The "Electronic *Biologia Centrali-Americana*" and the "*Biologia Centrali-Americana* Centennial"

A joint project of
Smithsonian Institution (National Museum of Natural History,
Smithsonian Institution Libraries, and Smithsonian Tropical Research Institute)
Natural History Museum (London)
Missouri Botanical Garden
National Commission for the Knowledge and Use of Biodiversity, Mexico (CONABIO)
Instituto Nacional de Biodiversidad, Costa Rica (INBio)
American Museum of Natural History
Harvard University (Museum of Comparative Zoology)
Royal Botanic Gardens, Kew
Museo Entomologico de Leon, Nicaragua
Global Biodiversity Information Facility

Abstract

The collaborating institutions aim to create a model for the global effort to facilitate access to the data necessary for understanding the world's biota. The flora and fauna of the Mesoamerican region will be used as a basis, starting with an important and out-of-print scientific work, the *Biologia Centrali-Americana* (BCA). This is the first part of a multi-phase project that will create an extraordinary new set of resources and knowledge tools in electronic form for biodiversity studies centered on Mexico and Central America. The methodologies developed in this project will be applicable to other digitization efforts.

The Smithsonian Institution Libraries (SIL) has obtained necessary funding for the first phase of the project. This is the 'Electronic *Biologia Centrali-Americana*' (EBCA), in which the full text of the BCA will be scanned, rekeyed, and coded in eXtensible Markup Language (XML). The original BCA was derived from scientific surveys and explorations conducted during the latter part of the 19th century and early 20th century. Many of the leading biologists of the time provided specimens and descriptions for the many volumes. The illustrations are in many cases the only images that exist of the biota of the region. The EBCA will be designed so that links can be made from the coded text to other vital biological datasets, including specimens, lists of names of taxa, taxonomic literature, and current treatments of the biota. This will form the '*Biologia Centrali-Americana* Centennial' (BCAC). Both first and subsequent phases of the project will benefit everyone who deals with biodiversity in Mexico and Central America, including biologists, conservation groups, land planners, natural resource managers, and quarantine officials.

The project will be the keystone and model for several other major bioinformatics projects to be pursued by the leading biological repositories in the world, in order to speed the pace of scientific investigation into the nature of the rapidly changing natural environment.

Background

A repeated message from those interested in conservation of biodiversity around the world, especially those in biodiversity-rich but resource-poor countries, is the need for taxonomic information. This is necessary for a wide range of environmental management and conservation purposes, as well as being a basic tool for education and enjoyment of the natural world. This issue has been identified as a part of the 'taxonomic impediment' - the lack of taxonomic information, skills, personnel and capacity inhibiting

many developing countries from implementing policies and practices of sustainable management and conservation of biodiversity. In particular, under the Convention on Biological Diversity (CBD), methods of dealing with the taxonomic impediment have been elaborated under the "Global Taxonomy Initiative" (GTI). Within the Work Programme of the GTI the need to make available the contents of taxonomic literature and details of material held in collections outside the countries of origin is highlighted several times.

At an October 2001 conference organized by the Smithsonian Institution and funded by the Andrew W. Mellon Foundation, *Toward Collaborative Biodiversity Informatics: Mobilizing Collections and Research Data*, key representatives from the American Museum of Natural History, The Natural History Museum (London), the Royal Botanic Gardens (Kew, UK), Missouri Botanical Garden, and the Smithsonian Institution (National Museum of Natural History, Smithsonian Tropical Research Institute, and the Smithsonian Institution Libraries) met to address the problems of managing and accessing the information embedded in large biological repositories.

Natural history museums and similar large biological repositories and their libraries hold a wealth of inadequately accessible resources that describe and explain the diversity and depth of life on earth. Mining these data for research, conservation, drug discovery, protected area management, disease control, etc., is difficult, time consuming, and often leads to redundant efforts. What should be a seamless, open "book" of knowledge consists, instead, of disparate, unintegrated sets of data - some in electronic form but most still on paper, published and unpublished. The conference focused on the four following types of biological datasets:

1. *Specimen collections in large repositories*. Many large biological repositories, including all of the participants, are converting manual records about their vast collections of biological specimens into integrated electronic collection information systems.
2. *Taxonomic databases* that record the names, classification, synonymy, geographic distributions and relationships of biological organisms.
3. *Published taxonomic literature*, including journal articles, monographs, and other forms of publication that name and describe taxa, details of collection, and other information.
4. *Geographical information systems* (GIS) that link geographic place names and other geographic data elements with precise geospatial coordinates. Once large numbers of specimens have been georeferenced, aggregate studies may be performed, such as species distribution over time.

These datasets are part of a larger, worldwide effort to enable easy access to the complete range of data required to understand individual species and their environmental and evolutionary relationships. This will require the establishment of cross-linkages between, and simultaneous access to, data sets from such information sources throughout the world.

To deliver an immediate product, provide a testbed for developing protocols and technologies for linking these datasets in a coherent, usable way, and act as a model for further projects, the conference participants agreed on an immediate high-priority project based on the *Biologia Centrali-Americana* (BCA), one of the most important compendia of the flora and fauna of Mexico and Central America. The project has been developed as the 'Electronic *Biologia Centrali-Americana*' (EBCA).

History and Importance of the *Biologia Centrali-Americana*

According to Dr. Sandra Knapp, Research Botanist, The Natural History Museum (London), and one of the participants in the October 2001 Conference,

"…the inaccessibility of literature relating to tropical plans and animals…leads to waste of effort on the part of taxonomists from biodiversity-rich countries, either through re-description of that which has been done before, or in just trying to get access to what they need. As more and more literature becomes available over the Internet, it is crucial that old literature be concomitantly more available, as the combination of old and new is essential to the good practice of taxonomy and systematics. Almost uniquely, systematics relies upon its past to define its future."[1]

The *Biologia Centrali-Americana* is a fundamental work for the study of neotropical flora and fauna and includes nearly everything known about the biological diversity of Mexico and Central America at the time of publication. The BCA was privately issued in installments between 1879 and 1915 by F. Ducane Godman and Osbert Salvin of The Natural History Museum (London). "The work consists of 63 volumes containing 1677 plates (of which more than 900 are coloured) depicting 18,587 subjects. The total number of species described is 50,263 of which 19,263 are described for the first time."[2] The contents were derived from scientific survey and explorations conducted during the latter part of the 19th and early 20th centuries. Many of the leading biologists of the time provided specimens and descriptions for the many volumes. The illustrations are, in many cases, the only images that exist of the biota of the region and as such could be compiled for use in an electronic field guide if available in a digital and portable format. The specimens described are deposited in many places including The Natural History Museum (London), Royal Botanic Gardens (Kew, UK), Missouri Botanical Garden, American Museum of Natural History, Harvard University, and the National Museum of Natural History (Smithsonian).   Since the BCA appeared, a few select volumes have been republished but never the entire series.

Based on research in the OCLC international bibliographic database, which contains cataloged holdings from tens of thousands of libraries, the entire 63-volume BCA is held by only eight libraries. Many other libraries hold individual volumes. Some Central American countries, whose flora and fauna are so well documented in BCA, lack a complete set; thus the BCA is not generally accessible to taxonomists working in the region.

The BCA is important to biodiversity studies and national biological surveys at such scientific centers as the Smithsonian Tropical Research Institute, INBio (Instituto Nacional de Biodiversidad) in Costa Rica, CONABIO (The National Commission for the Knowledge and Use of Biodiversity) in Mexico, and the Humboldt Institute in Colombia. However, Dr. Knapp reports:

"….This project is not just about providing the original data for biologists in developing countries to use to empower them to adequately meet their obligations under the Convention on Biological Diversity. This is the first step in helping to remove the taxonomic impediment--and it is critical that the large museums of the world, who hold so much of the data in the form of literature and specimens for the rest of the world, be the ones to make this major attempt."[3]

The Electronic *Biologia-Centrali Americana* (EBCA)

The EBCA will replicate the full text of the BCA. This is important not only in its own right, but as a testbed for the linkages with other data. The full text of the BCA will then be available to any researcher with an Internet connection anywhere in the world. Researchers who must now travel significant distances to use these texts can access them from their desks. Downloaded elements or parts captured on disk and used on non-connected machines, will enable workers in remote locations to work with the

---

[1] Email from Dr. Sandra Knapp to Thomas Garnett, 2 Dec 2001
[2] Prospectus, Biologia Centrali-Americana, p. 4.
[3] Knapp, op cit.

animals and plants. Additional follow-on projects may include production of portions of the digitized BCA on searchable cd-roms or on a chip that can be loaded into a handheld computer for use in the field.

The EBCA will be in two parallel formats. Firstly, the volumes will be reproduced as a set of pdf files, to enable the original images and text pages of the BCA to be seen. A copy of the BCA held in the Smithsonian Institution Library will be used, although comparisons of the quality of plates with copies held in the Natural History Museum, London, will be made to ensure that the best versions are made available. Secondly, the text will be rekeyed and coded in eXtensible Markup Language (XML). This will not only enable the text to be searchable, but also make it possible to emplace direct links to ancillary information such as glossaries, gazetteers, bibliographies and other resources to enable users to work with the BCA more effectively. Links will also be made from the EBCA text to databases of BCA specimens and species held in different collections, including those of Smithsonian's National Museum of Natural History, Harvard University, The Natural History Museum (London), The Missouri Botanical Garden, and the Royal Botanic Gardens, Kew.

EBCA Project Benefits

Use by researchers in the region and worldwide
The BCA is essentially an early database of biodiversity of Mexico and Central America and thus forms the basis for subsequent advances in knowledge, as well as the foundation for studying changes caused by natural or human factors (e.g., invasive species). For many groups of organisms it is still the most recent review for the region. It also contains information that will help countries of the region identify the composition of their flora and fauna. Making the information held in the BCA widely available electronically, especially to Mexico and the countries of Central America, will enable a significant increase in capabilities to identify and work with the biodiversity of that region. The extraordinarily detailed and finely engraved plates of the BCA form the basis of an unparalleled field guide to the fauna and flora. The linkages put in place to extant databases, together with additional collection databasing, and updates of nomenclature, will provide baseline data on country biota for nations who at present would find that information virtually impossible to obtain.

Users requiring access to the information in the BCA will include taxonomists, conservationists, ecologists, agriculturalists, quarantine officers, interested amateurs, and those advising governments and others on the distributions of animals and plants in the various countries. Historians of science and other fields will find data available for their research. Libraries who lack copies of the originals will also benefit.

As a model
The BCA, while certainly one of the most comprehensive and broad-ranging, is only one of a large number of similar works that would benefit the research community through being more accessible. By providing a model of how a work like this can be successfully translated to a digital form, the stage will be set for others to engage in similar projects. Institutions pursuing similar projects are already showing interest in using BCA as a model.

As a basis for research and development for digitization of taxonomic literature
The digitized BCA will engage the systematic biology community in general by providing an experimental testbed for determining the best ways of linking different, yet complementary, sets of data. A large part of the testbed will consist of defining the practices and standards needed to ensure effective crosswalks between, and links among, relevant biological data systems, including appropriate specimen images, especially of types. In particular, developing an XML-based standard for coding taxonomic literature will be defined as part of this project. As a recent report from the Council on Library and

Information Resources stated, successful digitizing projects "will fit technologically into ("interoperate" with) broader collections so that they can be accessed through a variety of search engines and portal services far beyond a single campus."[4]

As a catalyst and vehicle for making taxonomic and associated information more accessible
Dynamic links will be made to information held in different web sites or held in different ways. This project is expected to be the keystone for several other major bioinformatics projects to be pursued by the leading biological repositories in the world. For example, the American Museum of Natural History has committed to fund a project to rekey and code the text of *A Gazetteer to Accompany the 'Insecta' Volumes of the 'Biologia Centrali-Americana'*, by Richard B. Selander and Patricia Vaurie.[5] This will be linked to the EBCA and to community-based GIS systems, such as the Alexandria Digital Library Gazetteer at the University of California, Santa Barbara.[6] In another project, the Missouri Botanical Garden plans to link the EBCA to the *Flora Mesoamericana* (FM), a joint effort of NHM (London), the Missouri Botanical Garden, and the Instituo de Biología of National Autonomous University of Mexico, to provide the historical context. These links are now being transcribed manually, for the FM's approximately 60,000 names and 17,000 species. The collaborators plan to generate several additional grant proposals to other funding sources to help forward this work.

The *Biologia-Centrali Americana* Centennial (BCAC)

The catalytic possibilities noted above are perhaps only the beginning. The BCA (and thus the EBCA) is a publication of the late 19th and early 20th centuries, and the information contained is fixed in time. Moreover, while encompassing available knowledge about a vast portion of the biota of the region, the BCA does not include all of it, even that available at the time of publication. However, the availability on line of the basic taxonomic information provides a core to which many other data can be linked. Through these links, the BCA can be brought up to date in its content, and expanded to include data and tools never dreamed of by its original compilers. The EBCA is being designed to enable such links to be made, so as to maximize the accessibility of information on Mesoamerican biota and provide a set of tools for researchers, environmental managers and advisors, conservationists and educators, as well as being a basic tool for education and enjoyment of the natural world. This broader project has been termed the '*Biologia Centrali-Americana* Centennial', or 'BCAC'.

The BCAC is envisioned as a project to which many will contribute, and which will have a very wide 'ownership'. The partners in this process are expected to include those listed above but also other organizations especially those in Mexico and Central America. The BCAC will solicit contributions from the taxonomic and wider community (not only including taxa included in the original BCA, but expanding to all other groups, including marine taxa). No one team will be able to provide all of the information that will serve to complete it, and indeed it may never be 'complete', in he sense that it covers 100% of all Mesoamerican species and all of the possible information and analytical tools. However, as increasing numbers of workers use it and contribute to it, not only will it grow in content, but also more uses of it will be devised and developed.

---

[4]*CLIRinghouse* (April 2002), Washington, D.C., Council on Library and Information Resources.
[5] Richard B.Selander and Patricia Vaurie. *A Gazetteer to Accompany the "Insecta" Volumes of the "Biologia Centrali-Americana,"* New York: American Museum of Natural History, 1962. American Museum Novitates. no. 2099.
[6] The Alexandria Digital Library is a working digital library with collections of geographically referenced materials and services for accessing those collections. The ADL Gazetteer contains 4.5 million place names. It is headquartered on the campus of the University of California at Santa Barbara and hosted by the Davidson Library (http://fat-albert.alexandria.ucsb.edu:8827/gazetteer/)

Two classes of information will be accessible through the BCAC: links, defined by BCAC collaborators, to existing and novel data placed on the web (e.g., descriptions, keys, images, checklists, discussions of taxa, databases), and dynamic interfaces to a wide variety of data sources, (e.g., specimen databases, name servers, digital gazetteers, mapping tools, a wide variety of search engines for generally available data and images).

As an example of the first class of data, the BCAC will provide access to digitized publications of descriptions of Mesoamerican taxa published before the BCA, after the BCA (including newly-described taxa), and in groups not included in the BCA. Some of these will be contributed directly to the BCAC by partner taxonomists, and the possibility of establishing the BCAC as an electronic journal for the publication of such information is being discussed. Examples of such added descriptive work are given on the project site (http://www.sil.si.edu/BCAProject/prototype.htm).

With the development of tools to facilitate global access to specimen databases and taxonomic name servers, such as are being established by GBIF[7], dynamic linkages will be added to the BCAC. In combination with appropriate tools, these will enable a number of additional possibilities, such as the "on the fly" preparation of check-lists of fauna and flora at all levels from local to regional and distribution maps. Such maps could be time-sliced, or be linked to climatic and ecological data enabling predictive analysis of species ranges. These products will be based on the most recent identifications in all available collections databases and, where possible, with updated information available from taxonomic name servers and/or electronic gazetteers.

The BCAC will be available in parallel with the EBCA. The BCAC will include additional taxa, and, at least for some time, the EBCA will have taxa that will not be treated in BCAC. For those taxa that do overlap, the BCAC will include a series on intermediate web pages linked from/to the taxa in the EBCA. These will provide links to BCAC content and reverse links to the EBCA and BCA. The current model of this page, with explanations, is provided in Appendix 1 below. Examples of BCAC pages are on line at the project web site (http://www.sil.si.edu/BCAProject/prototype.htm).

Project elements and functionality of the EBCA and the BCAC

A. *Existing funding from the Seidell Endowment will support the following tasks for the EBCA:*
1. The cataloging of the BCA in the Smithsonian Institution Research Information System (SIRIS) will be updated. Because the records in SIRIS were converted from catalog cards created under differing cataloguing practices and do not always reflect current locations, the records are frequently inaccurate and misleading. A cataloguing contract will be employed for this step.
2. The entire set of 58 biological volumes of the 63-volume set of the *Biologia Centrali-Americana* will be scanned, and a digital edition of the BCA created and mounted on a server at the Smithsonian. This digital edition will provide high-resolution, bit-mapped images of the entire text and figures with a navigational device. All of the pages will be made freely available through the Internet to anyone with a web browser.
3. A web application supporting xml and serving and linking the full text of the BCA and related data will be implemented. As far as possible this system will be standards-based and integrated into the Smithsonian information technology architecture. The system will support searching, updating, linking, and management of the EBCA data.
4. The text of the BCA will be rekeyed.

---

[7] The Global Biodiversity Information Facility; see http://www.gbif.org

5. While being rekeyed, the text will be marked up and coded in such a way as will allow full-text searching, easier on-screen reading and navigation, easy reformatting of portions of the text for multiple uses (e.g., Adobe Acrobat files) including downloading and exporting, and integrating with other systems, etc. The general portions of the text (e.g., title pages, introductory sections) will be marked up using a Document Type Description (DTD) developed by the American Museum of Natural History (AMNH) in a project funded by the Andrew Mellon Foundation. Those portions of the text that comprise the taxonomic treatments will be coded using a DTD for taxonomic literature developed as part of this project, in coordination with representatives of the taxonomic community. This schema for coding taxonomic literature will allow for integration with other taxonomic and scientific datasets and the thus coded EBCA text will form a major part of the basis for the BCAC.

6. The document coding and system development will facilitate linking (existing funding may only allow the project to link to some of these as part of EBCA) of BCA to:
   a. collections information systems, first the NMNH Multimedia Catalogue and, as funding and availability allows, other specimen databases including those held by CONABIO, Missouri Botanical Garden (W³TROPICOS), and the NHM (London), including images they contain, especially of type specimens;
   b. taxonomic databases such as the Integrated Taxonomic Information Systems (ITIS), Species 2000, the International Plant Names Index (IPNI), etc.;
   c. standards-based electronic gazetteers such as the Alexandria Digital Library Gazetteer.

7. To locate items within the EBCA, the following kinds of search and browsing capabilities will be supported:
   a. Browse BCA by Volume: The format is still under development, but might be in the form of a graphic representation of the volumes and subdivisions with tables of contents including the major taxonomic group name for each heading added if not present. When the user selects a volume, they go to the title page of that volume in jpg with browsing tools and a choice to get a pdf, or the EBCA coded text version with the rest of the page text also accessible (see main prototype).
   b. Browse BCA illustrations: The user will be presented with an expanding branching diagram (tree) that starts with Mammals, Birds, Fish, Amphibians and Reptiles, Insects, Spiders, Plants, etc. and then breaks down appropriately from there. When a taxon is selected, one or more pages of plates will be available for scrolling. Images will be marked in such a way that the user can then get to the appropriate part of the EBCA coded text from which they can then move to other sections (see main prototype). Alternate methods for the non-taxonomist user, such as by flower color, might also desirable, but this concept is yet to be explored.
   c. Browse BCA by Taxon: The user will be presented with the taxon tree explained above. Links will take the user directly to the EBCA coded text from which they can then move to other sections (see main prototype). Alternate methods for the non-taxonomist user might also desirable, and need exploration.
   d. Browse BCA by Author: The user will be presented with a list of authors of treatments and taxa organized alphabetically, with choices of where in the BCA text the author appears. Links will take the user directly to the EBCA coded text from which they can navigate.
   e. Browse BCA by Geographic Region: The user will be presented with a page with a map of Mesoamerica with country outlines and names. Each country outlined, when clicked, will take a user to a blow up with a map of primary political subdivisions named and outlined. Users can click on either the map or the name to get lists of taxa that they can then use to link to treatments in BCA.
   f. Search BCA: A search box on the front page that allows immediate access to data. There will be a drop down list that refines the possibilities for this search.
      i. Taxon: A search on names at the level of family, genus, or species.
      ii. Author: A search on Authors of treatments and Authors of taxa in BCA.

        iii. Collector: Search on all collectors' names in BCA.

        iv. Geographic region: Search on all geographic regions, country or smaller.

8. Within the EBCA, the following kinds of links will also be supported from the appropriate places in the EBCA coded text:

    a. Links to the facsimile pdf and/or jpeg image description of the BCA.

    b. Links to pdf and/or jpeg image of BCA plate. Ideally this will be cropped to show image(s) of the target species only (this is particularly relevant for images with 5 or more species), with a toggle to allow a view of the whole plate. The methodology and economics of developing this model rather than simply providing an image of the entire plate are being investigated.

    c. Links within the EBCA coded text to and from taxon names embedded in keys, discussions, and elsewhere to the taxon name and the associated synonymy list.

    d. Links within EBCA coded text to discussion(s) if one or more are present. In some cases discussions concerning a particular species are present in more than one place, for example in the text under the species, in the text under a genus, or as a footnote. Discussion elements may be varied, e.g., description; additional morphological information; comparisons between taxa; geographic information (in which case there should be some sort of link to the specimen data reached through 8d below); synonymic information; biological/ecological information; economic information; and discussion of extralimital species. Exploration of costs/benefits from various ways of marking and linking these elements are ongoing.

9. From the EBCA text, the potential to include links to a number of additional resources and external sources will be provided through the coding and system engineering/software development noted in point 3 above although implementation in various groups will depend on available funding and collaborators:

    a. Links to BCAC treatments of each taxon as they are prepared and included.

    b. Links to full Bibliographic record (see 'BCAC Bibliography' below).

    c. Links from geographic and specimen information in the BCA to specimen data and to gazetteer(s) (see BCAC links below).

B. *Projects anticipated for the BCA Centennial (the following elements are part of the larger project and will be contributed by or funded by other agencies). As the digitized BCA volumes are made available, additional funds will be sought from other sources for collaborative projects with many partners.*

1. The text, *A Gazetteer to Accompany the "Insecta" Volumes of the "Biologia Centrali-Americana"* will be rekeyed and coded.[8] This will be linked to community-based GIS such as the Alexandria Digital Library Gazetteer. This project element is the responsibility of the American Museum of Natural History.

2. Practices and standards to ensure effective crosswalks between and links among relevant biological databases, obtain appropriate specimen images, etc., will be defined. This project element will be carried out in close association with the Global Biodiversity Information Facility.

3. Contributions of descriptions, keys, additional information etc from the taxonomic and wider community (not only including taxa included in the original BCA, but expanding to all other groups, including marine taxa. For any one taxon there may be more than one treatment or treatment element in the BCAC. Each entry will be credited to the source.

4. Within the BCAC, the same search and browsing capabilities as for the EBCA will be supported (see previous section).

---

[8] Selander and Vaurie, op cit.

5. Within and beyond the BCA Centennial, the following kinds of links will also be supported, reachable through a front page for each taxon (see Appendix 1 for a concept of this) (see also Appendix 2. Conceptual model of linkages in the BCAC):

   a. Links to the EBCA:
      i. Links to the EBCA coded text, pdf and/or jpeg image description in the BCA.
      ii. Links to pdf and/or jpeg image of BCA plate. Ideally this will be cropped to show an image of the target species only (this is particularly relevant for images with 5 or more species), with a toggle to allow a view of the whole plate. The methodology and economics of developing this model rather than simply providing an image of the entire plate are being investigated.
      iii. Links to data relating to specimens cited in the "*Hab.*" block of EBCA coded text, which lists Country + locality + (Collector) for each specimen/sample. These will be static links to data relating to specimens which have been verified as BCA material by some authority. N.B. the "*Hab"* leader does not occur in botanical texts, though the section is present.
      iv. Links to any key in the EBCA coded text which keys out the species concerned (or other taxon).
      v. Links to a key in the EBCA coded text which keys out subtaxa of the taxon concerned.
      vi. Links to the EBCA coded text to the species name and the associated synonymy list.
      vii. Links to EBCA coded text to discussion if one is present. In some cases discussions concerning a particular species are present in more than one place, for example in the text under the species, in the text under a genus, or as a footnote. Multiple links may be needed in some cases. Exploration of costs/benefits from various ways of marking and linking the different kinds of discussion elements listed above are ongoing.

   b. Links to non-BCA descriptions:
      i. Links to descriptions entered by collaborating taxonomist(s) in the BCAC. The format of description could be either free-style or subject to editorial constraints; this matter is still under discussion. The description itself may contain links (e.g., to images in the BCAC, a glossary of terms stored elsewhere).
      ii. Links to any descriptions available on the web in other sites, entered by collaborating taxonomist(s).
      iii. Links to original published descriptions or protologues, including all text and associated illustrations, if available. This might be placed onto BCAC by collaborating taxonomist(s), or might be hosted elsewhere on another server, requiring a link outside the SI site.
      iv. Dynamic links to descriptions available elsewhere on the web, mediated through the proposed GBIF programme on creating a digital library of biodiversity data or other systems as they become available.
      v. Dynamic links to original descriptions (protologues) available elsewhere on the web, mediated through the proposed GBIF programme on creating a digital library of biodiversity data or other systems as they become available.

   c. Links to non-BCA images:
      i. Links to images (photographs, line drawings) if entered by collaborating taxonomist(s), in the BCAC. Contributors may wish to embed links from the text of their descriptions and keys to individual images.
      ii. Links to other images, placed on the site by a collaborating taxonomist(s).
      iii. Links to images of the type specimens either placed on the BCAC by a collaborating taxonomist(s), or available on an externally-hosted site (in which case the link will be placed by a collaborating taxonomist(s). There may be multiple images. In the case of species with one or more heterotypic synonyms links may be required to all of the types, in which case an intermediate screen allowing a choice to be made will be required.

      iv. Dynamic links to the web available image search tool(s) automatically entering the name to be searched for, e.g., Google at present, others as they become available.

      v. Links through a dynamic process to type images made available elsewhere on the web. At present a dedicated tool is not available.

d. Links to non-BCA specimen data:

      i. Links to specimens examined lists / database entered by collaborating taxonomist(s), restricted to specimens from Mexico and Central America.

      ii. Links to specimens examined lists / database entered by the collaborating taxonomist(s), but encompassing specimens of the species collected from both within and outside the region.

      iii. Links to type specimen data entered into BCAC by collaborating scientist(s), or direct link to data elsewhere on the web if available.

      iv. Dynamic links, developed in two stages: initially this will be a list of links to collection databases of the major specimen-holding institutions for Mexican and Central American specimens, which can be put in globally. Phase 2 will depend on dynamic searches being possible through GBIF, Species Analyst, REMIB, etc, and will require links to them.

      v. Dynamic links, developed in two stages: initially this will be a list of links to collection databases of the major specimen-holding institutions for species with Mexican and Central American distributions, which can be put in globally. Phase 2 will depend on dynamic searches being possible through GBIF, Species Analyst, REMIB, etc, and will require links to them.

      vi. Dynamic links, developed in two stages: initially this will be a list of links to collection databases of the major type-holding institutions for Mexican and Central American species, which can be put in globally. Phase 2 will depend on dynamic searches being possible through GBIF, Species Analyst, REMIB, etc, and will require links to them.

e. Links to non-BCA Keys:

      i. Links to a key in the BCAC entered by collaborating taxonomist(s) which keys out the species concerned (or other taxon).

      ii. Links to a key in the BCAC entered by collaborating taxonomist(s) which keys out subtaxa of the taxon concerned.

      iii. Dynamic links will be developed in two or possibly three stages: first, links to extant keys on the web in other sites entered by collaborating taxonomist(s). Second, dynamic link to search for keys published elsewhere on the web. The tool to provide this is not currently available. Finally, key(s) produced within BCAC dynamically from description elements entered by collaborating taxonomist(s) (this depends on the development of software and an appropriate editorial policy).

      iv. Dynamic links will be developed in two or possibly three stages: first, links to extant keys on the web in other sites endorsed by collaborating taxonomist(s). Second, dynamic link to search for keys published elsewhere on the web. The tool to provide this is not available. Third, key(s) produced within BCAC dynamically from description elements entered by collaborating taxonomist(s) (depends on the development of software and an appropriate editorial policy).

f. Links to names:

      i. Link to synonymy lists entered into BCAC by collaborating taxonomist(s).

      ii. Link to local names entered into the BCAC by collaborating taxonomist(s).

      iii. Dynamic links to name servers for synonymy lists, expected to be developed in two phases: First, links to on-line databases held elsewhere, entered either by collaborating taxonomist(s) or as global list. Later, links to GBIF, Species 2000, ITIS, and/or IABIN to do a dynamic search of the authority file. Output will be synonymic lists, and possibly alternate classifications.

      iv. Dynamic links to local names via names server, expected to be developed in two phases: first, links to on-line databases held elsewhere, entered either by collaborating taxonomist(s) or as global list. Later, links to GBIF, Species 2000, ITIS, and/or IABIN to do a dynamic search of the authority file.

    g. Links to other data additional to that presented in the EBCA:

      i. Links to discussion entered in the BCAC by collaborating taxonomist(s).

      ii. Links to distribution information or list entered into BCAC by collaborating taxonomist(s).

      iii. Links to BCAC distribution map: Two alternates are possible (and may both be used):

        a. Static map submitted by collaborating scientist(s);

        b. Maps generated dynamically from specimen data and distribution list entered into BCAC by collaborating taxonomist(s) (5.d.i and 5.d.ii, above). Separate maps can be generated from each of 5.d.i and 5.d.ii above, or a combined map could be generated (a choice could be given). The mapping facility on Tropicos can be used as an example (and use of it perhaps through collaboration with Missouri). If there are multiple treatments, maps should be generated separately from each.

      iv. Link to annotations from the BCA volumes collected and entered by collaborating taxonomists and librarians. Annotations should be attributed to author (if known) and volume set. It would be useful to have at least some images of the script for each supposed author.

      v. Bibliography of species entered by collaborating taxonomists and librarians.

      vi. Distribution produced dynamically from the information collected through the operation of 5.d.iv and 5.d.v. Choices might be offered as to the geographical extent of the checklists, in a way similar to that offered through TROPICOS. The software to do this should be part of the BCAC suite, and might be accessed through collaboration with Missouri.

      vii. Distribution map(s) produced dynamically from the information collected through the operation of 5.d.iv and 5.d.v. The source of the mapmaker software, and whether it should be part of the BCAC on-site suite of software or accessed remotely, needs to be considered.

      viii. Links to web pages outside the BCAC with data on the taxon. This would be achieved through a search engine either dedicated to the system or, for example, Google.

      ix. Dynamic links to abstracting services such as Biosis, CABI etc. There might be a subscription cost implication to this.

    h. Links to information on conservation status:

      i. Dynamic link to CITES to establish CITES listing of species.

      ii. Dynamic link to IUCN to establish red list status of species.

      iii. Dynamic link to any extant regional lists of species and their conservation status.

      iv. Dynamic links to any extent national lists of species and their conservation status. A choice would be offered as to the country or countries listed.

6. Other long-term goals include the following items available from a tool bar anywhere within the EBCA and BCAC:

    a. Spanish – English Toggle: a toggle between various parts of EBCA and BCAC as completed; this may be done incrementally as funding and man-power allow; a mechanism is for implementing is needed. Automating the translation feature would be a nice addition.

    b. Index to authors of treatments and taxa (BCA and BCAC) that will include:

      i. a list of authors' names

      ii. where they worked

      iii. where specimens were deposited

      iv. BCA and/or BCAC treatments authored

      v. provision for places to hold other kinds of information available in the reference literature, e.g., TL-2.

     vi. a cross-referenced list of abbreviations used in BCA.

  c. Index to collectors listed in BCA (and BCAC?) that will include:

      i. a list of collectors' names

     ii. where specimens were deposited

    iii. provision to generate lists of localities collected at

    iv. provision to generate lists of taxa collected

     v. provision to generate lists of taxa authored in BCA and/or BCAC.

  d. Links to external indexes of related information as those are made available digitally (e.g., authors in TL-2, Collectors volumes of Index Herbariorum).

  e. Links to good reference maps of countries and regions covered: physical, political, climate, soil, vegetation, etc.

  f. A search tool for finding localities on maps—if or when such a thing becomes available.

  g. Links to gazetteers and gazetteer search tools, such as the AMNH gazetteer mentioned above, the Alexandria Digital Library, NIMA, Getty, etc.

  h. Create regional lists: Provision to generate lists, by region, of taxa. This should include the ability to do lists of families (all or within a particular higher taxon), genera (all or within a particular higher taxon), species (all or within a particular higher taxon). Other criteria may be useful in defining these lists as well.

  i. Create national lists: Provision to generate lists, by country, of taxa. This should include the ability to do lists of families (all or within a particular higher taxon), genera (all or within a particular higher taxon), species (all or within a particular higher taxon). Other criteria may be useful in defining these lists as well.

  j. Bibliography for BCA (i.e., a bibliography to the thousands of works cited within the text of the BCA). This should be generated through the markup of volumes; will need to be edited by area specialists as these are identified and volunteer to do so. To speed up the editing process, the following tool should be made available. Because there are a finite number of citations possible for the journal/book/book series field, some of which can be imported and others can be entered (full names, standardized abbreviations, and variant abbreviations), the system should be set up to so that the EBCA coded text should be validated against the 'standards' and, if there is no match, be flagged as needing attention from reviewers. If flagged, the entry can either be accepted (as a new journal/book/book series for the bibliography) and added to the 'acceptable' list or rejected and corrected (or linked as a non-standard, but used abbreviation.

  k. Bibliography for BCAC (i.e., bibliographies for each BCAC taxonomic treatment). This will be contributed and edited by collaborators.

  l. Links to information about related field notebooks, including listings in library catalogues (especially NHM, London), and a list of and links to digitized field notebooks as those become available. The latter should be indexed by collector, date, locality, and taxon.

  m. Links to information about related correspondence, including listings in library catalogues (especially NHM, London), a list of and links to digitized materials as those become available. The latter should be indexed by author, recipient, and date.

7. Electronic attributed marginalia. As taxonomists have added notes in the margins of books in institutional libraries for centuries, provision to include these as linked, attributed information from throughout BCA and BCAC should be made. Not only to capture the marginal notes in existing printed literature, but also to allow users of the electronic BCA and BCAC (and e-journal) to add such marginalia (properly attributed, and reviewed by the editors before it becomes widely available).

8. Long-term "archival" storage of image and coded text files will be arranged through the OCLC Digital and Preservation Co-op that SIL has joined as a charter member. OCLC is a tested and

trusted third-party, which has been at the forefront in enhancing access to information resources for the last 40 years. As a charter member of the OCLC Digital and Preservation Co-op, SIL, with peer institutions, will be defining the structure, features, and policies of the OCLC Digital Archive, the first version of which is planned to open this spring.

9. BCAK(ids). A possible development is to create specialized site elements for use in education. While this is only in the early discussion phase, several ideas have been suggested and are being investigated. These include:

   a. An interactive site for key construction, using morphological character subsets from the BCA and BCAC. A dynamic 'score' for effectiveness could be generated by appropriate software. There would be links to the appropriate images for optional use.

   b. An interactive mapping system allowing local specimen data to be added as part of, e.g., school projects, associated with local/domestic determinations and surveys. This may be more immediately accessible with large / showy taxa (e.g., many vertebrates, butterflies).

   c. The facility for following collecting trips by contributors to the BCA, based on their field notebooks and specimen data. Images of the sites at the time of the BCA and now could be added to the site.

Key Contributors

Smithsonian Institution Libraries (SIL)
- Tom Garnett, Assistant Director for Digital Library and Information Systems, will manage the EBCA project
- Martin Kalfatovic, Head of SIL's New Media Office, will serve as project manager and Contracting Officer's Technical Representative (COTR) for the scanning and rekeying contracts, as well as supervisor of project personnel
- Sherry Kelley, Assistant Director for Technical Services and Administration will be the COTR for the cataloging clean-up project

Smithsonian Institution, Office of the Chief Information Officer
- Dawn Leaf, Director, Systems Architecture & Product Assurance, will coordinate OCIO review of and participation in the information technology portions of this project

Smithsonian Institution, National Museum of Natural History
- Anna Weitzman, Research and Collections Informatics Manager, serves as chief scientific advisor to the EBCA and coordinate the taxonomic portions of the work of the Taxonomic DTD working group and the Scientific Advisory Committee to the EBCA; also serves as coordinator for scientific planning of the BCAC
- Dennis Hasch, Technical Webmaster
- Scott Miller, Acting Chair, Department of Systematic Biology

Smithsonian Institution
- Leonard Hirsch

Natural History Museum, London
- Christopher Lyal, Leader of the Beetle Diversity and Evolution Programme, serves as external scientific advisor to the EBCA and coordinator for scientific planning of the BCAC
- Sandra Knapp

Missouri Botanical Garden
- Gerrit Davidse

CONABIO
- Jorge Soberón Mainero

Taxonomic DTD working group
- Stan Blum, CAS
- Gerrit Davidse, MO
- Tom Garnett, SIL
- Donald Hobern, GBIF
- Raul Jimenez, CONABIO
- Martin Kalfatovic, SIL
- Chris Lyal, NHM, GTI
- Erick Mata, INBio
- Tom Moritz, AMNH
- Anna Weitzman, NMNH
- Betty Harvey, XML contractor, SI-OCIO

Scientific Advisory Committee
- Eldredge Bermingham, STRI
- Gerrit Davidse, MO
- Michael Dix, Universidad del Valle de Guatemala
- Sandy Knapp, NHM
- Chris Lyal, NHM, GTI
- Jean Michel Maes, Museo Entomologico de Leon, Nicaragua
- Antonio Mijail Perez, Universidad Centroamericana, Nicaragua
- Scott Miller, NMNH
- Jorge Soberon, CONABIO
- Jesus Ugalde, INBio
- Anna Weitzman, NMNH

Appendix 1. Draft of summary page for BCAC taxa entries and links.

Taxon name

	Descriptions	Images	Specimens	Keys	Names	Other data	Conservation Status
BCA	*BCA original* (EBCA)	*BCA image*	BCA (EBCA)	BCA – to taxon (EBCA); BCA – within taxon (EBCA)	BCA (EBCA)	BCA discussion (EBCA)	CITES
BCAC (manually entered)	*BCAC*; *Other descriptions*; Original description / Protologue	BCAC images; Other images; Type image(s)	From Mexico and Central America; All specimens; Type specimens(s)	BCAC – to taxon; BCAC – within taxon	BCAC synonyms; Local names	BCAC discussion; BCAC distribution; BCAC mapmaker; Annotations; Species bibliography	IUCN; Regional lists; National lists
BCAC (dynamic)	Other descriptions; Original description / Protologue	Images; Google; Type image(s)	from Mexico and Central America; All specimens; Type specimens(s)	Others – to taxon; Others – within taxon	Name server; Local names	distribution; mapmaker; Other web pages; Abstracts	

APPENDIX 2: DRAFT Conceptual Model for *Biologia Centrali-Americana Centennial* Project (BCAC)

Name Server

ITIS, GBIF ECAT, Species 2000 and/or others or internal temporarily or external if/when possible

- Name (incl. valid names and synonyms; subdivided into name (bi- & tri-nomials split) and author string)
- Nomenclatural comments
- Colloquial names (many to many)
- Alternative Classifications (many to many)

Name Server provides list of names and synonyms back to BCA web interface (incl. ability to choose classification)

User selection of names to search from BCA site

Selected names allow searching of other sources

Post- and Pre-BCA taxonomic treatments

- Other descriptions of same species (INBio, Flora MesoAmericana, other treatments, etc.)
- Images from treatments
- Species Web Pages (various)
- Specimen databases incl. images
- Collective specimen data resources (e.g., REMIB, Species Analyst, DiGIR)
- Other sources of taxonomic names and synonyms

Ability to describe contents of resource with link to it if available

- Spanish Language Descriptions

BCA taxon, key, text section entry links to

BCA Taxon name links to

Multiple identifications may be used to link to

Other sources of images, e.g. Google

Conservation status of taxon (e.g., CITES)

Direct link from "preferred" taxon name to other treatment

Specimen data collected and passed back to BCA site

Specimens and type specimens and images with direct links to their home collections

Direct links from specimens examined

Link from key, taxon name to taxon record

Specimen data collected and passed back to BCA site

- Mapping of specimen data by taxon and through time (precisely those specimens listed in BCA in one and including others through time in others)

Combine species names and distribution information

Links from geographic data

Geographic data sources

- Alexandria Digital Library and/or Getty and/or ?
- AMNH BCA Gazetteer update and georeferenced
- Place names and synonyms
- Coordinate point/polygon data
-

BCA Central Data components

- Taxon Name (for all levels–family, genus, species, etc.)
- Author
- Synonyms/Authors
- Publication
- Key
- Description [text only, no links]
- Discussion (incl. Notes, Ecology, Distribution) [text only, no links]
- Specimens Examined
- Illustration (link from citation to actual)
- relative taxonomic position (probably derived from relative position in text)

Species information from same region from external sources (e.g., species described later or not recognized in BCA)

BCA Centennial (Other long-term goals)

- Checklists of taxa from regions and subregions within MesoAmerica
- Interactive key generation (assumes morphological data from somewhere)
-

Link by Author Name

Correspondence

link to relative position in text, possibly multiple

Attributed annotations in various copies

Link by Taxon and Collector (also date?)

Link to bibliographic record (TBD)

Bibliographic Record (EMu, SIRIS, other library systems)

Field notebooks

Other needs:

- Provision of ability for other websites to link to BCA
- Ability to take on-line suggestions for update and resources to include in BCA (and send them to the appropriate contact?)
-

16