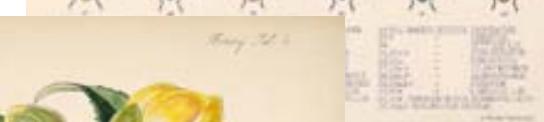
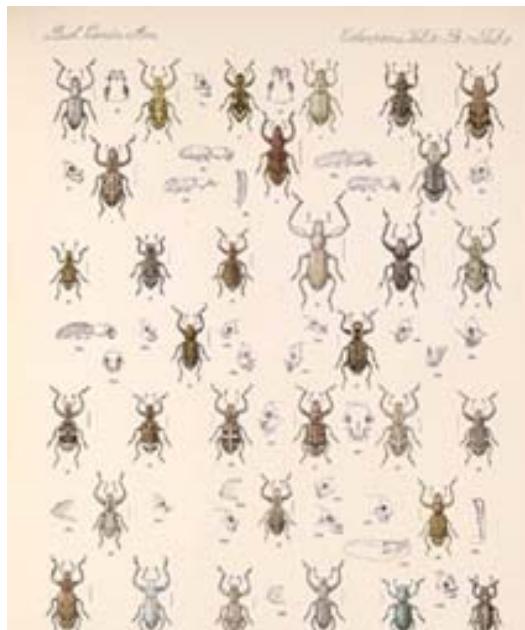
Collaborative project includes, but *not* limited to:

- Smithsonian Institution (NMNH, SI Libraries, STRI)
- Natural History Museum (London)
- The National Commission for the Knowledge and Use of Biodiversity, Mexico (CONABIO)
- Instituto Nacional de Biodiversidad, Costa Rica (INBio)
- Missouri Botanical Garden
- American Museum of Natural History
- Royal Botanic Gardens, Kew
- Museo Entomologico de Leon, Nicaragua
- Global Biodiversity Information Facility



# The Biologia Centrali-Americanana Centennial Project

- All pages now turned into JPEGs
- Project web page:  
[http://www.sil.si.edu/  
bcaproject](http://www.sil.si.edu/bcaproject)



1. *Trichosanthes verticillata*. (Tob. XX, fig. 22, &c.)

*Schizanthus verticillatus*, Sch. in Schlecht. Gen. Comm. in, p. 719 (post) (see esp. id. id., p. 171, § 1).

*Trichosanthes* (Trichosanthes), in Mex. (Méj.), Guanajuato, Teguaj, Edia (Saltillo, Coahuila (J. H. Smith), Jalapa (M. Trujillo), Mexico city, Cárdenas de Plazas, Oaxaca (Méj.), Cuernavaca, Cuicatlán (U.S. Nat. Mex.), Puerto de Ixtla (Hidalgo), Tehuantepec (Méj. Ixtl.).

The name *verticillata* is here applied to the form common in Veracruz and Oaxaca; the form subsequently described in Schlecht's work is no doubt referable to *T. micros*. The present species (which is synonymous with *T. verticillata*, Sch., by Leconte, and reported as possibly identical with *T. micros*, Leuc., by Casey) has the rostrum short and sharply bent downward from the base in both sexes (fig. 22), the pubescence usually with a large space on the flanks almost bare, and the depressed space on the basal half of the abdomen of the male thickly clothed with coarse, long, radiating scales. *T. micros* is a little less elongate than *T. micros*, the rostrum is less curved (appearing more sharply gibbose at the base), the median space on the basal segments 2 and 4 is denuded at the base only, and the depression of the male is elevated





*“Although the ‘Biologia’ contains the record of such a large number of species, it is but a fragment of what may yet be obtained. The whole work must be looked upon as only a contribution to our knowledge of the subject, and I hope it may be an incentive to others to carry it further.”*

- F. DuCane Godman



[BCA DIGITAL EDITION](#)

[ABOUT THE BCA](#)

[PROJECT STATUS](#)

[PROJECT BACKGROUND](#)

[PARTNERS](#)

[KEY CONTRIBUTORS](#)

[RESOURCES](#)



Smithsonian Institution Libraries



Smithsonian  
*National Museum of Natural History*



[credits](#) || [permissions](#)

# *The Electronic Biologia Centrali-Americana*



- [Navigation page](#)
- [Image example](#)



[Go to Page](#)[Reset](#)

CHIROPTERA.

[previous](#) | [next](#)

## 16. STURNIRA.

*Sturnira*, Gray, Ann. & Mag. Nat. Hist. x. p. 257 (1842).

*Sturnira* differs conspicuously from the preceding genera in the rudimentary state of the interfemoral membrane—which is reduced to a narrow fringe along the hind legs, hidden by the fur. The chin is marked with a flat median protuberance, surrounded by smaller warts; the molars are longitudinally grooved; and, as in the next genus, the males are decorated with an epaulet-like tuft of light-coloured hair on each shoulder. Two species have been generally recognized, but only one is accepted by Mr. Dobson. Its colour is variable, usually brown, more or less washed with red; and the forearm measures about 1<sup>11</sup>·70.

### 1. *Sturnira lilium*.

*Phyllostoma lilium*, Ét. Geoffroy, Ann. du Mus. xv. p. 186 (1810, descr. orig.)<sup>1</sup>.

*Sturnira spectrum*, Gray, Ann. & Nat. Hist. x. p. 257 (1842, descr. orig.)<sup>2</sup>; Zool. Voy. ‘Sulphur,’ p. 22, pl. vi. fig. 1<sup>3</sup>.

*Stenoderma chilense* (Gervais), Gay, Hist. de Chile, Mamm. p. 30, pl. i. fig. 1 (1847, descr. orig.)<sup>4</sup>.

*Sturnira chiloensis*, Frantzius, Arch. f. Naturg. xxxv. 1, p. 262 (nec *Vespertilio chiloensis*, Waterh.)<sup>5</sup>.

*Sturnira lilium*, Dobson, Cat. Chir. Brit. Mus. p. 538, pl. xxx. fig. 4<sup>6</sup>.

Hab. GUATEMALA (*Mus. Berol.*); HONDURAS (*Dyson, Mus. Brit.*)<sup>6</sup>; COSTA RICA (*Frantzius*<sup>5</sup>,

& *Hoffmann, Mus. Berol.*)<sup>7</sup>; WEST INDIES<sup>6</sup>; PERU; BRAZIL; PARAGUAY; CHILI<sup>4</sup>.

# *The Biologia Centrali-Americana* Centennial Project



In the next phase:

- Data in the BCA text will be searchable and will be able to be addressed with web tools
- Non BCA data and images will be accessible from the BCAC via hyperlinks
- Ultimately, data from all sources will be interoperable and treatable by web-based analytical tools
- The first step is an XML schema for Taxonomic Literature (TaXMLit)



# *The Biologia Centrali-Americana Centennial Project*



## XML Definition

- "language"**      **A Standard Methodology with Formal Syntax**
- "markup"**        **for Adding Information to a Document Relating to its Structure and/or Content**
- "eXtensible"**     **by Applying Identifiers for Elements of Information in a Neutral Way, Stored in a Neutral Form, Independent of Systems, Devices, & Applications**

NAVEA NSWG Carderock Code 205 <http://navycals.dtic.mil> gamertf@nswccd.navy.mil

19



XML is a way to *structure, describe,* and *interchange* electronic data



# The Biologia Centrali-Americana Centennial Project

## 2. Species Trichobaris mucorea

*Baridius mucoreus* Lec. Proc. Acad. Phil. 1858 p. 79 '1868 p. 364 2.

*Trichobaris trinotata* var. *mucorea*, Lec. Proc. Am. Phil. Soc. xv 288 3

*Trichobaris mucorea*, Casey Ann. N. York Acad. Sci. xv. pp. 562, 564 4

Hab. NORTH AMERICA, Southern California and Arizona 4, Texas; LOWER CALIFORNIA 4. MEXICO, Mexican boundary (Morrison), Ventanas (Forrer), San Blas (U.S. Nat. Mus.), Durango city (Höge).

Specimens of this species (♂ ♀) from San Blas and other localities in N.W. Mexico agree perfectly with those before me from California and Texas. The vestiture of the ventral depression of the male, as stated by Casey, is uniform with that of the rest of the under surface, and the median space on the segments 3 and 4 is almost entirely bare. The San Blas examples are labelled as having been found on tobacco. *T. mucorea* is known in the United States under the name of the "Tobacco-stalk weevil," and it is also said to attack *Solanum carolinense* and *Datura stramonium* and *D. tatula* [cf. Bridwell, U.S. Dep. Agric., Div. Ent., Bull. no. 44, pp. 44 46 (1904)].



```
</TaxonDiscussion>
</TaxonTreatment>
<TaxonTreatment RankDesignation="Species">
  <TaxonNumber>2. </TaxonNumber>
  <TaxonName PublishedTextAfter=".">
    <GenusName>Trichobaris</GenusName>
    <SpeciesEpithet>mucorea</SpeciesEpithet>
  </TaxonName>
  <CitationGroup>
    <PrimaryCitations>
      <PrimaryCitation>
        <TaxonName>
          <GenusName>Baridius mucoreus </GenusName>
        </TaxonName>
        <TaxonAuthors>
          <TaxonAuthor>Lec.</TaxonAuthor>
        </TaxonAuthors>
        <Publication> Proc. Acad. Phil. </Publication>
        <Volume>1858</Volume>
        <Pagination>p. 79 '1868</Pagination>
        <Volume>1868</Volume>
        <Pagination>p. 364 2.</Pagination>
      </PrimaryCitation>
    </PrimaryCitations>
    <Synonyms>
      <Synonym KindOfSynonym="Original Name of Accepted">
        <TaxonName>
          <GenusName>Trichobaris </GenusName>
          <SpeciesEpithet>trinotata.</SpeciesEpithet>
          <RankBelowSpeciesAsStated>var.</RankBelowSpeciesAsStated>
          <EpithetBelowSpecies>mucorea </EpithetBelowSpecies>
        </TaxonName>
        <Publication>Lec. Proc. Am. Phil. Soc.</Publication>
        <Volume>xv</Volume>
        <Pagination> 288 </Pagination>
        <CrossReference CrossReferenceID="someID">3</CrossReference>
      </Synonym>
      <Synonym KindOfSynonym="Original Name of Accepted">
        <TaxonName>
          <GenusName>Trichobaris </GenusName>
          <SpeciesEpithet>mucorea</SpeciesEpithet>
        </TaxonName>
        <TaxonAuthors>
          <TaxonAuthor>Casey</TaxonAuthor>
```



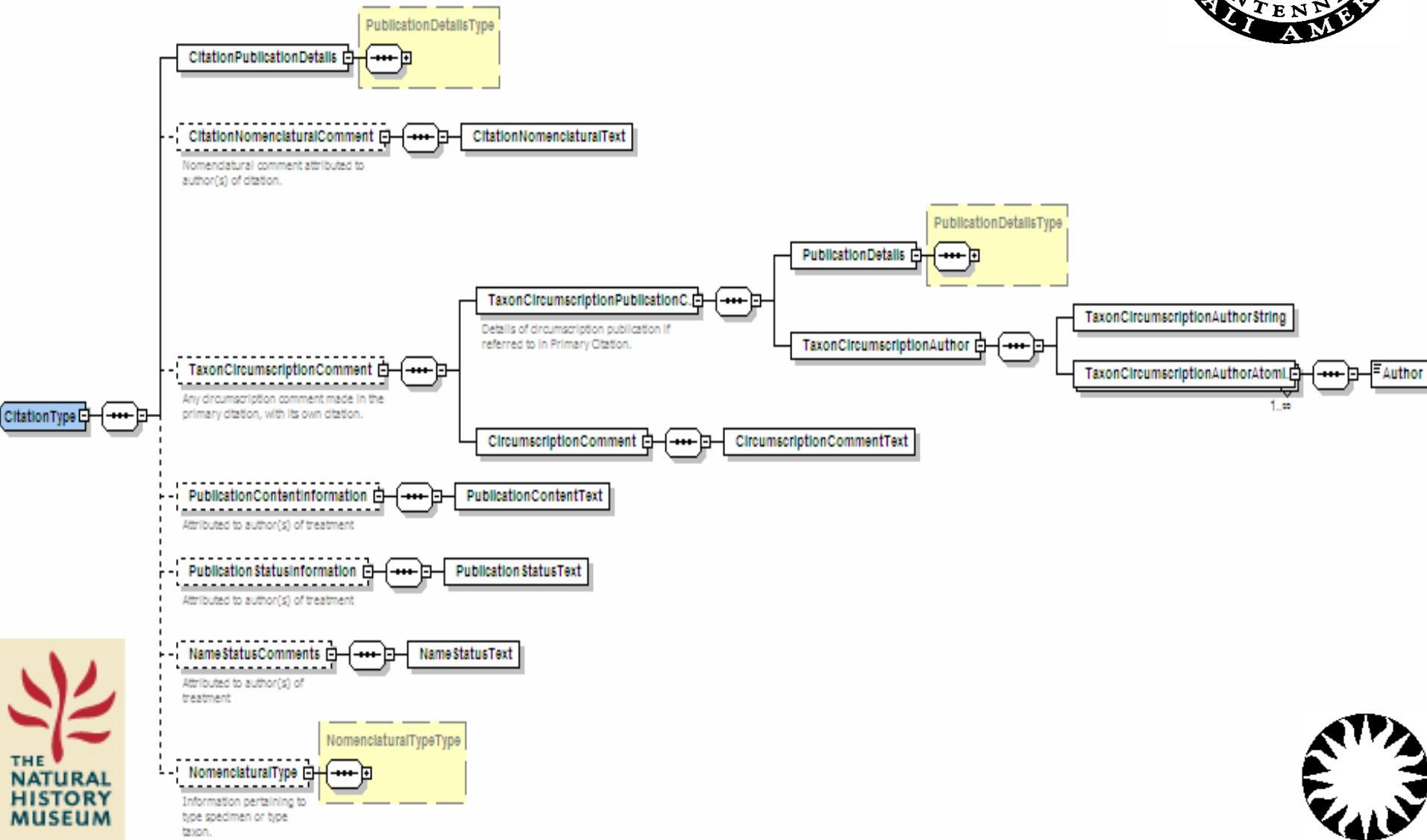
# *The Biologia Centrali-Americana* Centennial Project



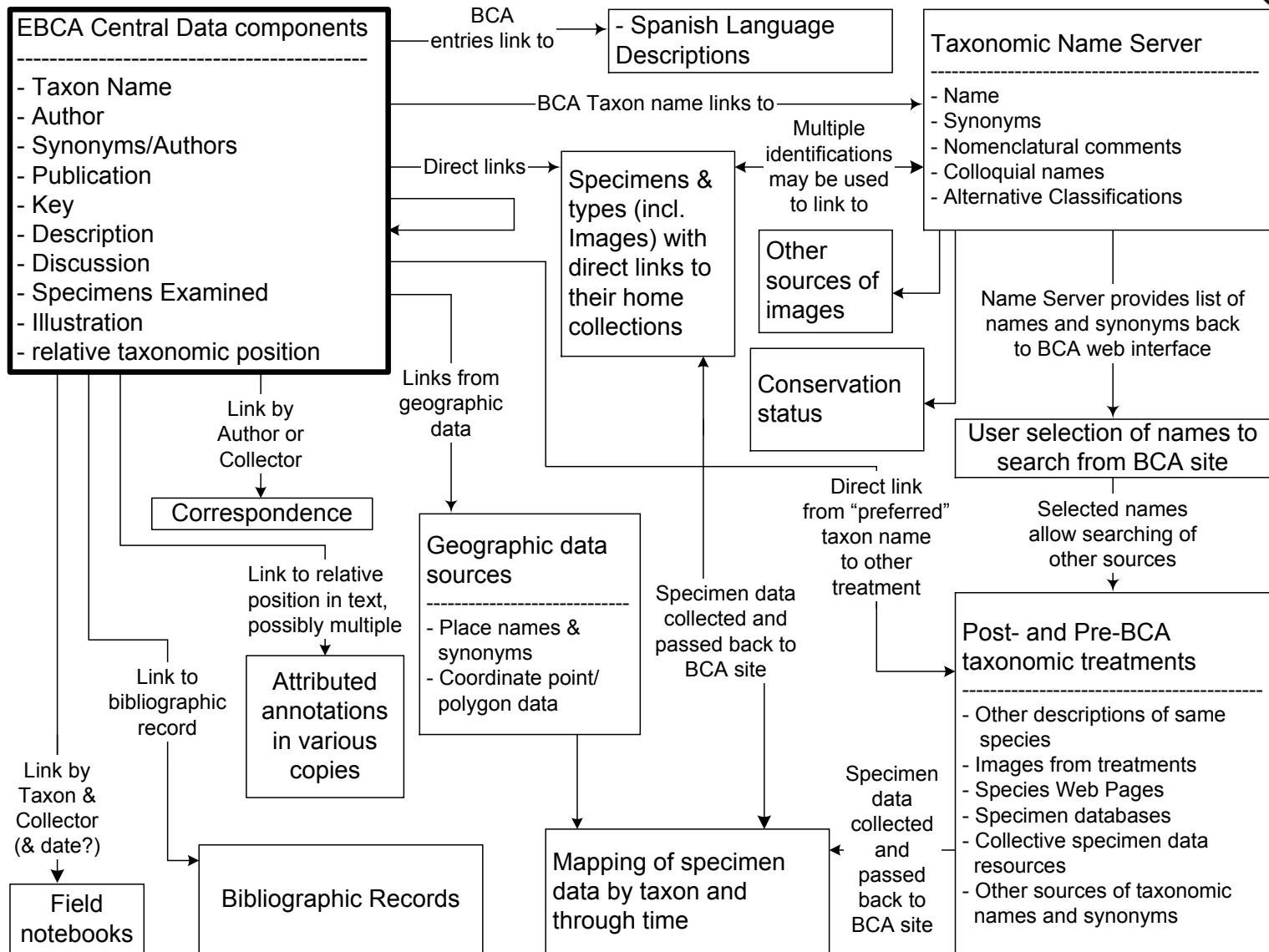
- Marked up text gives a flexible output, which can be called on in appropriate formats
- More importantly, it can be used to link to and be addressed with other digital data
- Standards are needed  
TDWG, GBIF, ABCD, taXMLit, SEEK, etc



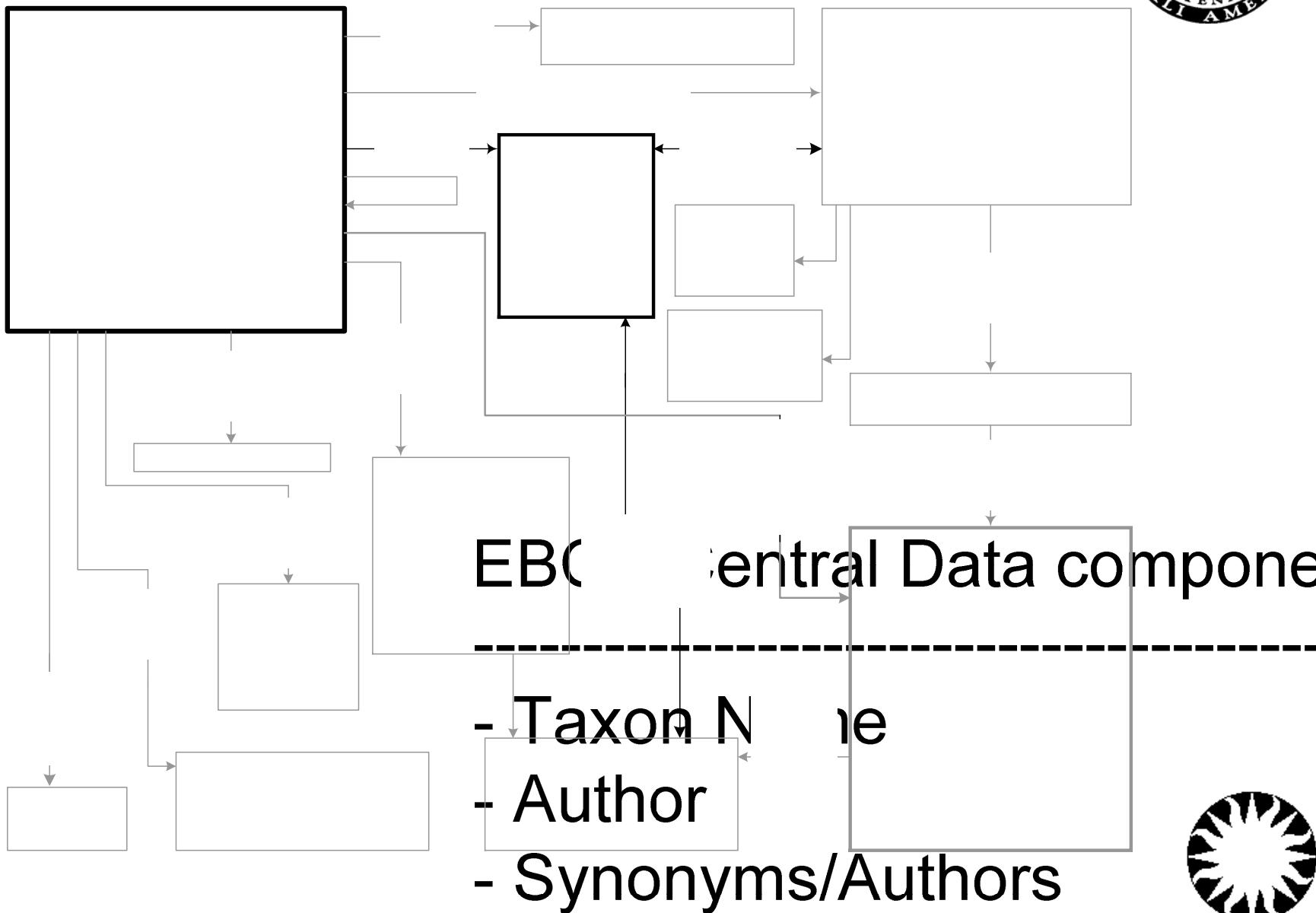
# The *Biologia Centrali-Americana* Centennial Project



# The Biologia Centrali-Americana Centennial Project – data model

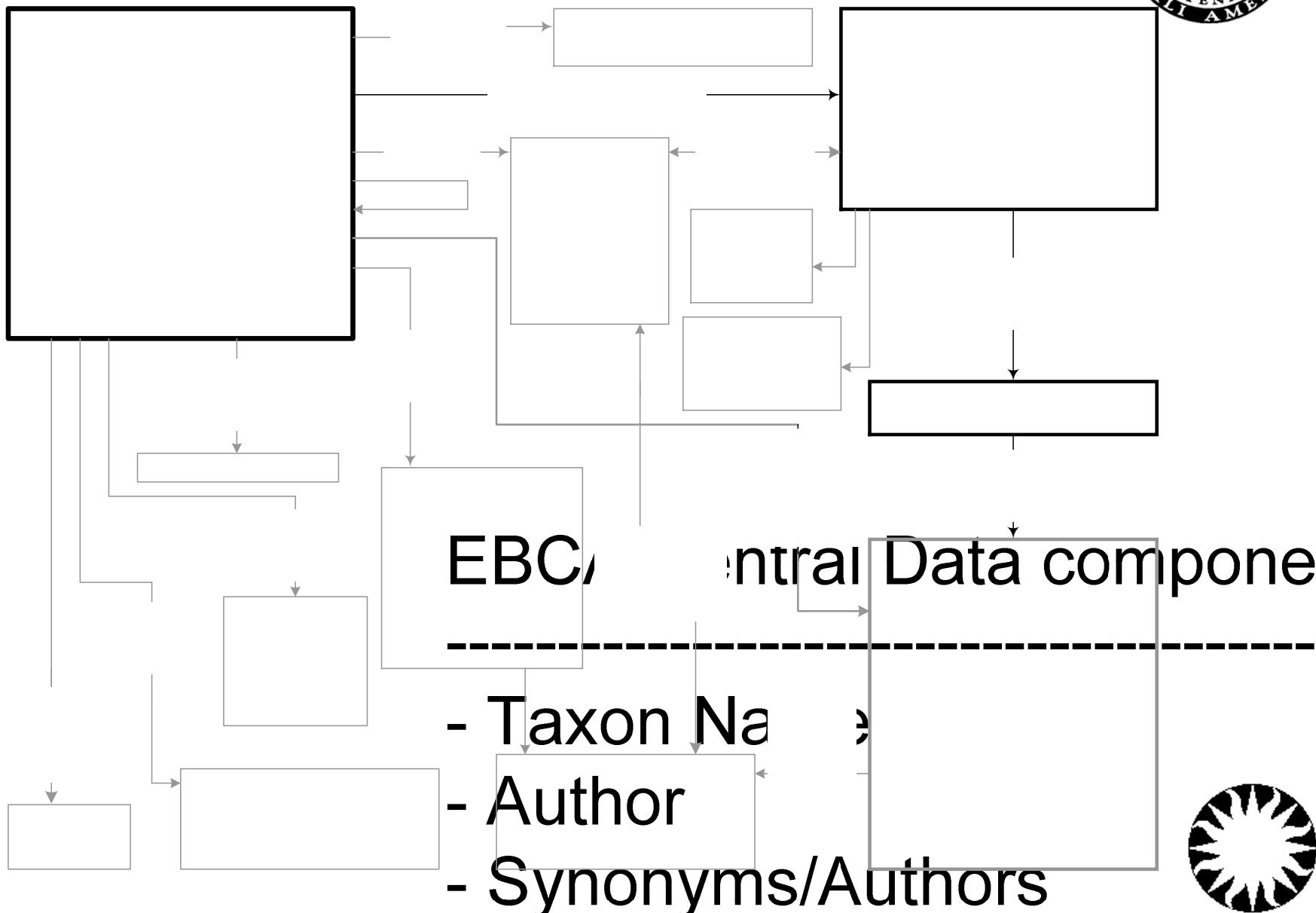


# *The Biologia Centrali-Americana* Centennial Project – specimens



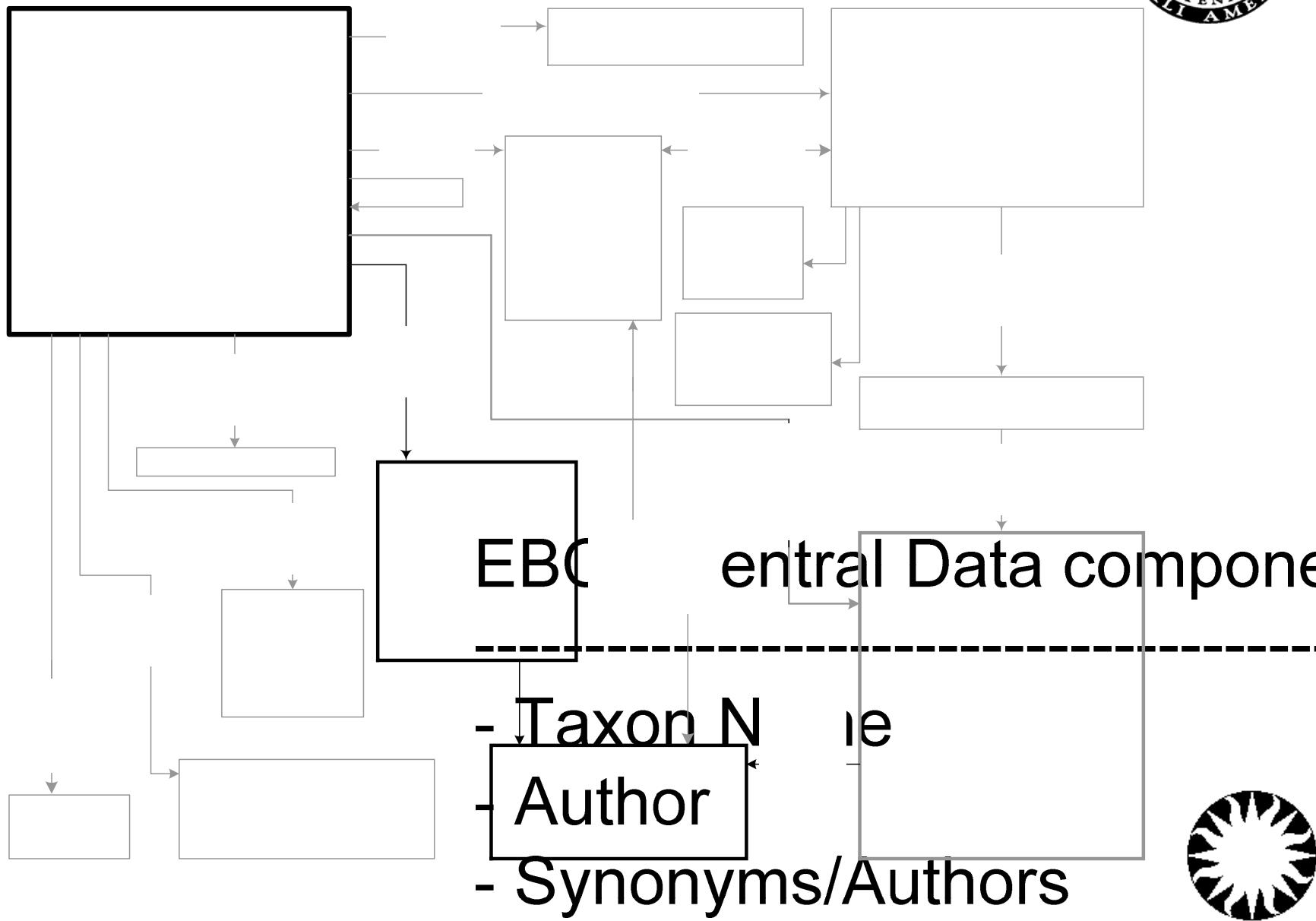
# *The Biologia Centrali-Americana*

## Centennial Project – Name Server

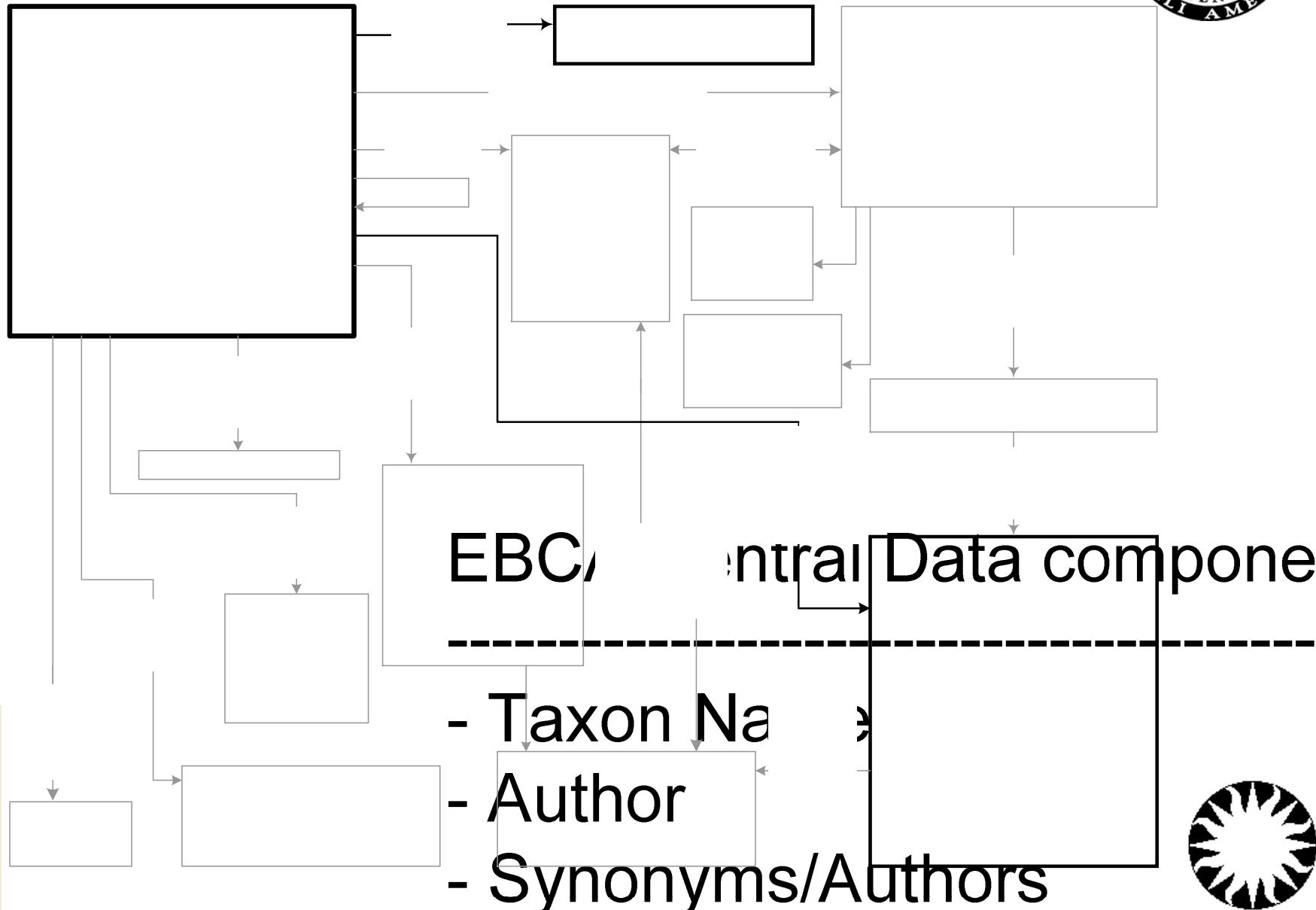


# *The Biologia Centrali-Americana*

## Centennial Project – Geography



# *The Biologia Centrali-Americana* Centennial Project – other treatments



# *The Biologia Centrali-Americanana Centennial – prototype designs*

## Electronic Biologia Centrali-Americanana

[Browse BCA by Volume](#)

[Browse BCA illustrations](#)

[Browse BCA by Taxon](#)

[Browse BCA by Author](#)

[Browse BCA by Geography](#)

[Search BCA by Taxon](#)

SEARCH

[Search BCA by Author](#)

[Search BCA by Collector](#)

[Search BCA by Geography](#)



## Biologia Centrali-Americanana Centennial

[Browse BCAC by Taxon](#)

[Browse BCAC by Author](#)

[Browse BCAC by Geography](#)

[Search BCAC by Taxon](#)

SEARCH

[Search BCAC by Author](#)

[Search BCAC by Collector](#)

[Search BCAC by Geography](#)

# *The Biologia Centrali-Americana* Centennial Project



Next stages: core capabilities

- XML schema implemented for all BCA biological volumes and made available on web
- Links between BCAC and collection databases (partner institutions, GBIF, BioCASE, REMIB etc) to call up data for BCA species
- Links to Taxonomic Name Servers and others to enable species dictionary component (GBIF, uBio, CoL etc) including colloquial names
- Links to BCA locality gazetteer (AMNH)
- Links to extant national and regional checklists
- Links to specimen images
- Link to web-based analytical tools and other datasets (GIS)



# *The Biologia Centrali-Americana* Centennial Project



Next stages: core outputs

- Species lists at multiple geographic levels, linking valid/current names and synonyms
- Specimen database for Mesoamerica
  - From linked collections;
  - From the BCA itself!
- Online descriptions and images of the majority of Mesoamerican biota



# *The Biologia Centrali-Americana* Centennial Project



Next stages: develop information base for selected taxa

- Focussed multi-institutional specimen data collection
- Schema applied to other taxonomic literature of selected groups and linked to BCAC
- Links to key resources for selected groups (e.g. *Flora Mesoamericana*)
- Add facility to upload new descriptions, taxonomic acts and other data to BCAC (e-publishing)

