White Paper
DEMCA: Documenting the Ethnobiology of Mexico and Central America:
A Digital Portal for Collaborative Research
Jonathan D. Amith (PI)
2020-09-30

0. Overview

This project is part of a long-term, ongoing effort to develop an encyclopedic, multimedia study of the nomenclature, classification, and use of flora and fauna among Indigenous communities in Mexico. There are two aspects to this project. The first is substantive. Over the past decade Amith has researched traditional ecological knowledge of flora and fauna in several central Balsas (state of Guerrero) Nahuatl communities, in four Yoloxóchitl Mixtec (coastal Pacific) communities, and in Nahuat[l][1]- and Totonac-speaking villages in northern Puebla. This has led to the creation of a lot of material relevant to ethnobiological research and the fields of lingusitics, anthropology, and biology. At a very basic level, Amith's studies provide data on local traditional ecological knowledge across a range of Mesoamerican languages, cultures, and communities. But this data has implications that reach far beyond the study of any single locality. Viewed collectively, this data opens a window into the cultural and linguistic history of Indigenous communities: (1) the paleo-ecology and location of ancestral homelands based on an analysis of protoforms from cognate sets; (2) the detection of linguistic and cultural borrowings (loans, calques, uses, and beliefs) between native groups in movement; (3) insight into patterns of migration and contact; (4) the history and nature of linguistic and cultural divergence; and (5) those factors that most influence the retention or loss of words that reference local flora and fauna.

The second focuses on the creation of a an online portal (DEMCA: Documenting the Ethnobiology of Mexico and Central America; demca.mesolex.org; built with two NEH Office of Digital Humanities grants). This portal will (a) manage and analyze ethnobiological data; (b) offer multifaceted mechanisms for data discovery and display; (c) facilitate direct participation of Indigenous scholars and communities in grass-roots documentation of local knowledge of natural history; and (d) provide a wide range of stakeholders (researchers, Indigenous communities and scholars, the general public) with open access to comparative (ethno)biological material (including downloadable databases, photo documentation, sound files and transcriptions/translations).

The present white paper reports on the structure and content of the DEMCA portal and offers the general public a guide to the types of botanical (over 12,000 collections and 22,000 in situ photos of plants) and ethnobotanical (plant names and uses in Indigenous communities and audio recordings with line-by line playback while the cursor follows the transcription and translation of this material).

1. BENEFITS OF BASING DEMCA ON THE SYMBIOTA CONTENT MANAGEMENT SYSTEM

The DEMCA portal is based on innovation and adaptation of Symbiota software (see https://symbiota.org/docs/). As the Symbiota homepage states, this software is focused on creating an online resource that is most heavily focused on providing tools for the study of biodiversity:

> Symbiota is an open source *content management system* for curating specimen- and observation-based biodiversity data. It is built on the premise that a collaborative partnership of biodiversity informaticians, collection managers, and biodiversity research communities will be most effective in creating high quality biodiversity research resources with publicly useful portals.

Given its goals, Symbiota is an extremely powerful tool for managing biological (particularly specimen and observation) data. It is used by many collections (see https://symbiota.org/docs/symbiota-introduction/active-symbiota-projects/) for online presentation of their holdings including the very large and active Smithsonian Tropical Research Institute online database, a Symbiota-based portal at https://stricollections.org/portal/.

[1] The term Nahuat[l] with the final consonant in brackets is used because the area studied includes communities that lack the <tl> phoneme (e.g., *tata*, 'it burns') and those that do not (e.g., *tlatla*, 'it burns').

The choice of Symbiota software for DEMCA's content management system (CMS) has yielded immensely positive results in a very short time. Functionalities such as advanced specimen and image search and display, including mapping; generation of lists of documented taxa; flexibility in incorporating multiple projects (such as distinct community studies) and allowing for searches across any group of selected projects; restricting access to any subset of material (e.g,. the geolocation of protected species), were all included in the Symbiota software package. Moreover, many of these functionalites can easily be utilized in the context of new ethnobiological data. For example, to avoid exploitation and commercialization of protected species (such as orchids), Symbiota has quite an advanced system to limit access to sensitive data, in this case the coordinates of documented collections of protected species. Ethnobiological data can also be sensitive, for example in the case of data on medicinal use of plants, knowledge to which communities and individuals might want to restrict access. Symbiota software has a mechanism to restrict access based on taxa (for example, hiding the coordinates of protected species); DEMCA will adapt this functionality to ethnobiology my hiding the description of use for all plants tagged as medicinal.

Symbiota also has an advanced system for localizing and images and creating an output (at present a dymically generated webpage) based on a search or browse. The image search functionality of DEMCA is illustrated in figures 1 and 2. Figure 1a is the search interface. Presently it generates (figure 1b) a portfolio of all portal photos that meet the search criteria. This output represents another DEMCA-Symbiota functionality that could be adapted to community-based and ethnobiological initiatives by facilitating the generation of extensively illustrated species "plaques". The plaque is in effect an assembly of all photos from all collection and observation events for *C. icosandra* (a total of 19 collections). Clicking on any image remits the user to the collection details related to the field photo clicked (fig. 1c). As can be seen from this set of images, the DEMCA-Symbiota portal includes functions that are extremely relevant to creating resources that could be highly beneficial to a range of stakeholds. The content management exists; the step that is lacking is generation of pdf files that capture the search results in a permanent and printable format.

In addition to the "image search" interface (fig. 1a) there is also an image browse interface (fig. 2a). The browse interface offers the user the ability to browse by family, genus, or species and then select a species. Instead of a "plaque" of images of that species, the user is remitted to a Taxon page for the species (fig. 2b). This Taxon page will have multiple tabs: Description, Vernacular names, Vernacular uses, Multimedia (which links to audio playback about this species), Plant name summary (summary of all Indigenous names documented for this species), Plant name analysis (morphological analyses and glosses for all Indigenous plant names), and Calques/loans (which presents a list and analysis of plant names that are calques or loans of names in a distinct language). The first three tabs contain information that is very relevant to many stakeholders. Again, as with the Search image function, the Brown images offers users a way to view a cohesive set of information related to any particular species. At present this functionality is limited to online presentation. But a downloadable and printable pdf would not be difficult to implement. This provides another example of how the DEMCA-Symbiota CMS can fairly easily be adapted to providing printable resources for off-line (and community) use.

Figures 1a, 1b, and 1c: Search input and results for *Conostegia icosandra* as a function of DEMCA-Symbiota

Figure 1b: Search for *Conostegia icosandra*

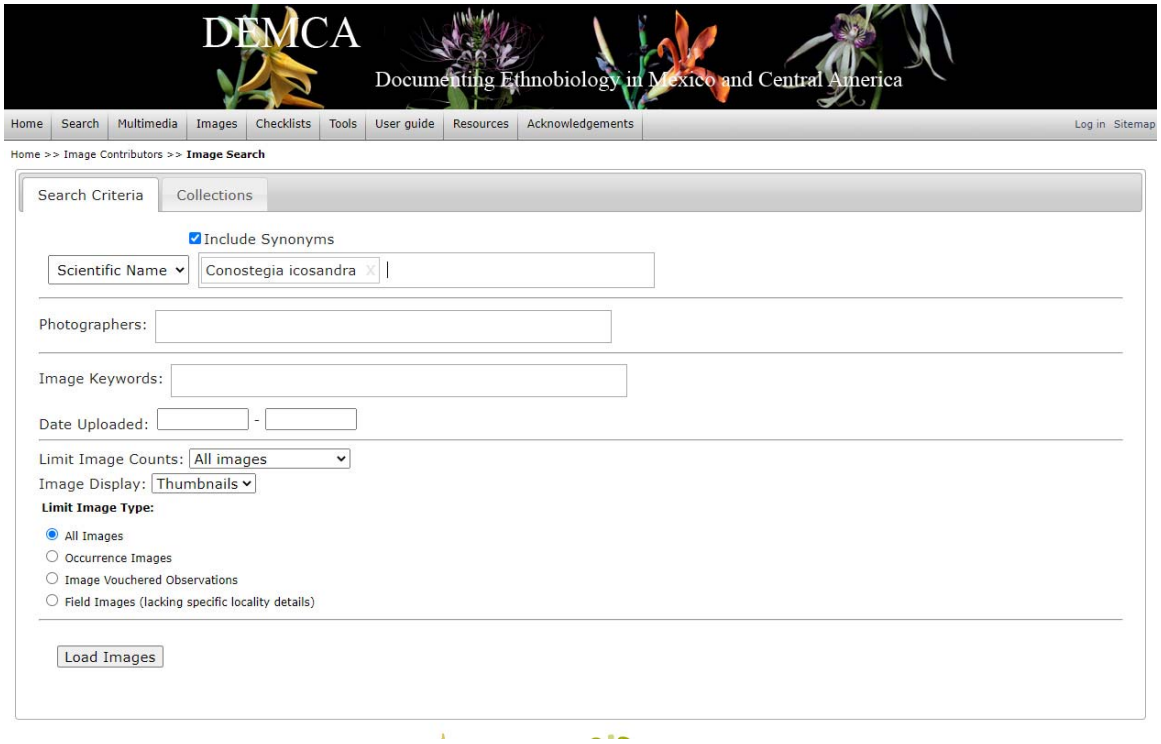
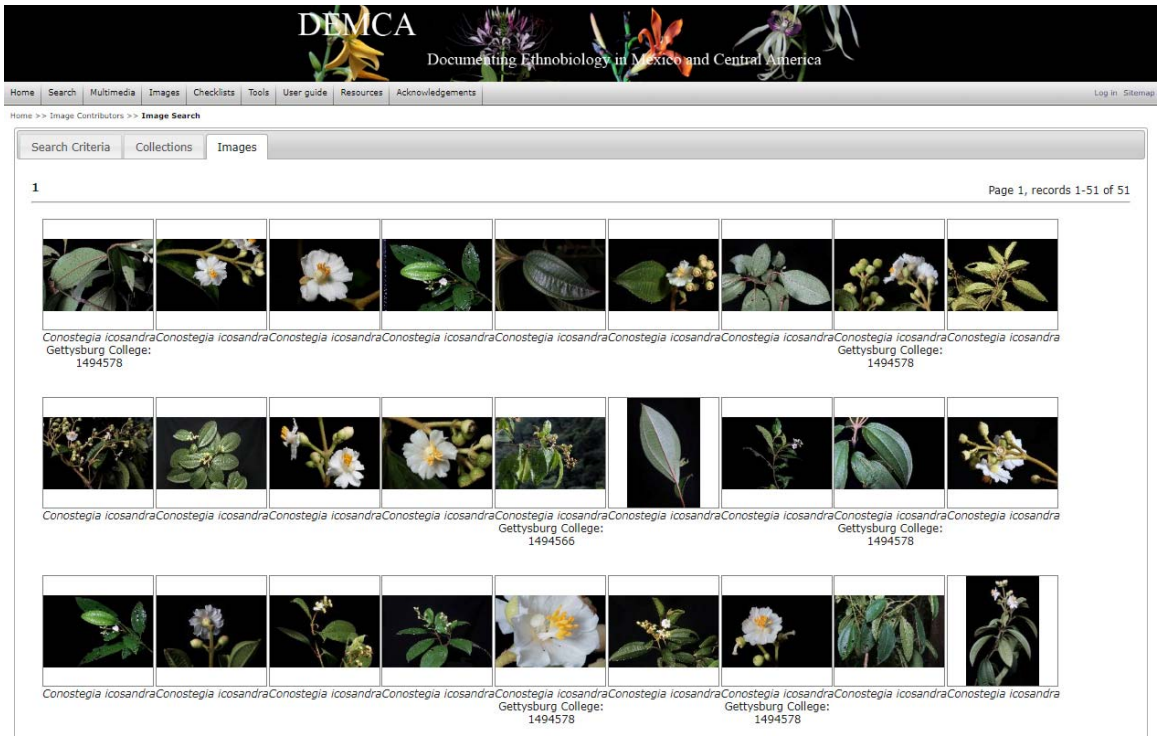


Figure 1b: Search results for *Conostegia icosandra*: 51 photos from 19 collection events



1c Links to collection basic data

Details | Map | Comments | Linked Resources

Share 0 | Tweet

Gettysburg College

Floristic and ethnobotanical study of the Nahuat community of Xinachapan and nearly villages (munic. Huitzilan), Sierra Nororiental de Puebla

Taxon: *Conostegia icosandra* (Sw. in Wikstr.) Urban
Family: Melastomataceae
Determiner: F. Almeda (2019-04-29)
Collector: Jonathan Amith y Hermelindo Salazar 80219
Date: 2016-08-03
Verbatim Date: 2016-08-03
Locality: México, Puebla, Huitzilan de Serdán, PUEBLO: Xinacachapan; LOCALIDAD EXACTA: A 1 hr de camino de Xinacachapan, hacia XAltipac, hacia abajo, en el último cafetal hacia el río Atenoh.
19.96835 -97.67239
Elevation: 813 meters
Verbatim Elevation: 813 m
Habitat: VEGETACIÓN y HÁBITAT: Vegetación Secundaria Bosque perennifolio con apariencia de no haber sido perturbado en mucho tiempo.

Specimen Images

Open Large Image

Open Large Image

Open Large Image

Open Large Image

Open Large Image

Open Large Image

Open Large Image

Open Large Image

Usage Rights: CC BY-NC (Attribution-Non-Commercial)
Record Id: 3346c6d2-c728-407e-a1b1-c2932977e054
Occurrence ID (GUID): 3346c6d2-c728-407e-a1b1-c2932977e054
For additional information on this occurrence, please contact: Jonathan D. Amith (demca.biology@gmail.com)

Figures 2a and 2b: Browse functionalities and links to Taxon pages

Figure 2a: Browse features for species starting with "R"

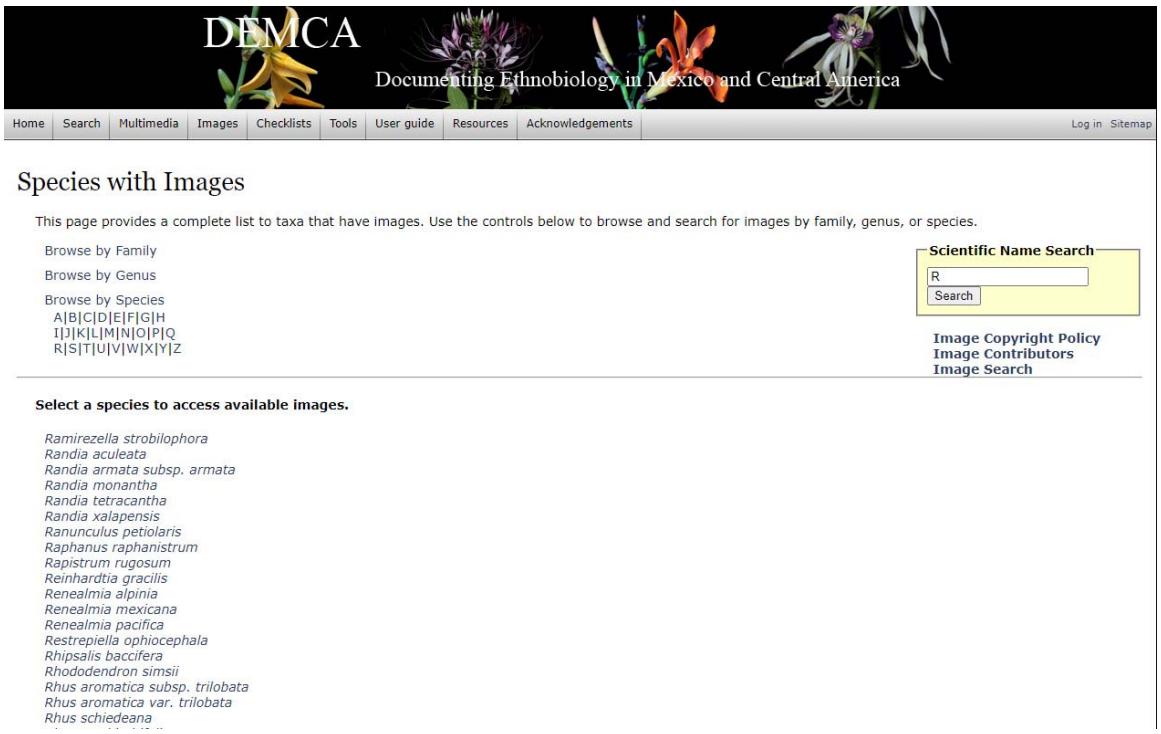
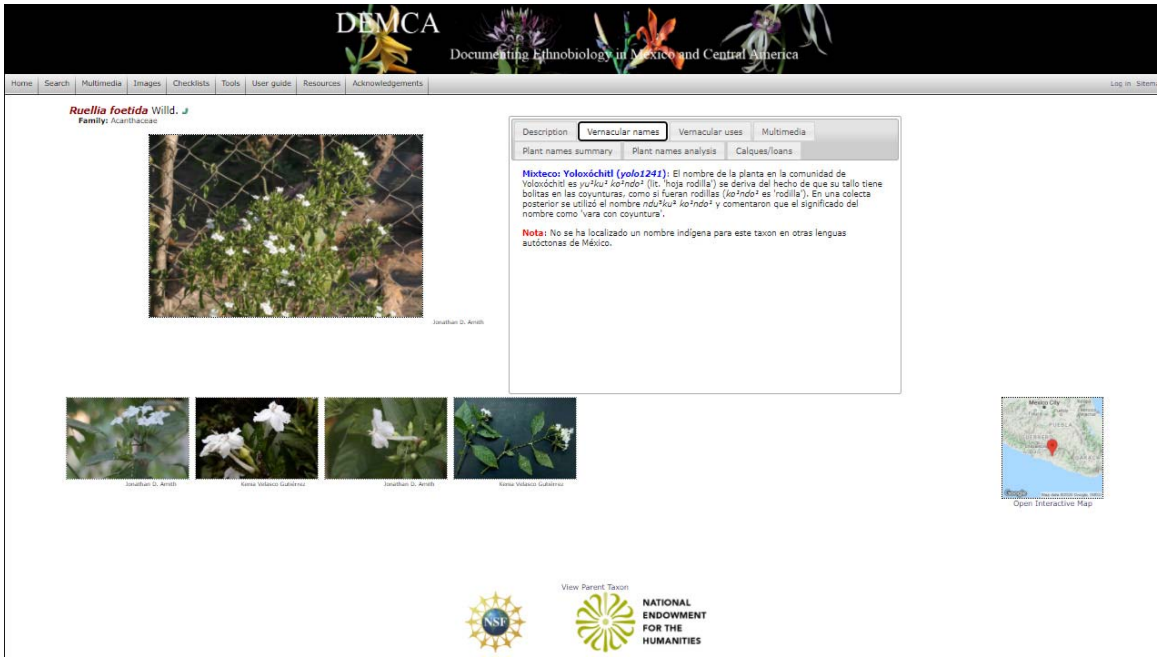


Figure 2b: Remision from species on browse list (in this case, *Ruellia foetida*) to Taxon page



Figures 3a, 3b, and 3c: Map search and display followed by link to collection record with in situ photographs

Figure 3a: Map search interface note the possibility of limiting searches to "specimens with ethnobiological data" or "specimens with multimedia (audio) files. Note also (visible at the lower left) the very top line of a interface to limit searches to plants with specific criteria (e.g., all *Piper* that are used in living fencing)

Fig. 3b: Search results. Each individual collection is marked by a clickable orange dot that remits the user to details of the collection. The search not only maps the results but generates a list of all records that meet the search criteria (middle tab to left: Records) and produces an inventory of all *Piper* species in the projects searched (29 species of *Piper* in the present case).

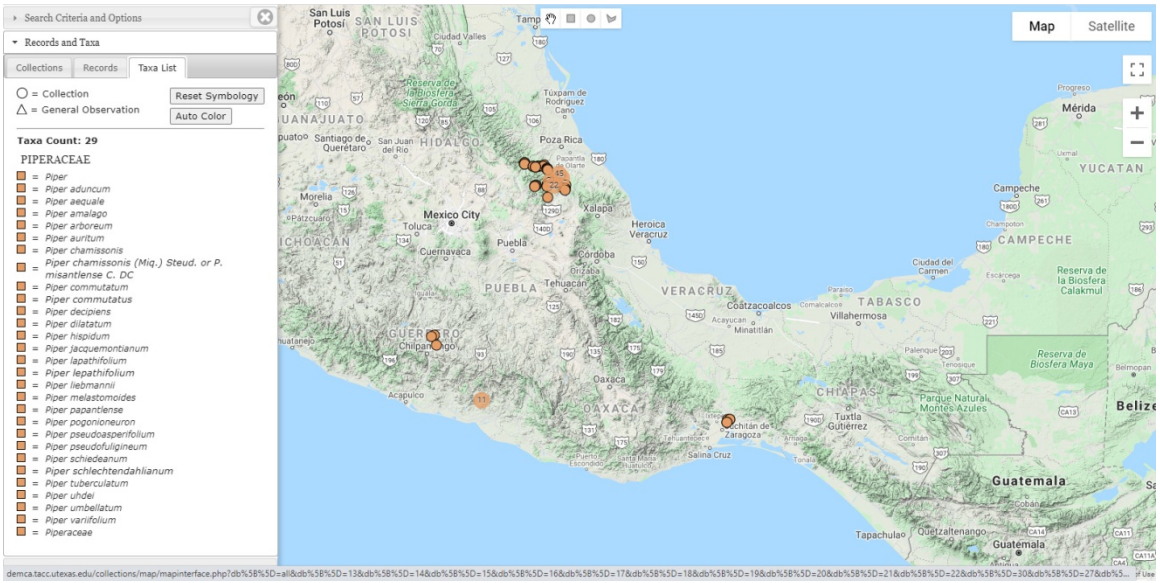
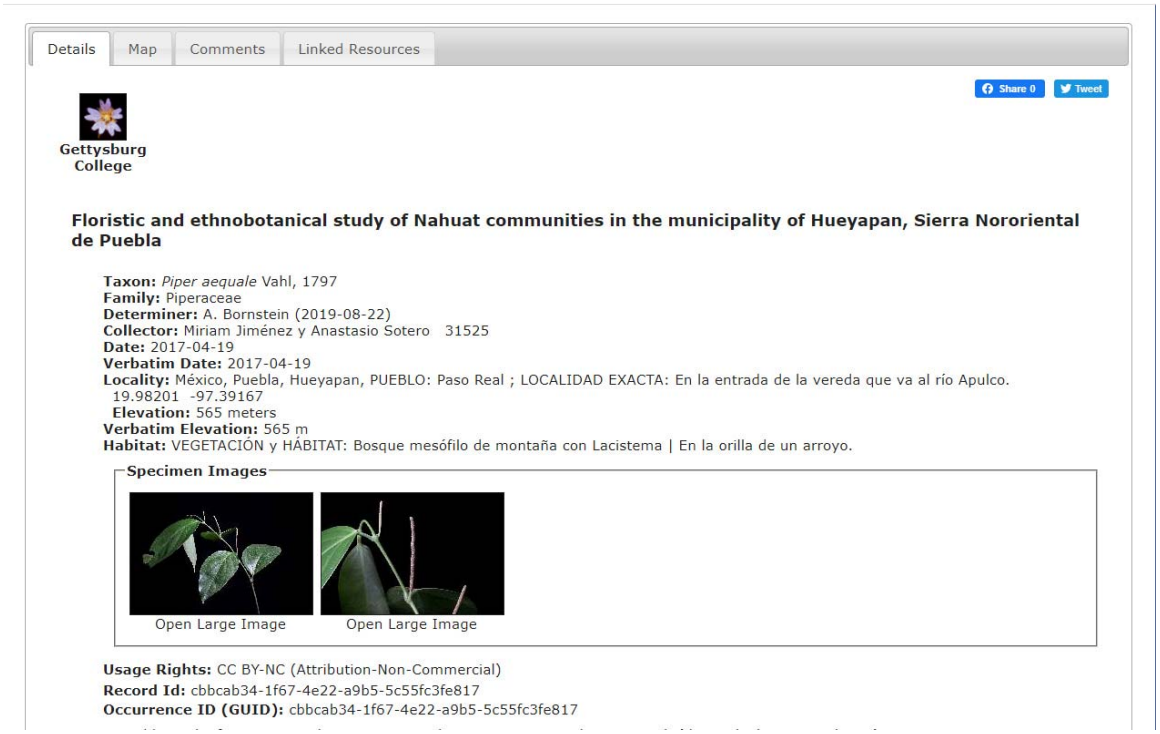


Figure 3c:



Another very useful tool in DEMCA-Symbiota is the mapping function. Symbiota itself has a highly developed system that allows for generating geographic restrictions on searches. For example, a researcher interested in the occurrences of a particular species above 1,000 meters could set this parameter in a filter. Much more basic, however, is a map display in which the georeferences from collection records are used to map the location of each collection event. Figures 3a, 3b, and 3c demonstrate the way in which this presently works. A user would access the Map search from the DEMCA home page (demca.mesolex.org), click on Open to access the search interface presented in figure 3a.

Similar to the relevance of the Image Search and Image Browse functions for generating online or downloadable resources relevant to multiple stakeholders (particularly the general public and Indigenous communities interested in high quality photographic documentation), the built-in mapping functions of Symbiota will be of great benefit to cultural and linguistic studies. For example, a common interest in historical linguistics are cognate sets that are used to create isoglosses: mapped presentations of cognates. DEMCA-Symbiota will both contain the relevant data and the software mechanism to map cogates via isoglosses.

The preceding discussion presents some of the advantages of implementing a content management system (Symbiota) that is open source, robust, and designed for the management, discovery, and sharing of biological data: Many of its functional tools—image search and browse, mapping—can be adapted and extended to accommodate the needs of ethnobiological research and education. A more basic advantage is the fact that it's entire raison d'être is precisely, as mentioned, the management, discovery, and sharing of biological data. The following section looks at this and the ease with which Amith has been able to ingest extensive botanical and ethnobotanical data from his own and other's research into the portal. The third section then looks at how new software was developed to enhance a modified Symbiota system (i.e., DEMCA, or DEMCA-Symbiota) so that it would be able to integrate ethnobiological data from the humanities and social sciences.

2. DEMCA-SYMBIOTA PROJECT CREATION AND DATA INTEGRATION

2.a Projects and data online.

Table 1 summarizes the contents of 14 projects that have been created in the DEMCA portal. Not included in this list, which represents only Amith's collections, is an Isthmus Zapotec project directed by Gabriela Pérez-Báez. All

Table 1.a Collections and photos from Amith's projects
https://demca.mesolex.org/portal/collections/index.php

Collection name	Communities/municipalities in collection		Total collections	Total in situ photos
Ayotoxco (Nahuat)	Village of Cuauhtemoc (and others)		638	2,089
Balsas Valley Nahuatl	Village of San Agustín Oapan Village of Ameyaltepec Village of San Juan Tetelcingo Village of San Miguel Tecuiciapan		1,171	0
Cuetzalan del Progreso (Nahuat; 95 percent of collections are in the municipality of Cuetzalan)	Villages in mpio. Cuetzalan del Progreso Villages in mpio. Jonotla Villages in mpio. Tuzamapan Villages in mpio. Nautzontla Villages in mpio. Tetela de Ocampo Villages in mpio. Xochitlán		2,088	3821
Ecatlán (Totonac)	Ecatlán (municipality of Jonotla)		246	288
Ferns in the Sierra Nororiental de Puebla (thematic study)	Various villages in the municipalities of the Sierra Nororiental		1,222	accessible through searches across all collections
Hueyapan (Nahuat)	Village of Nexpan (and a few others)		247	839
Huitzilan (Nahuat)	Village of Xinacachapan (and others)		656	2,130
Nanacatlán (Totonac)	Village of Nanacatlán		229	0
San Luis Acatlán (Mixtec)	Village of Yoloxóchitl Village of Cuanacaxtitlan Village of Buena Vista Village of Arroyo Cumiapa		1,428	3,724
Tepetzintla (Nahuatl and Totonac)	Village of Tepetzintla (Nahuatl) Village of Tlaquimpa (Nahuatl) Village of Tonalixco (Totonac)		242	779
Tlatlauquitepec (Nahuat)	Various villages in mpio. Tlatlauquitepec Various villages in mpio. Yaonahuac		103	287
Zacapoaxtla (Nahuat)	Various villages (mostly San Juan Tahitic) in the municipality of Zacapoaxtla)		1,433	3,919
Zautla (Nahuat)	Various villages in mpio. Zautla Various villages in mpio. Xochiapulco		764	1,872
Zongozotla (Totonac: 5 municipalities)	Zongozotla Cuauhtempan Zoquiapan	Atlequizayán Xochiapulco	1,773	4,787
TOTALS			11,589*	**22,527+**

* Reflects total 12,240 minus 651 ferns, which are listed in the both the fern thematic collection and in village-based collections

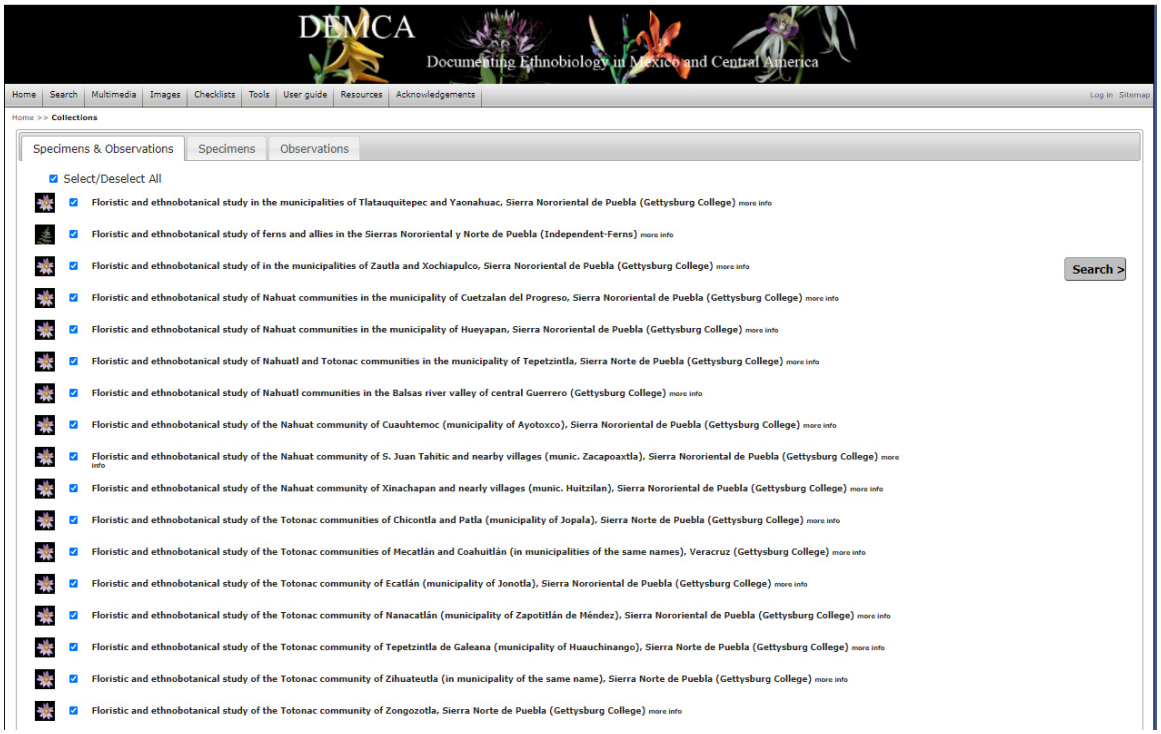
Table 1.b: Additional collections and photos from a comparative Totonac project with David Beck
(808 total collections)

Chicontla and Patla, municipality of Jopala	276 collections	being selected and edited
Zihuateutla, municipality of Zihuateutla	87 collections	being selected and edited

Tepetzintla, municipality of Huachinango	120 collections	being selected and edited
Mecatlán and Coahuitlán (municipalities of the same name)	325 collections	being selected and edited

12,397 plant collection metadata has been incorporated into the portal along with 22,527 in situ photos. Once vouchers have been mounted and processed at the Mexican national herbarium high resolution photos of the mounted specimens will be given to Amith and added to the DEMCA portal.

Figure 4: Screenshot of 17 of 19 collections in the DEMCA portal. Searches can be by individual project or groups of projects



The modular nature of DEMCA-Symbiota has important implications for grassroots, community involvement in documenting local ethnobiological knowledge. Such community-based initiatives, hopefully led by Indigenous scholars, is a middle ground between the usual institutional basis for these types of projects and the more diffuse citizen science efforts that are often "one-offs" with no integrated research or documentary goals.

Although this report will not go into detail on project creation, a few observations are important. First (figure 5), the "super-administrator" of the DEMCA portal can assign rights to individuals to create, populate, and edit a project. Figure 5 illustrates a project creation homepage where the project administrator names and describes the project, acknowledges support, and establishes rights to use. Figure 6 is how the project appears to a user.

Those who have been given permission to administer or edit project contents have active permissions and they may access the project control panel (figure 7) to add to, delete, or edit contents in any part of the project. As figure 7 illustrates, moreover, is that the administrative/editing project dashboard includes both functionalities that are integral to Symbiota and new content management tools that were specifically developed for ethnobiological projects and data. This new set of tools is highlighted in green overlay in figure 7.

2.b Symbiota-based project management tools: Administrative Control Panel and Data Editor Control Panel

The page in figure 7 offers three control panels, each of which will be briefly discussed in how it relates to DEMCA and the incorporation of ethnobiological data in this new portal. This section, 2.b., presents the control panels that are directly supported in the extant Symbiota

Data Editor Control Panel Administrative Control Panel Ethnobiological Control Panel (unique to DEMCA)

Figure 5: Project creation template

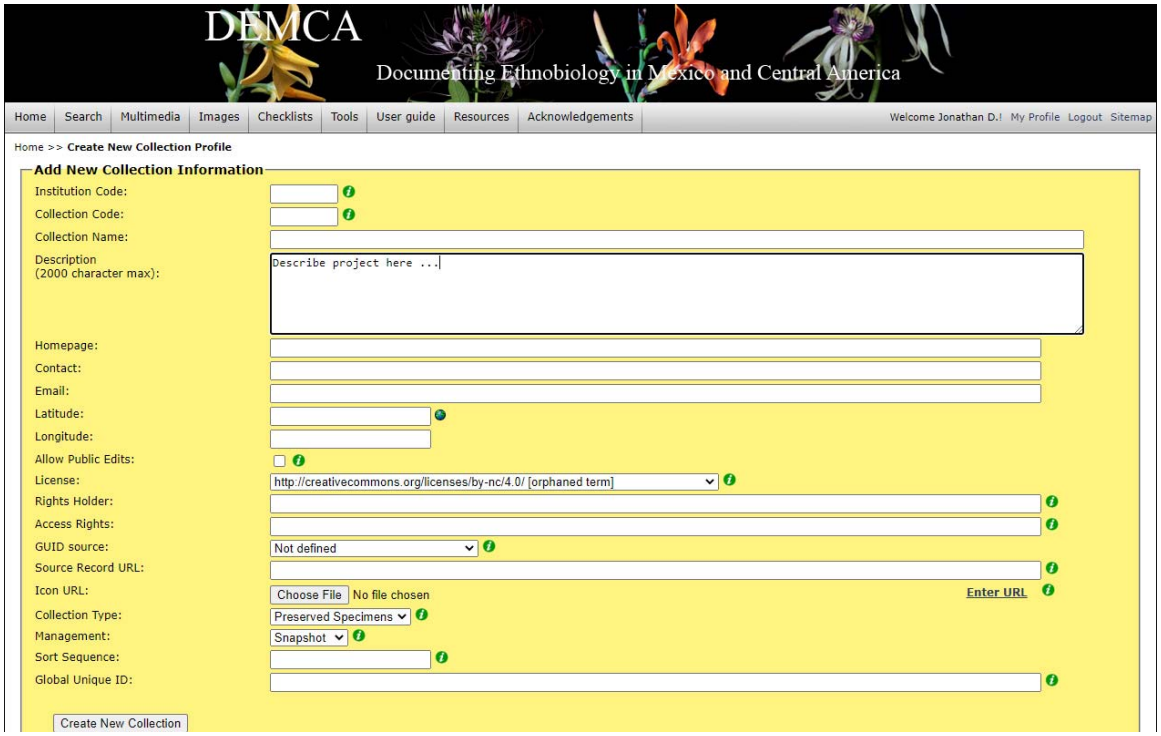
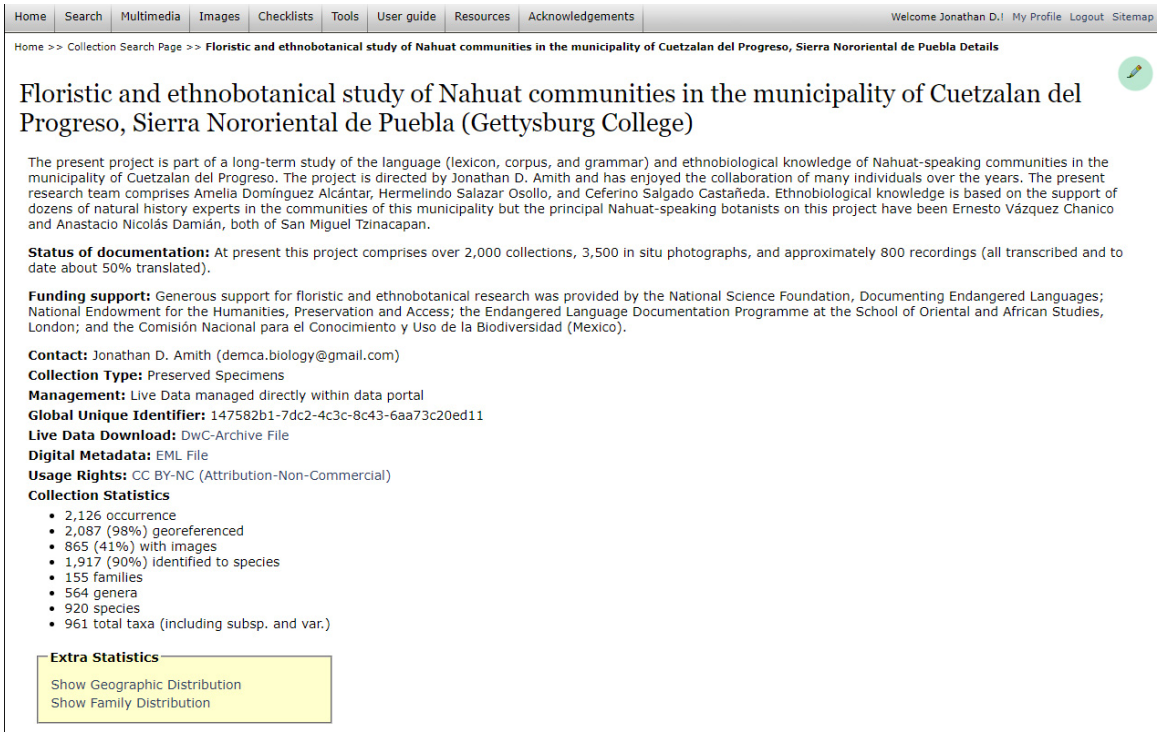


Figure 6: Project home page as it appears to the public (pencil icon in green highlight allows editing of content)



The **Administrative Control Panel** provides general project oversight for all those granted administrative and editing privileges. It allows these individuals to themselves control privileges. It also facilitates back-up of the project via a "Download Backup Data File". But most importantly it has what is called a Processing Toolbox (5th item in the Administrative Control Panel) that offers (1) batch processing of images linked to collection or observation events, (2) crowd-sourcing, and (3) exporting.

Amith has modified the batch processing of images to facilitate the uploading of a csv file that contains the necessary full path to photos stored on any server and collection/observation record. It also includes a keyword description of the image and photographer credits. It allowed Amith to link in a matter of minutes over 24,000 images to his collections.

Crowd-sourcing is particularly useful for community-based initiatives and will be available when Amith secures funding to support community-based initiatives for ethnobiological documentation. Most importantly, it allows the project management team to review photos and data before incorporating them into project metadata, particularly as Observations. As the crowd-sourcing panel (figure 9) illustrates, the project administrator authorizes "volunteers (data providers) and editors to review submissions for processing. A potential enhancement in future development of DEMCA will be to create a mechanism for the uploading sound files (e.g., the native name of the plant documented via photo) and ethnographic notes that will be important initial documentation of plant nomenclature, classification, and use. In essence, DEMCA crowd-sourcing merges grass-roots community-based projects and centralized administrative control over content with the ad hoc mechanism of citizen science.

The final Toolbox mechanism is a user interface for potentially selective and filtered export of project recordings into a csv file. Figure 10 below shows some of the options that this functionality offers in terms of filters and logical operators. It is different from the lists generated by the Search Collections function and the total data back-up that is possible through the Administrative Control Panel. Export to csv or tab-delimited format. This would also be useful for community-based initiatives as Excel file checklists could be created to accompany the illustrated documentation generated by the Image Search and Browse functions already described

Figure 7: Project control panel

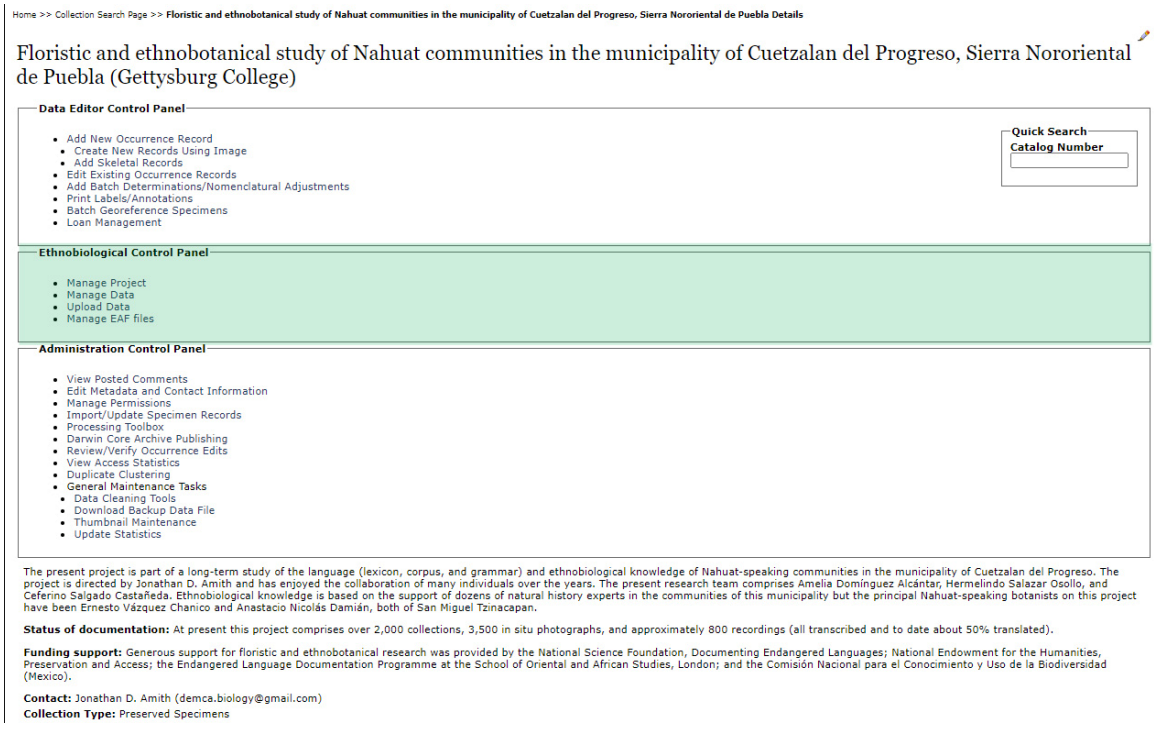


Figure 8: Processing Toolbox: Batch Processing Image Upload, Project Dependent Crowdsourcing, and Exporter

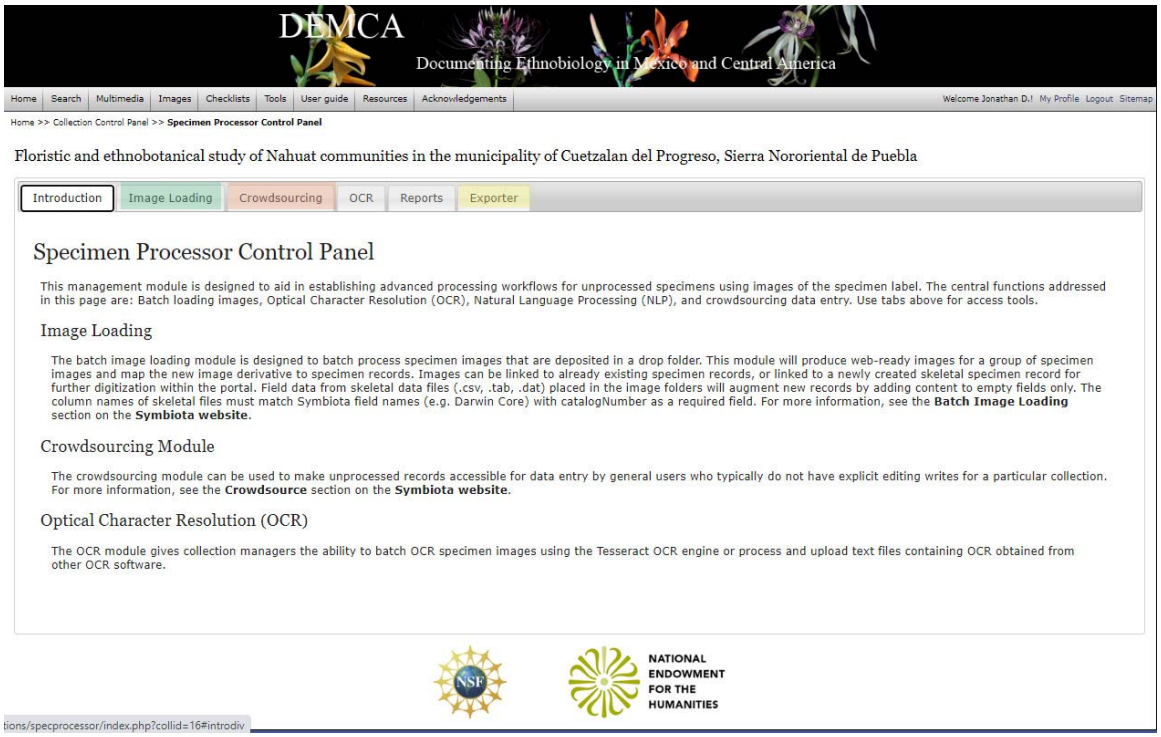


Figure 9: Crowd-sourcing module

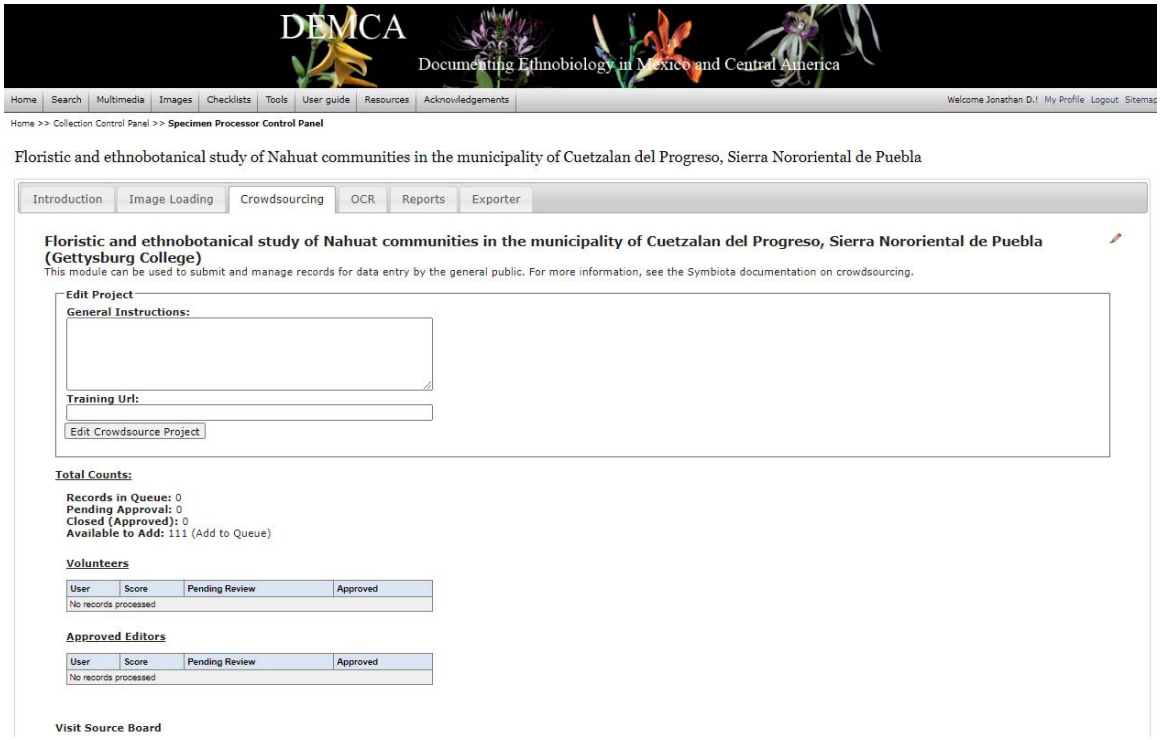


Figure 10: Filtered export of DEMCA collections to csv or tab-delimited databases

In sum, the two Symbiota-based control panels offer a range of functionalities that are extremely valuable content management tools for biological data. As the previous pages make clear, they not only constitute powerful back- and front-end tools that have been developed over many years of software programming but offer functionalities that, as the first pages of this report make clear, can be utilized in ethnobiological and community-based projects.

However, it is also the case the Symbiota and the Darwin-Core metabase structure that it utilizes is deficient in its ability to treat (manage and ingest) and offer a mechanism for discoverability and display of ethnobiological material. This limitation reflects the historical development of both systems: Darwin Core metadata standards and controlled vocabularies and the Symbiota CMS were developed to faciliate information flow among institutions with preserved specimens. Simple data such as vernacular name and language were part of the metadata structure and CMS but in a basically ad hoc manner that was very distinct from how a linguist or anthropologist would handle the same data or in a way that made it easy to retrieve and understand. The following section looks at the limitations of Darwin Core metadata standards/controlled vocabulary, limitations in Symbiota, and the solutions that this project developed.

3. DEMCA INNOVATIONS: ETHNOBIOLOGICAL PROJECT MANAGER DASHBOARDS

The previous sections documented the manner in which the functionalities of Symbiota serve the objectives of the DEMCA portal, for example in facilitating grass-roots and community based floristic and ethnobotanical initiatives. The remainder of this report will focus on the innovations that have been created in the DEMCA portal, innovations that are specifically designed to facilitate ethnobiological research.

The first major innovation has been the creation of a Ethnobiological Control Panel with four major dashboards:

a) Manage Project (§ 3.a)
b) Manage Data (§ 3.c)
c) Upload Data (§ 3.b)
d) Manage EAF files (presented in §5)

Each one will be dealt with in turn.

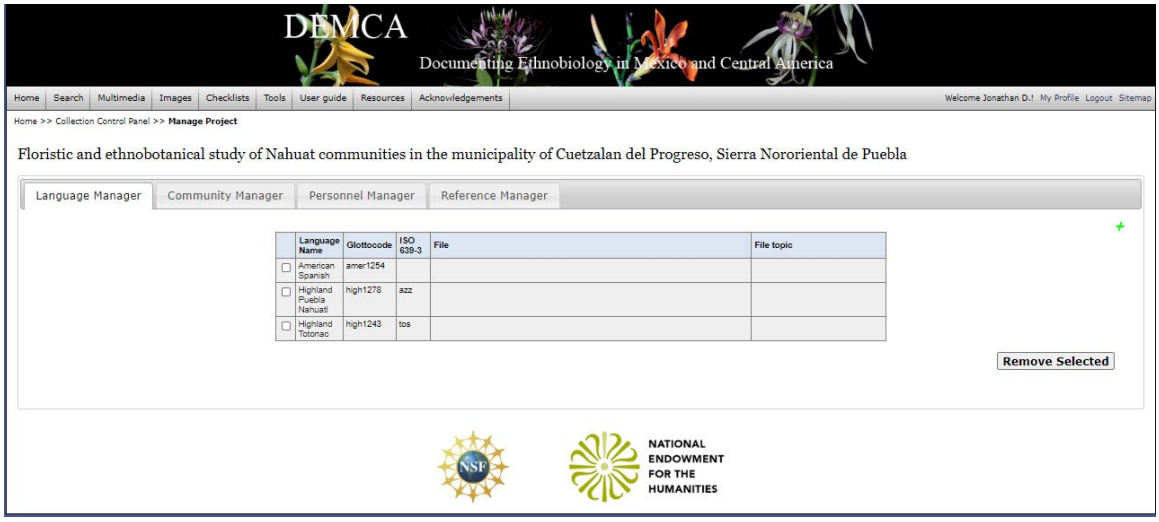
3.a Ethnobiological Control Panel: Manage Project Dashboard

The Manage Project toolset (figure 11), the first line on the Ethnobiological Control Panel, is the heart of the innovations in DEMCA to create a data portal capable of managing the varied types of data that such projects produce. First, however, it is important to present the four sets of data (figure 11) that are managed through the Manage Project dashboard

i. Language
ii. Community
iii. Personnel
iv. Reference

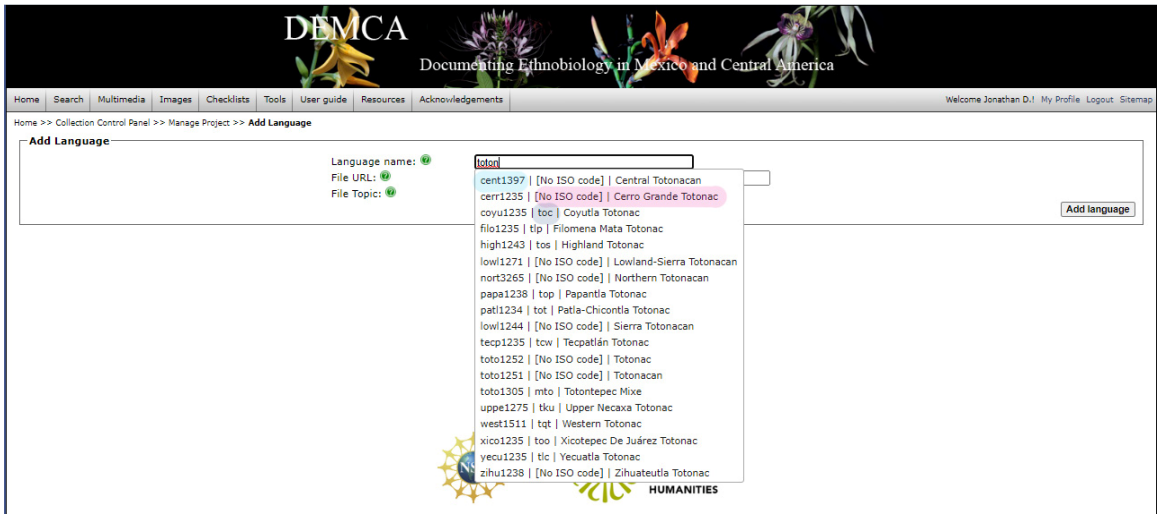
Management of the first three datasets ensures a standardization of data throughout the project. For example, if a consultant is named Osbel López Francisco, his name will always be spelled the same, his demographic data (birth year, village of origin, languages spoken, project role) will be consistent. Before any data is uploaded to the project data for language, community, personnel and references must be incorporated into the project database. All data uploaded at any time in the future must reflect the content of this data. If not, the administrator will be asked to correct the record. For example, If the consultant Osbel López Francisco (a Totonac speaker) is in the project database then an attempt to upload Osbel Lopez Francisco (without an accent in López) will activate a message: "Consultant Osbel Lopez Francisc is not in the project personnel database.". Or, if some time in the future one wants to associate a plant name for O. López with Nahuatl, this would be rejected unless Nahuat is added to the languages relevant to O. López.

Figure 11: Manage Project dashboard: Language, Community, Personnel, and Reference Managers



3.a.i Language Manager Figure 12 is the Language Manager. It is best to start a project database with Language as all communities and all consultants need to be associated with at least one language, and if this language is not already in the project dataset, it would need to be added.

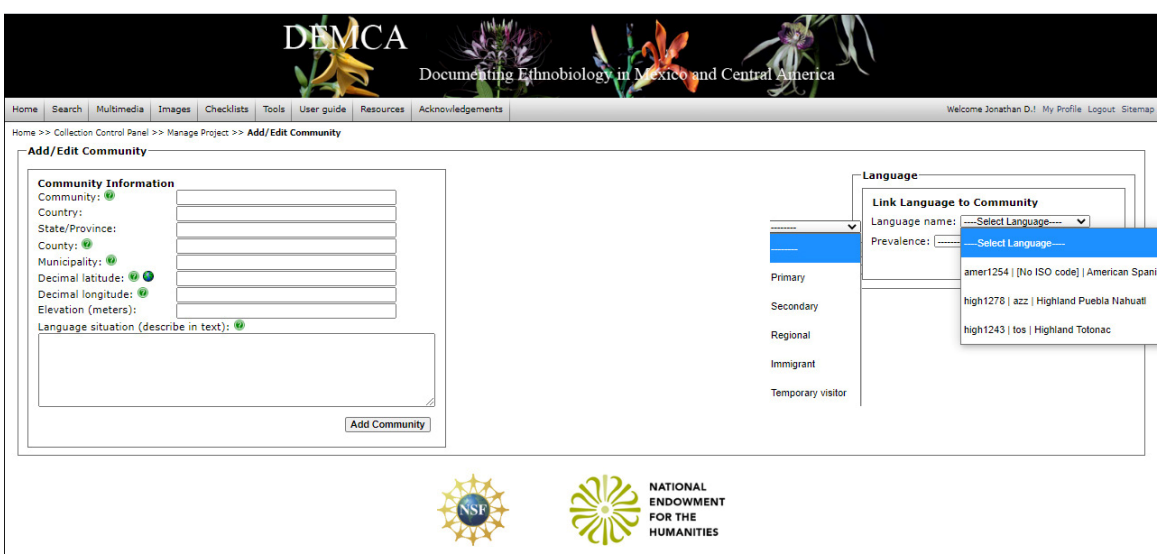
Figure 12: Language Manager: Details



The autofill function of the Language Manager creates a drop-list, in this case of all the languages that begin "toton". These are Totonacan languages. Note that the first column is the Glottolog code, the second column is the ISO 639-3 code, and the third column is the "common name" as Given in Glottolog (see for example https://glottolog.org/resource/languoid/id/filo1235). Note that the second option has "No ISO code". The reason for this is that ISO 639-3 only has codes for "terminal taxa" whereas Glottolog is hierarchical so the non-terminal nodes have no ISO equivalent. Note that the advantage of incorporating all terms into the database is that future user searches on any term will find material relevant to the desired language. Note also that to any language the project administrator may add a URL and a topic for the URL link. This is useful, for example, where there are online articles (perhaps even related to project results) that discussion the situation of the language or the communities in which it is spoken.

3.a.ii Community Manager The next module is the Community Manager (figure 13). This is a mechanism to both standardize community names and associate them with both permanent geographic data (coordinates, elevation) and languages (usually one Indigenous language and a colonial language.

Figure 13: Community Manager : Administtrative, Geographic , and Language Data for all Project Communities

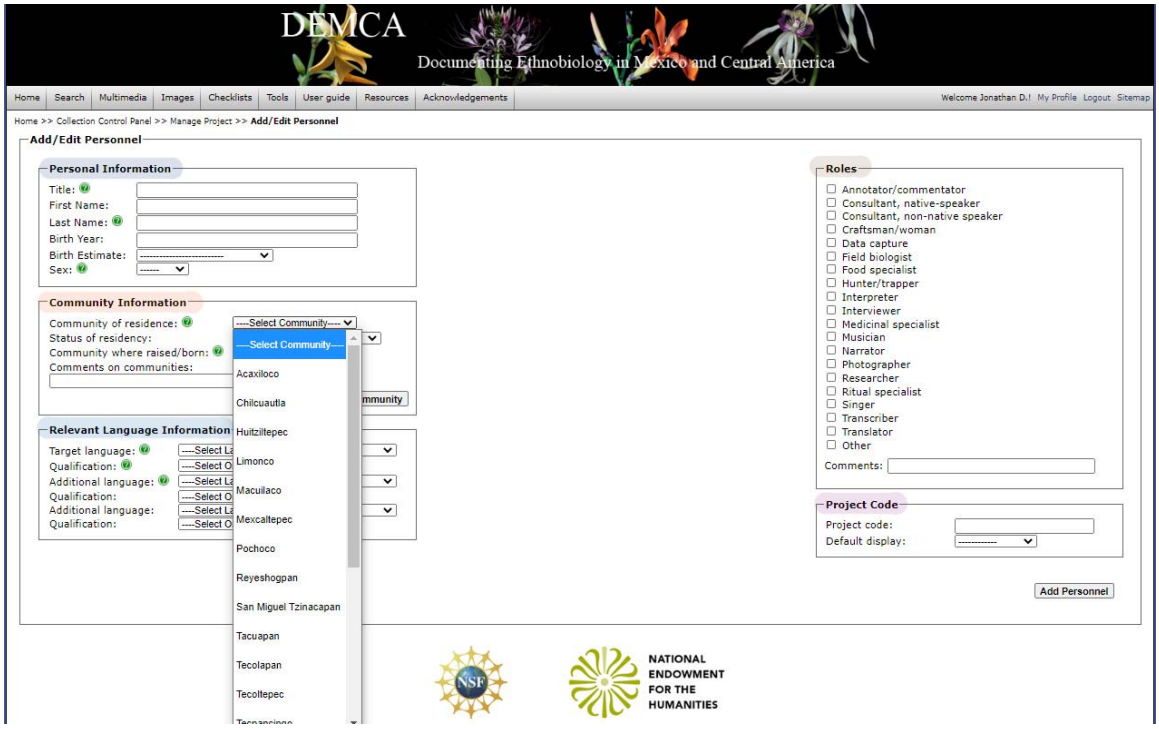


As can be seen, the Community Manager provides for input on the administrative hierarchy for any given terminal community; the latitude, longitude, and elevation; and from a drop-down menu the languages spoken in the community along with their status ('primary', 'immigrant', etc.). The drop-down menu of language options is derived from the languages that were input into the project and another drop-down menu. The input module also contains a space for a textual description of the language situation in the specific community.

3.a.iii Personnel Manager Finally, having documented the Languages and Communities in a project, the next step is to document all the personnel, each of whom is associated with a language and community that must be part of the project database. The personnel manager (figure 14) is the final major step is preparing the project for collection and observation data entry. As can be seen in figure 14, there are five major sections of the manager. The first is personal information (birth year and not birth date is used). The second and third establish the community and language of the individual. The community is filled from a drop-down list populated by communities already entered into the project via the Community Manager. In the personnel form the status of residence can be specified (e.g., an immigrant, worker, native) as well as the place of birth of the individual, should it be different. The language selection is limited to options in a pull-down menus generated by the languages already entered into the Language Module. Note again that specifications can be added such as whether the language is the speakers native tongue or not. The penultimate section links the role that the individual carried out. Finally there is a mechanism to protect the identify of the consultant should that be needed. The individual may be given an arbitrary code that precludes identification. The option for "default display" (which is not visible in the screenshot) offers three possibilities for: (1) individual's name; (2) individual's project code; (3) arbitary code generated by DEMCA.

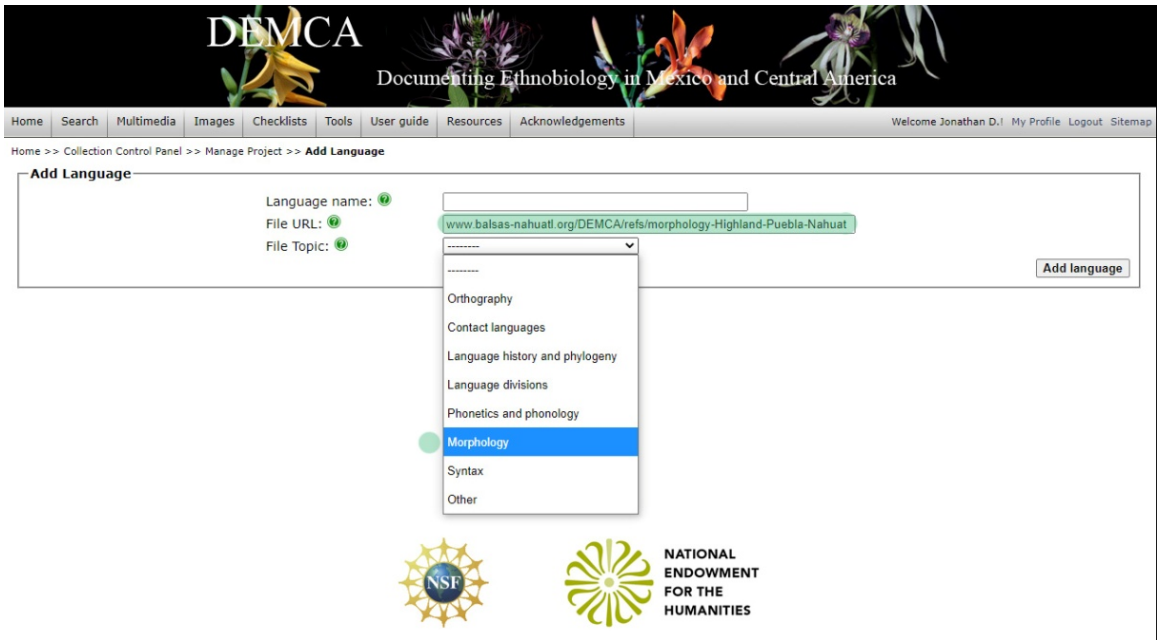
3.a.iv Reference Manager The final manager is the Reference Manager (figure 11, right tab). This is a mechanism to link references, including online references such as pdf files uploaded by the project team such as a pdf file of plant species and their Indigenous name, or perhaps a phonological description of the language so that users could understand any unusual orthographic conventions. Note, however, that the Language Manager also has a mechanism, added in the DEMCA software, that is distinct from the Reference Manage though it also

Figure 14: Personnel Manager



links files, particularly pdf files, that document specific points about the language relevant to the project. Figure 15 is an example of this way of linking references and selecting the relevant topic for each resource in this case a

Figure 15: Linking references through the Language Manager



The preceding discussion presented the DEMCA innovations to Symbiota software, innovations that have been designed specifically for integrating ethnobiological projects into the portal in a way that ensures consistency in the data relevant to Language, Community, and Personnel. Mechanisms have been put in place to ensure the privacy and anonymity of any native language consultant although in general such individuals want credit for their contribution to documenting traditional ecological knowledge and ethical practice is best served by providing this credit as would be due to any co-author of academic research.

3.b Ethnobiological Control Panel: Upload Data Dashboard

The third item of the Ethnobiological Control Panel (figure 11) is the Upload Data module. This is for batch uploading (via a csv file) of basic ethnobiological data from the field input. Figure 16 is representative of the ethnobiological data as it is entered into a database of a project that Amith is leading. Each consultant present at the time of collection is listed along with the Indigenous name (if any) she or he gave to the plant. The research team then adds a parse (morphological analysis) and gloss (meaning of each morpheme) along with two commentary fields, one on the plant name and one on the plant use. These commentary fields are often linked to each consultant although in the case of the data presented in figure 16 the comentaries on name and use are placed in a "synoptic" commentary section of the database. This occurs when the consultants basically agree in their commentaries on nomenclature and use.

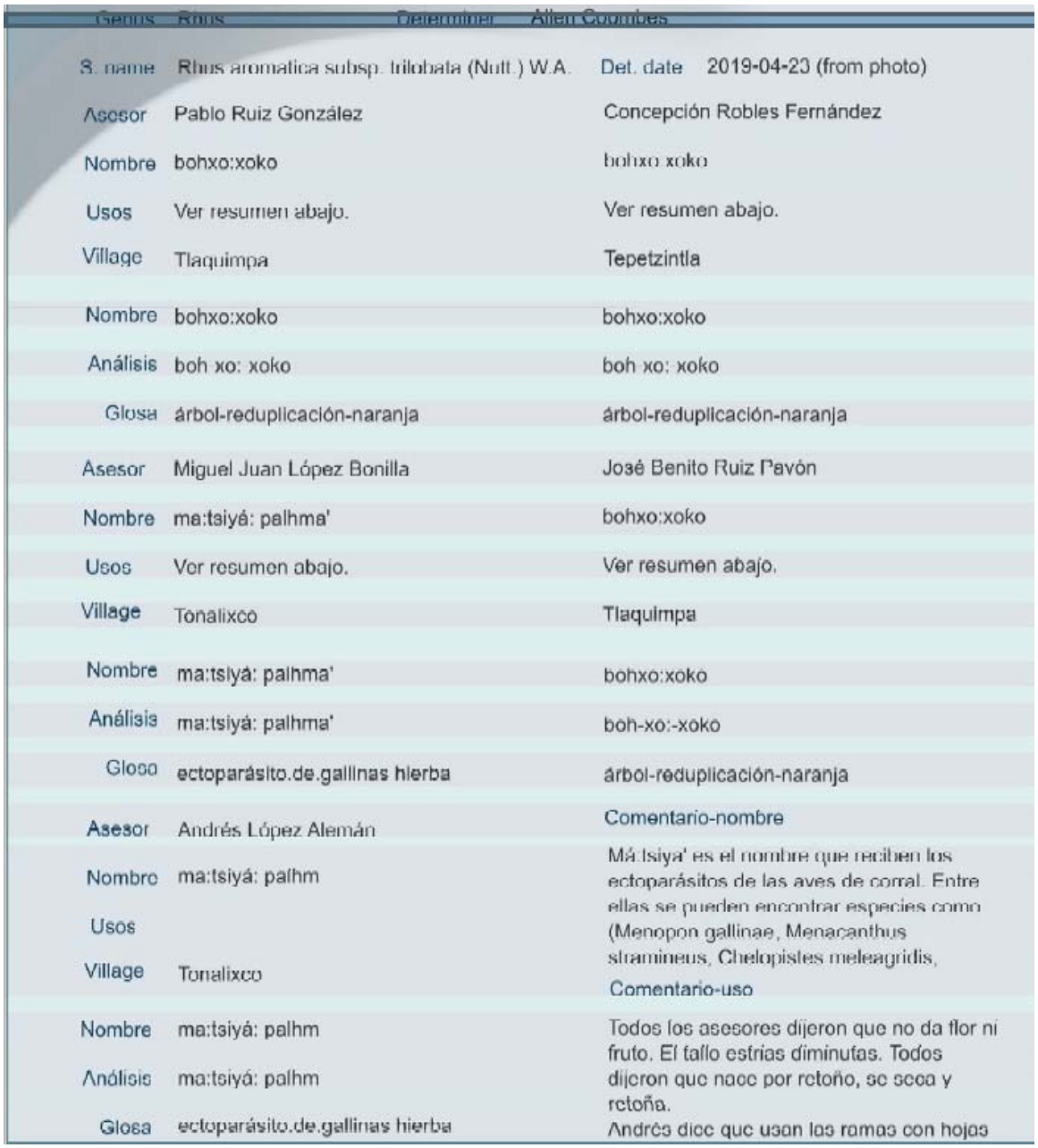
Whether or not a field researcher has separate data entry per consultant or a synoptic summary of data, DEMCA provides a mechanism for back uploading. Figure 17 is the Data Upload dashboard. It shows the two mechanisms for bulk upload of ethnobiological data to the portal:

- Synoptic Data Upload Form
- Vernacular Data Upload Form

The Vernacular Data Upload form is a graphical interface for uploading consultant specific data to the portal. That is, for any collection there will be one or more consultants present and each will be asked to give a name and use for the plant collected. In some instances, of course, the consultant may know no name for the plant, even though

he or she may know a use. In other cases only a use will be known, and no name. As figure 18 shows, the bulk upload module allows for mapping data in the csv file to fields in the portal MySQL database. Basically the workflow process is to extract the field data as entered into the project database (figure 16) and batch upload it to the DEMCA portal via the Vernacular Data Upload Form (when the data is linked to a particular native consultant) or the Synoptic Data Upload Form (not pictured) that ingests commentaries on names and use when they apply to all consultants present at a particular collection.

Figure 16: Data input form for field ethnobiological data



Genus	Rhus	Determiner	Allen Coombes
S. name	Rhus aromatica subsp. trilobata (Nutt.) W.A.	Det. date	2019-04-23 (from photo)
Asesor	Pablo Ruiz González		Concepción Robles Fernández
Nombre	bohxo:xoko		bohxo:xoko
Usos	Ver resumen abajo.		Ver resumen abajo.
Village	Tlaquimpa		Tepetzintla
Nombre	bohxo:xoko		bohxo:xoko
Análisis	boh xo: xoko		boh xo: xoko
Glosa	árbol-reduplicación-naranja		árbol-reduplicación-naranja
Asesor	Miguel Juan López Bonilla		José Benito Ruiz Pavón
Nombre	ma:tsiyá: palhma'		bohxo:xoko
Usos	Ver resumen abajo.		Ver resumen abajo.
Village	Tonalixco		Tlaquimpa
Nombre	ma:tsiyá: palhma'		bohxo:xoko
Análisis	ma:tsiyá: palhma'		boh-xo:-xoko
Glosa	ectoparásito.de.gallinas hierba		árbol-reduplicación-naranja
Asesor	Andrés López Alemán		Comentario-nombre
Nombre	ma:tsiyá: palhm		Má:tsiya' es el nombre que reciben los ectoparásitos de las aves de corral. Entre ellas se pueden encontrar especies como (Menopon gallinae, Menacanthus stramineus, Chelopistes meleagridis,
Usos			
Village	Tonalixco		Comentario-uso
Nombre	ma:tsiyá: palhm		Todos los asesores dijeron que no da flor ni fruto. El tallo estrías diminutas. Todos dijeron que nace por retoño, se seca y retoña.
Análisis	ma:tsiyá: palhm		
Glosa	ectoparásito.de.gallinas hierba		Andrés dice que usan las ramas con hojas

Figure 17: Data Upload dashboard

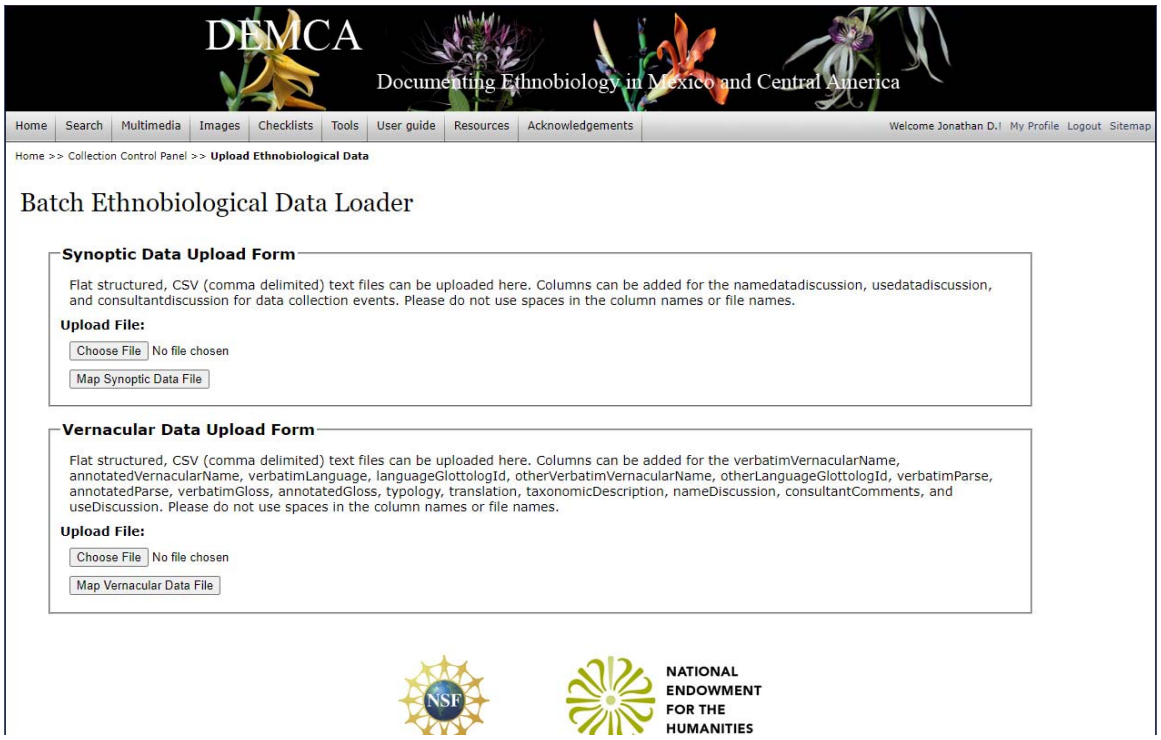
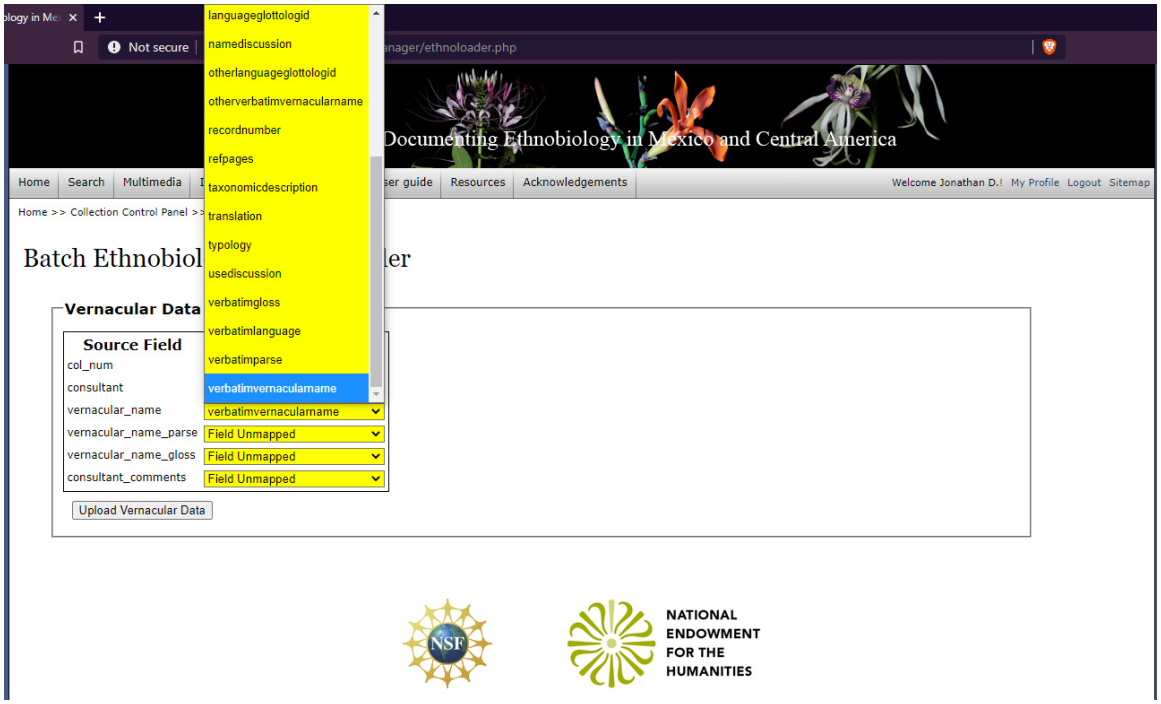


Figure 18: Vernacular Data Upload: File selected, analyzed, and ready for mapping fields



A more painstaking task is the codification of particular facets of nomenclature, classification, and use collection by collection. That is, once the basic botanical and ethnobotanical data has been bulk uploaded to the DEMCA portal the project administrator or editors can encode each entry with ethnobotanical data. This is accomplished via the Mangage Data interface in the Ethnobiological Control Panel (see second option figure 7).

3.c Ethnobiological Control Panel: Manage Data Dashboard

The Manage Data dashboard is essentially a way of tagging each collection or observation with information relevant to ethnobiological research. This information is divided into the following categories:

Semantics and morphology of plant/animal name
Part of plant/animal used
Use

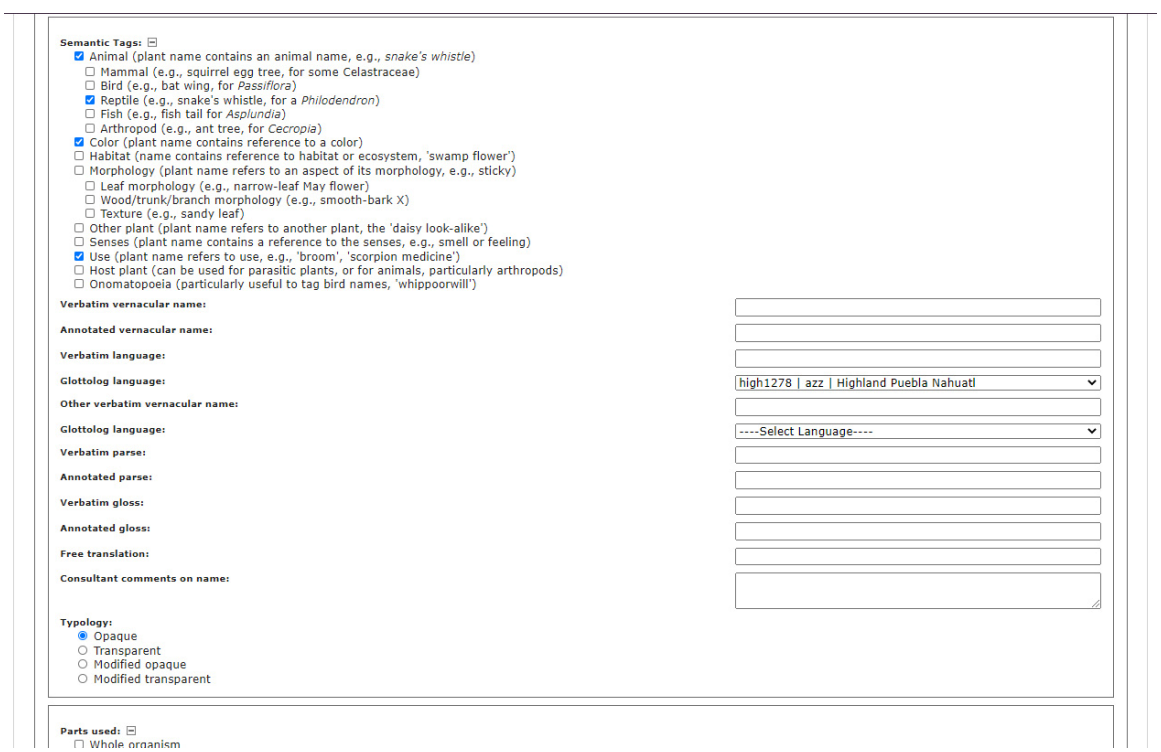
Figure 19 illustrates the way in which vernacular names of flora and fauna can be tagged. This is to supplement the data that is contained in the gloss of Indigenous names for which is often hard to maintain a controlled vocabulary that will enable discovery of the terms sought by a particular user. Note the example of Indigenous names for *Boa constrictor*, which is a calque ('loan translation') shared by many cultures:

Language	Highland Nahuat (high1278)	Highland Totonac (high1243)	Maya
Term	masa:kowa:tl	ju:kilu:wa'	*chij' kan*
Parse	masa:-kowa:tl	ju:ki-lu:wa'	*chij' kan*
Gloss	deer-snake	venado-culebra	venado serpiente

As this table makes clear the fact that both the Nahuat and Totonac terms share a meaning but are in distinct languages identifies them as 'calques'. The fact that the gloss is given in different languages, however, makes discovery non-trivial. This is compounded by the fact that even were all glosses to be in Spanish, different researchers might use 'culebra', 'víbora', or 'serpiente' for what it English would be 'snake' or 'serpent'. To obviate this problem DEMCA has created a series of tags (figure 19) for each vernacular (plant) name that includes the tag

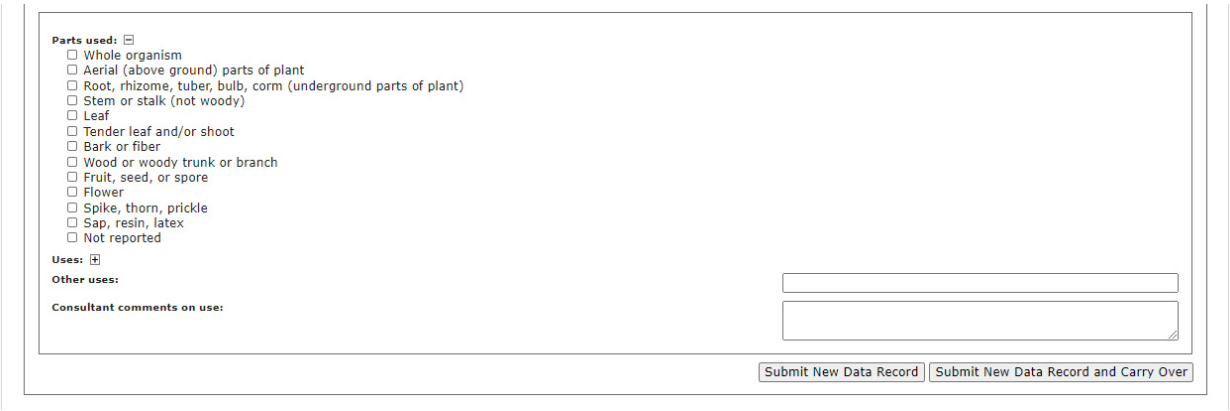
[ ] has a reptile in its name

Figure 19: Tagging of collections/occurrences with ethnobiological information: Name



When this is checked in data input, it is discoverable through the Search collections interface (discussed in continuation): A user will be able to search for all plants that have reptitiles in their name (or even limit this search to certain families of plants). Although this might seem inefficient even in a portal with tens of thousands of records with Indigenous names the number with reptile names would be manageable for browsing. Other tags may be important in distinct situations. For example, the Onomatopoeia tag (the last semantic tag) would be particularly useful in a search for bird terms that exemplify this type of name, which is common though has not been studied in comparative perspective (i.e., are there particular species that are more prone to, across cultures, have names that are onomatopoeic.

Figure 20: Tagging of collections/occurrences with ethnobiological information: Part used



A second area in which collections/observations are tagged is in the part of the plant/animal that is used (figure 20) and the final characteristic that can be tagged is use (figure 21).

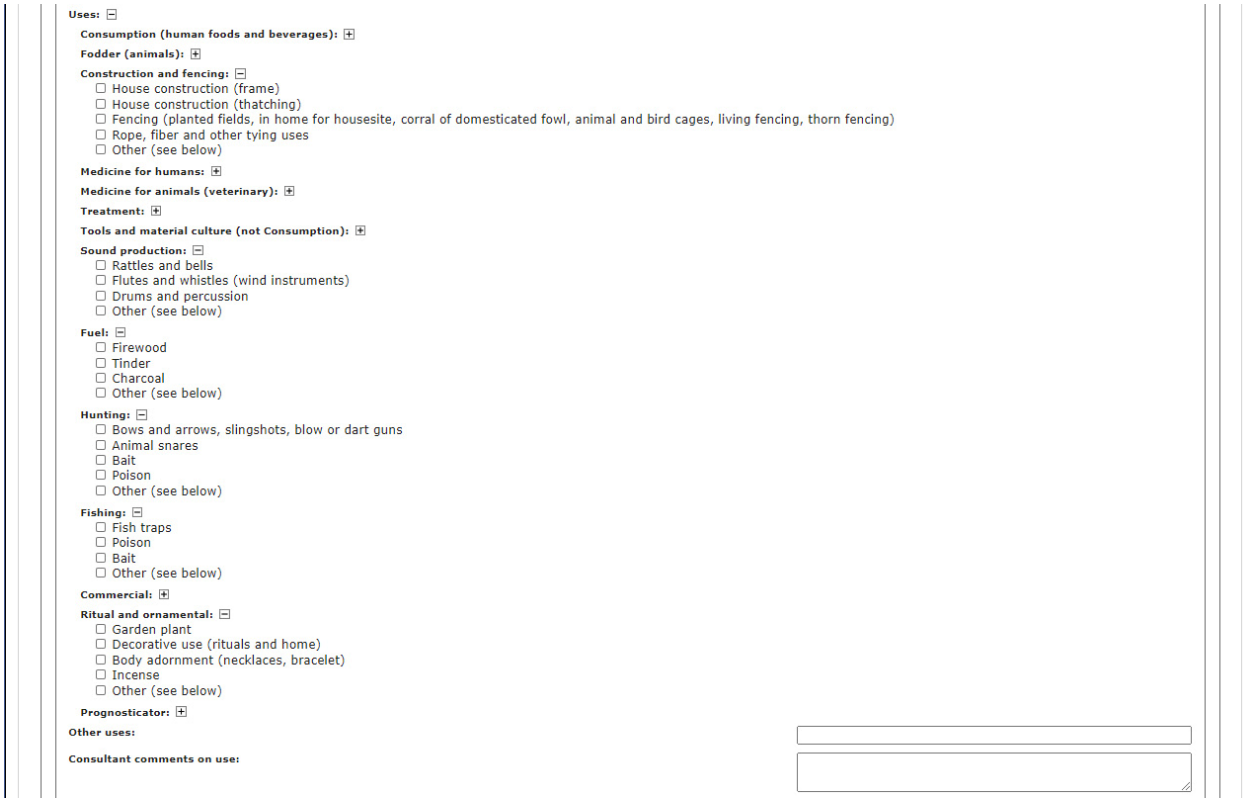
The use tag is a set of controlled vocabularies that has (provisionally) been divided into general rubrics such as ”house contruction", "fuel" and others. As can be seen in the two screenshot (figures 20 and 21), functionality is divided into two main sections, the second of which is multi-layered and expandable. The first section is Parts used. The list in fugure 20 is for plants. Future projects will have a similar set of tags for animals. The next section, figure 21, is Uses, with 13 subsections and a 14th “Other uses” for anything not covered. Each subsection can be expanded, as is the Treatments subsection above which when expanded reveals 16 possibilities. Clicking on any one of the 16 will automatically tag the collected plant as being used in “Treatment” (a general category). Note that these tags are linked to Occurrences. The Taxon page (figures 2b, 34, 35, 36a–g) has a tab for Vernacular uses for any given taxon. This is a synoptic summary statement assembled from many separate collections of the species and any other notes that may be found from other projects.

The next section of this report takes up the issue of searching on ethnobiological criteria, another added feature of DEMCA. Then in conclusion this report returns to the Ethnobiological Control Panel (fig. 7) and presents the final innovative software (line-by-line playback of transcribed/translated sound files) and how it fits it with the goal of the humanities- and social science–driven structure and content of the DEMCA portal.

4. Search capabilities for ethnobiological data

The combination of botanical and ethnobotanical searches creates an opportunity to study functionalities of particular groups of plants (e.g., all Asteraceae used for firewood, or as food or medicine). As the data set grows to include more and more sources, the possibility of comparative research on name (e.g., all plants with an animal as term part of their name; see figure 19), part used (e.g., all trees the sap of which is used in someone way; see figure 20), or the use of the plant (e.g., all trees documented as used in house construction, see figure 21). Not only can multiple ethnobiological criteria be used (e.g., all roots used as food) but the search can be filtered by biological criteria as well (e.g., all Asteracea used as fodder). Figure 22 shows now DEMCA allows for the integration of biological and ethnobiological criteria in a single Boolean search. In fact any combination of criteria is possible. Ethnobiological search capabilities are not easily achieved in any other (ethno)biological data portal.

Figure 21: Tagging of collections/occurrences with ethnobiological information: Use



The input categories for:

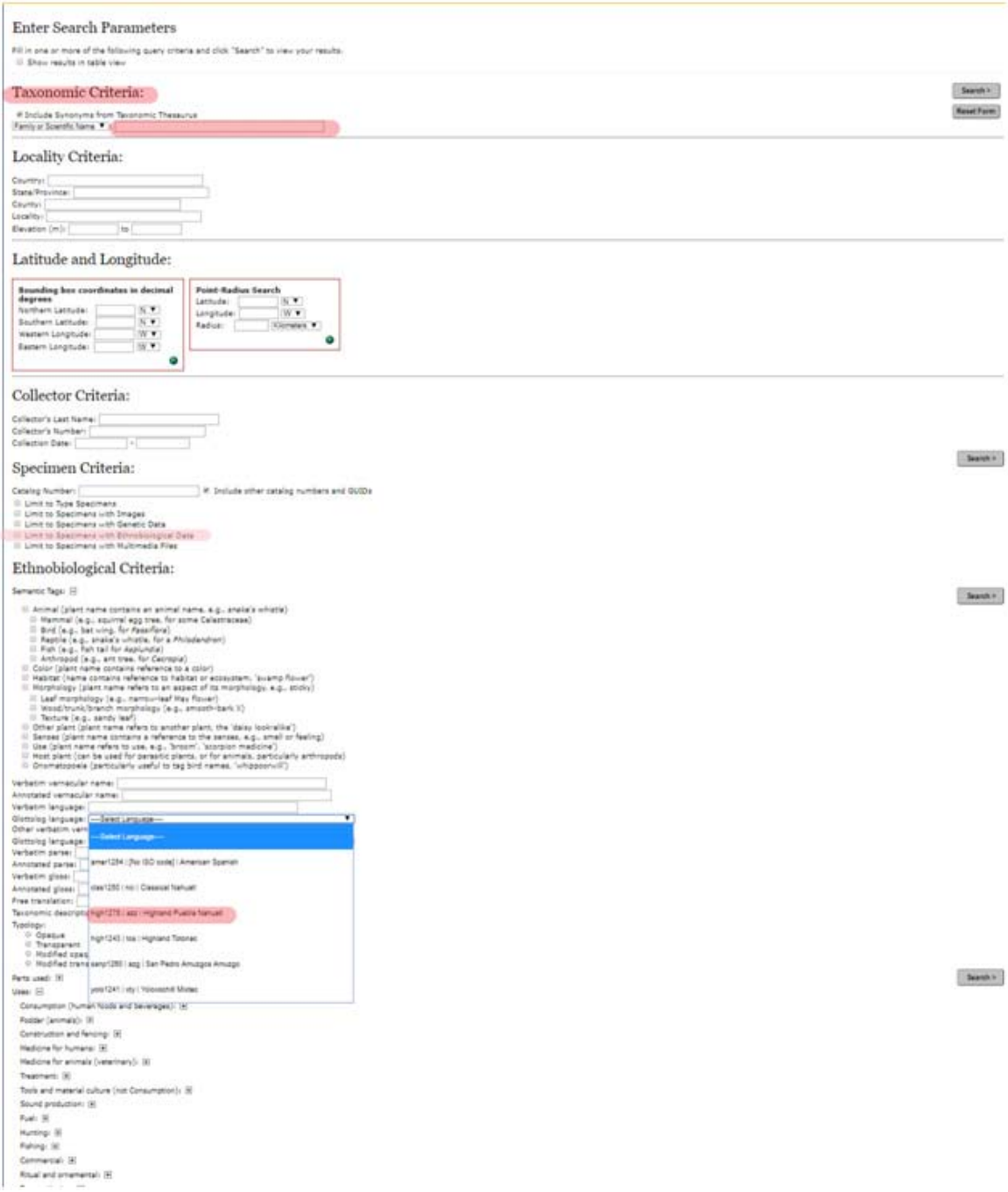
- nomenclature (the semantic criteria noted in figure 19 as well as the data recorded in the </vernacularName>, </vernacularParse>, and </vernacularGloss> fields of individual records.
- part of plant used
- use

As figure 22 illustrates, the search can begin by specifying Taxonomic criteria (e.g., a particular family or species; first two highlighted items at the top of the screenshot). This might be important in generating a list by family of all plants used for a particular purpose (e.g., fuel).

The third highlighted line is under the group of limitations to search results based on Specimen Criteria. For example, "Limit to Specimens with Ethnobiological Data" extends to DEMCA (ethnobiology) a functionality already present in Symbiota: Limitation of search results to any one of several secondary criteria. As can be seen, users can also limit searches to collections/observations with Multimedia files. The inclusion of online multimedia files is another innovation in DEMCA. The other three limits are all in the Symbiota software before the DEMCA enhancements: Limit to type species, Limit to specimens with images, Limit to specimens with Genetic Data.

Finally, the DEMCA Search interface allows for searches based on the collection/observation based criteria that was presented in figures 19, 20, and 21 relating to Name, Part of plant used, and Use. As figure 22 demonstrates the search can also be limited to a particular vernacular language, i.e., those records in which the vernacular name is of a particular language. With this filter, for example, a user can find all the plants with a vernacular name documented for high1278, Highland Puebla Nahuatl (Note that this search is in the collection/observation data, not the Taxon page, described in detail in §6).

Figure 22: Search template for ethnobotanical data



An example of a search from beginning to end is illustrated in the following screenshots (figures 23–27 and figure 2.b, which is the Taxon page for *Ruellia foetida*. The process begins, figure 23, with a search for all the Acanthaceae for which there are multimedia files (two lines of blue highlighting). The result are two collections of different species, one from the Mixtec community of Yoloxóchitl and the other from the Nahuat community of San Juan Tahitic (figure 24). Clicking on "Full record details" for any one of the two results takes the user to a screen such as that in figure 25, which is the Full record for collection #40202. One of the five tabs at the top of this Full record is

Figure 23: Search interface for collections of Acanthaceae with associated multimedia

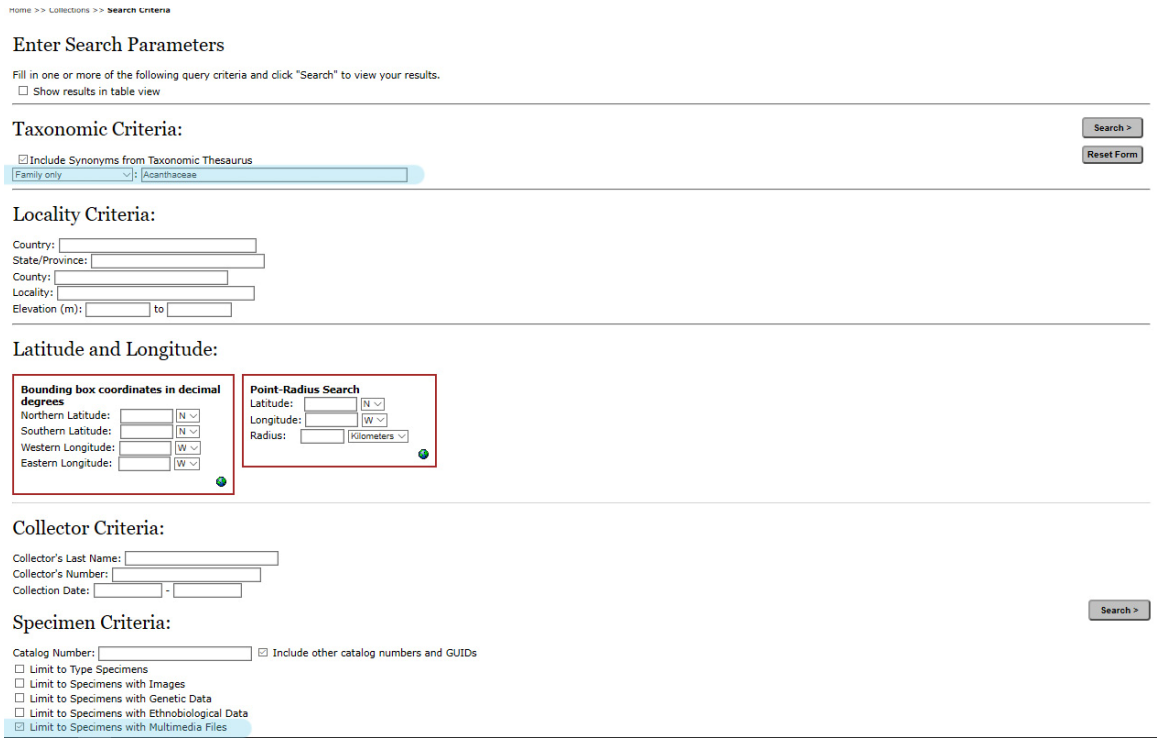


Figure 24: Results of search for Acanthaceae collections with associated media files

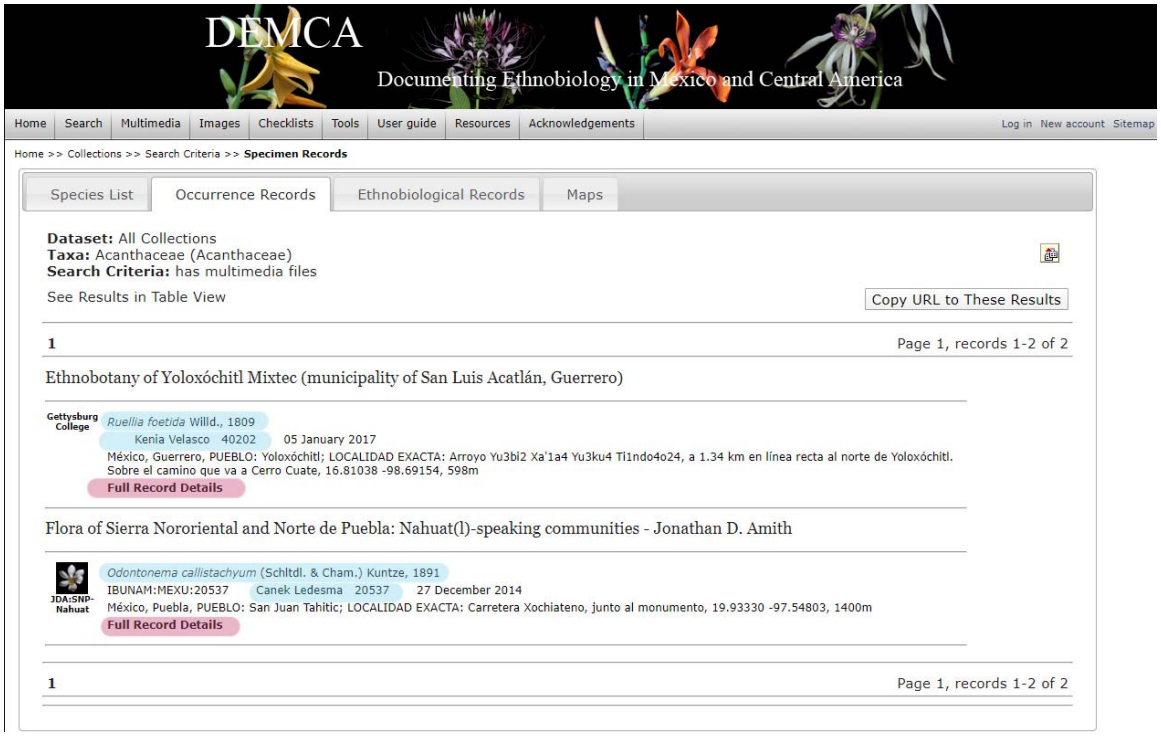


Figure 25: Full Record Details for *Ruellia foetida*

Details | Map | Comments | Linked Resources | Multimedia

Share 0 | Tweet

Gettysburg College IBUNAM : MEXU

Ethnobotany of Yoloxóchitl Mixtec (municipality of San Luis Acatlán, Guerrero)

Taxon: *Ruellia foetida* Willd., 1809
Family: Acanthaceae
Determiner: T. Daniel (2017-03-08 (from photo))
Type Status: NA
Collector: Kenia Velasco 40202
Date: 2017-01-05
Verbatim Date: 2017-01-05
Locality: México, Guerrero, San Luis Acatlán, PUEBLO: Yoloxóchitl; LOCALIDAD EXACTA: Arroyo Yu3bi2 Xa'1a4 Yu3ku4 Ti1ndo4o24, a 1.34 km en línea recta al norte de Yoloxóchitl. Sobre el camino que va a Cerro Cuate.
16.81038 -98.69154 WGS84
Elevation: 598 meters
Verbatim Elevation: 598 m
Habitat: VEGETACIÓN y HÁBITAT: Secundaria de bosque tropical subcaducifolio, a lado del camino con Nectandra, Andira, Ficus y acantáceas | A 5 m del arroyo.
Individual Count: 1
Preparations: Herborizado

Usage Rights: CC BY-NC (Attribution-Non-Commercial)
Record Id: d4eb4c24-2dff-421c-85b1-8a7d4211204c
Occurrence ID (GUID): d4eb4c24-2dff-421c-85b1-8a7d4211204c

For additional information on this occurrence, please contact: Jonathan D. Amith (nahuatl.biology@gmail.com)

Figure 26: Clickable link: Occurrence Record to line-by-line presentation of multimedia file for *Ruellia foetida*

Details | Map | Comments | Linked Resources | Multimedia

Col. #40202: *Ruellia foetida* Willd. (*Yu¹ku¹ ko¹ndo¹*) Grabación de campo

Fig. 27 Line-by-line playback of recording associated with Collection 40202 (*Ruellia foetida*) https://demca.mesolex.org/portal/ethno/eaf/eafdetail.php?mediaid=180

DEMCA
Documenting Ethnobiology in México and Central America

Home | Search | Multimedia | Images | Checklists | Tools | User guide | Resources | Acknowledgements | Log in | New account | Sitemap

Home >> View EAF Files >> Col. #40202: *Ruellia foetida* Willd. (*Yu¹ku¹ ko¹ndo¹*) Grabación de campo EAF Detail

Col. #40202: *Ruellia foetida* Willd. (*Yu¹ku¹ ko¹ndo¹*) Grabación de campo

(Open to customize display)

00:00 03:07

KVG Ejemplar 40202,

T Ejemplar 40202,
T familia Acanthaceae,
T eeeh, especie *Ruellia*
T probablemente *foetida*.
T Eeeh, hierba de sesenta, setenta centímetros
T o hasta un metro, de flor blanca
T y crece a orilla del camino.
T **Constantino Teodoro Bautista.**
T *Yu¹ku¹ ko¹ndo¹*, este que llamamos *ndu³ku² ko¹ndo¹*
T no sirve para nada sino solo
T para ensartar flores de cacaloxóchitl en los tiempos que ya se secan como por le mes de
T marzo o baril cuando se secan y quedan sólo las varas,
T las puras varas ahí paradas y les quitan las semillas para que
T que ensartan las flores de cacaloxóchitl para llevar al cerro de la lluvia, a la iglesia, al camposanto, le ensartan flores de cacaloxóchitl y
T y esa vara en el tiempo de [...]

for Multimedia. The Multimedia tab reveals whether any multimedia is associated with this collection (#40202). Note that audio media connected to a collection/observation record are by definition recordings made in the field at the moment of collection/observation. The filename by definition begins with the collection number, in this case 40202. Additional "studio" recordings of discussions among native speaker specialists are only accessible through the Taxon page (see dicussion §6)

The next section covers another innovation of the DEMCA portal, the ability not only to present line-by-line transcription/translation of native language recordings but to integrate these recordings in the DEMCA portal in three ways:

- Linked to collection events (see figures 25, 26 above)
- Linked to Projects (see figure 28, below)
- Linked to Taxon pages (see figure 34, 35, 36a-g, below)

5. Multimedia (audio) line-by-line playback

Ethnobiological projects often include multimedia documentation of traditional ecological knowledge. These may include videos of material cultural production (for examples, https://www.youtube.com/watch?v=lUQnMiF6OcI&t and https://www.youtube.com/watch?v=XzumbCmedSM&t among many others). These videos can be linked via URL to a DEMCA Taxon page). The most important innovation in DEMCA is, however, line-by-line playback of audio with accompany transcription and translation (for an example, see https://demca.mesolex.org/portal/ethno/eaf/eafdetail.php?mediaid=105&collid=0) This section will describe the upload and playback module and how all audio is linked to a particular project. The final section (§6) will discussion the Taxon page, a page that integrates biological and ethnobiological data in one location.

Multimedia upload and linking is accessible through the final tool in the Ethnobiological Control Panel: Manage EAF Files (figure 7). Note that the Ethnobiological Control Panel is accessed once an administrator has selected a particular project and then choses Manage EAF Files. Figure 28 is the initial dashboard for uploading an EAF file to the DEMCA portal so that it can be changed to HTLM format linked to an mp3 file already online. As is evident the list of EAF/Sound files displayed are those that are linked to the project in Nahuat-speaking the Cuetzalan del Progreso. Pulling the drop-down menu at the right of the Collection/Project line will reveal all other projects in which there are multimedia files. Clicking on the filename (e.g., *Ma:talin (Comelina sp.) [azz]*) will take one to the line-by-line playback page. Clicking on the pencil next to the filename will take a project administrator or editor (the pencil editing icon would not appear if a regular user accessed the project page of EAF recordings).

Figure 28: Initial dashboard for EAF file upload

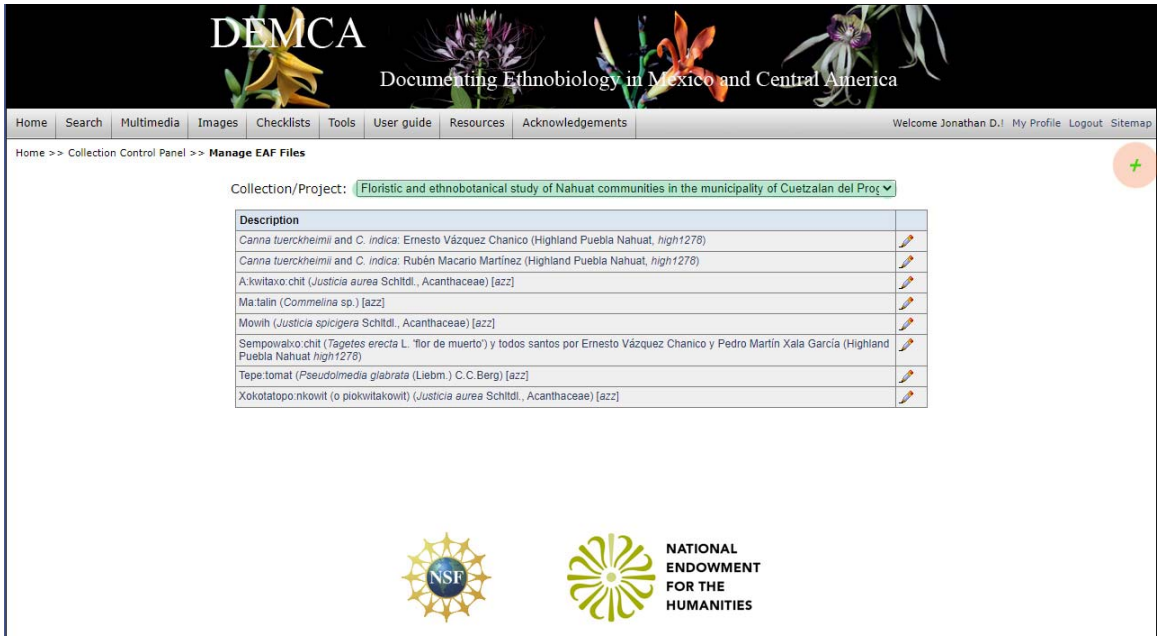


Figure 29: EAF file upload: Choose file

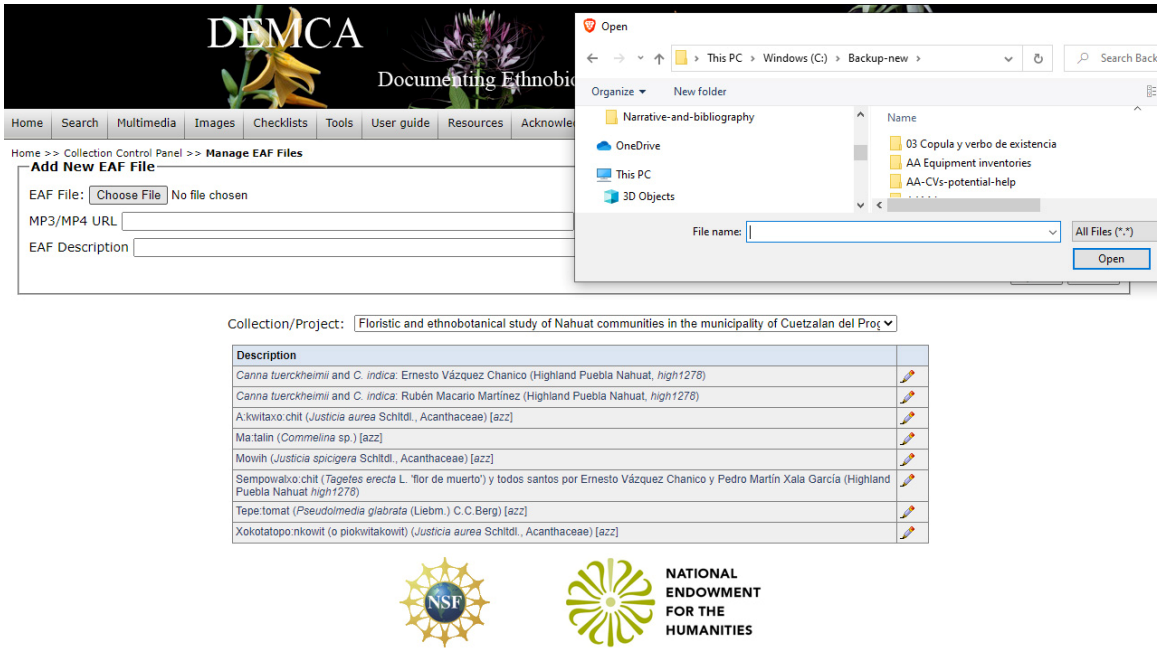
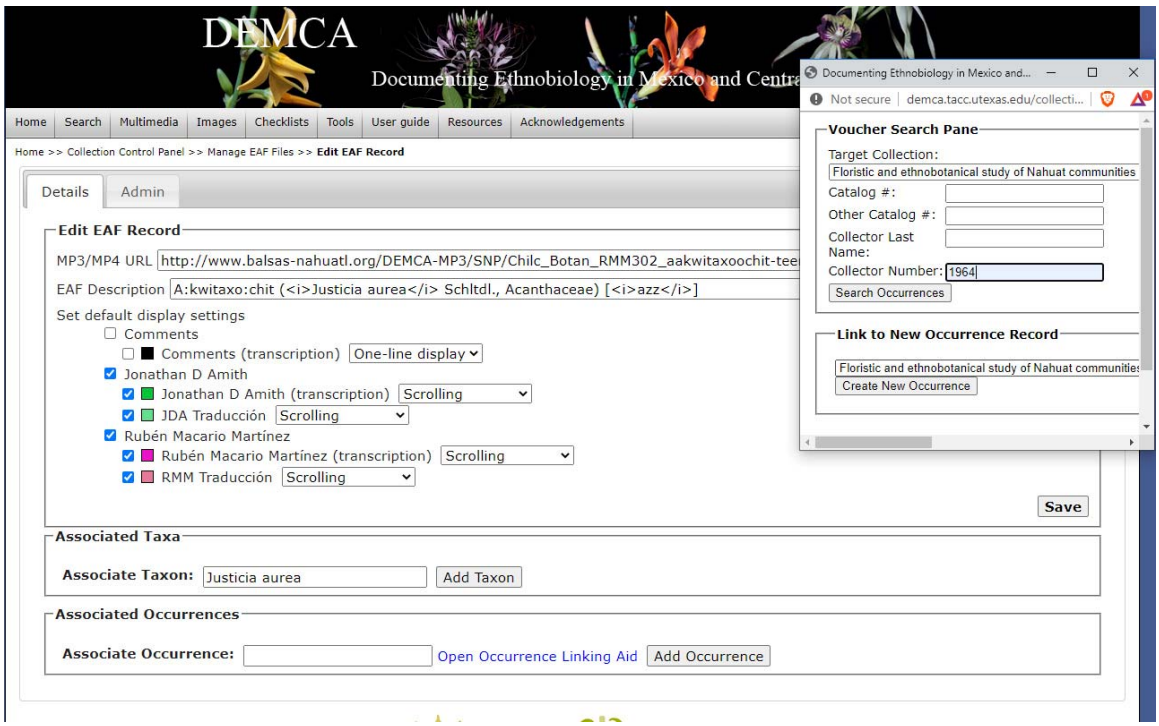


Figure 30: EAF file upload: Configure default display



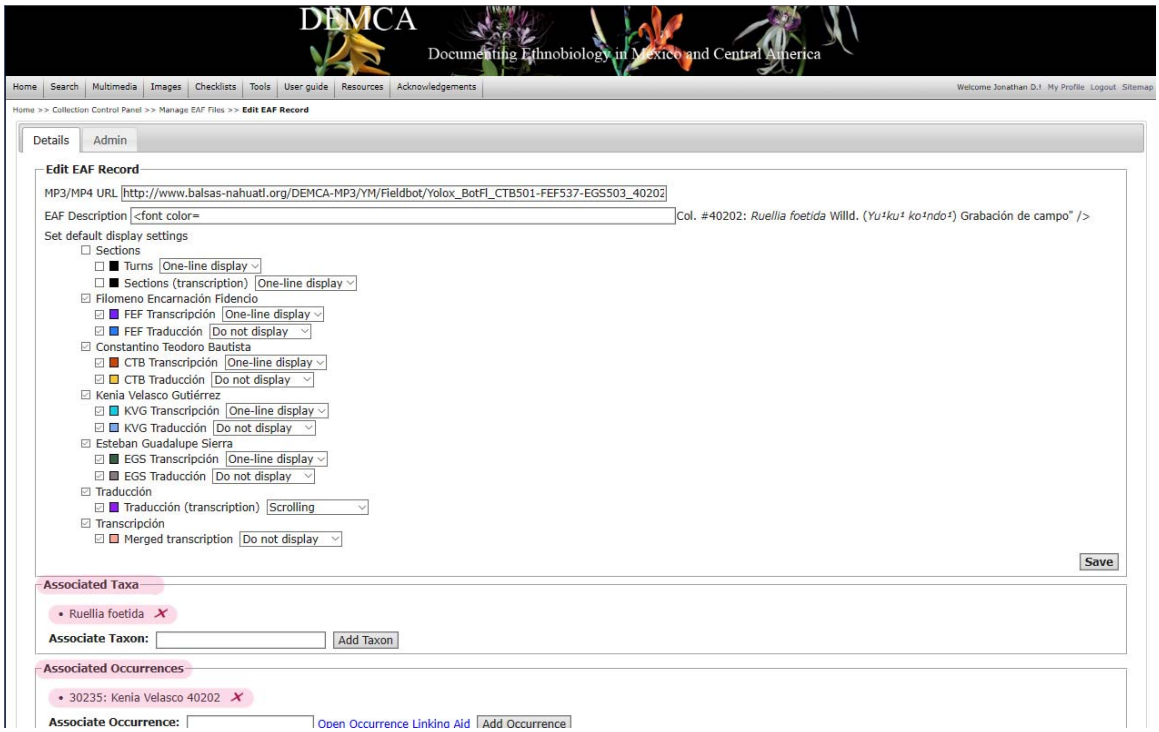
The EAF upload is fairly easy. The first step (figure 29) is to choose a file to upload. Once the file is chosen the project administrator/editor continues adding the remaining required information.

mp3/mp4 URL This is the full path to the location of the multimedia file, which can be on any server

1. EAF This is the title of the EAF HTML page (the two-line descriptive title on the EAF-HTML page in figures 31, 32).
2. Default display setting This is the view that they user will see as s/he access the EAF-HTML page (figures 31, 32)
3. Associated taxa These are the Taxon pages to which the EAF-HTML page will be linked. (see figure 32)
4. Associated occurrences This is the Occurrence record (collection or observation) to which the recording pertains. Only field recordings at the time of plant collection can be related/linked to Associated Occurrences (see figure 32)

1. Title of EAF completely open in regard to content though it will usually include the Indigenous vernacular name, the scientific name, the speakers and the Glottolog code for the language. HTML tags can be included in the title name and these tags are reflected in the display of the title as the italics in the title on figure 32 attest).
2. Default setting disply is where the administrator/editor selects both which EAF data is uploaded to the server (and there converted to HTML) and how the uploaded EAF data appears. In the example (figure 30) the Comments are left unselected and therefore the content of the comments field is not uploaded and not made available to the general public. The structure of the EAF file is read and as evident there are two speakers (Jonathan D. Amith and Rubén Macario) and each has a line corresponding to a transcription of what they said and then another line which is the translation. By checking the boxes (in blue) the administrator/editor ensures that the content is uploaded to the server. The administrator/editor can then choose the default display: (a) a check in the box to the left indicates that line of text (e.g., FEF Transcription) will be uploaded to the data portal, an absence of a check means that the material will not be uploaded to the portal; (b) the square next to the one for the check to upload data select the font color;

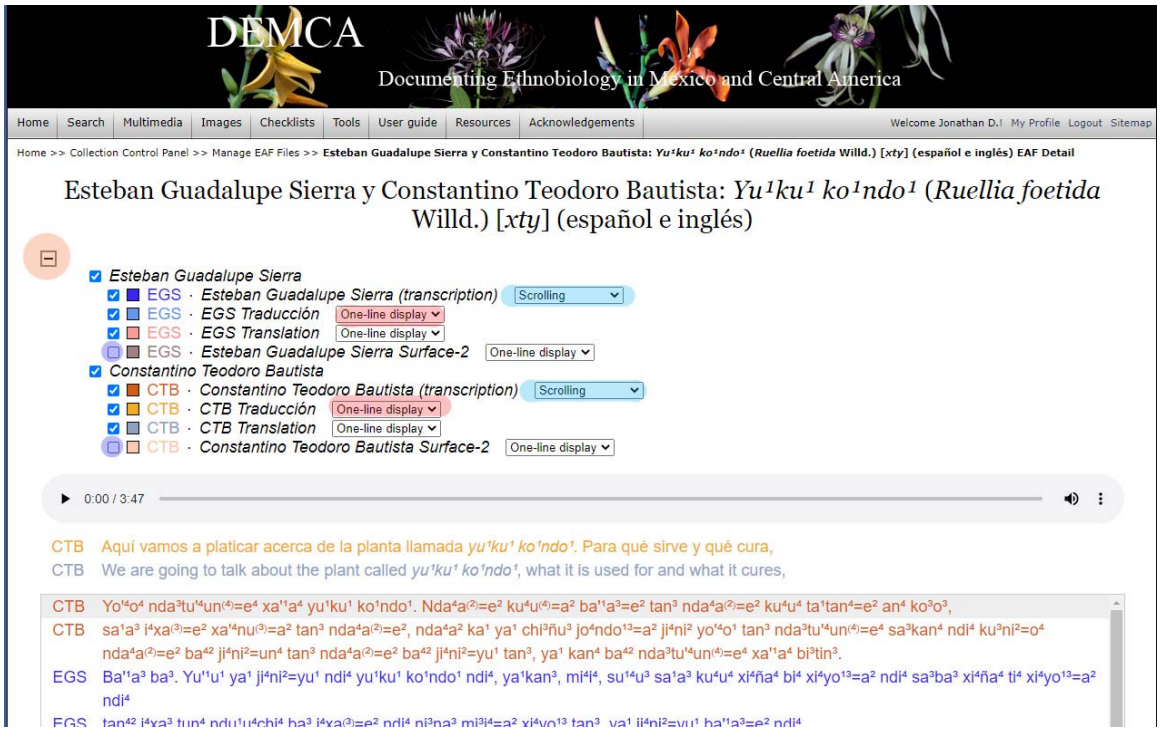
Figure 31: EAF file upload: Configure default display



(c) the final determination by the administrator/editor is whether the uploaded data should display or not. If the default is to not display the user will nevertheless see the option to display (and how to display) in the User confirmation interface (figure 32). For the file of figure 31 note that some lines are "Do not display", others are "One-line display" and the others are "Scrolling". The user controlled reconfiguration option is described below. The advantage of upload under "Do not display" criteria is that the upload could, for example, include several translations into different languages, or writing and a glossed lines of text.

3. Associated taxa is a mechanism by which the administrator/editor can link the EAF-HTML page to a Taxon page , a part of the DEMCA portal that is described in §6. In this case *Ruellia foetida* is linked to the species Taxon page. This means that it is automatically included in the multimedia tab for Ruellia foetida (figure 2b). If the recording is relevant to more than one species then additional Associated Taxa can be linked to the EAF-HTML page via this Dashboard.
4. Associated Occurrences is a mechanism by which the EAF-HTML page can be linked to Occurrences (i.e., plant collection or observation events). Only audio that is associated with such an event can be linked to a collection record. Figures 25 and 26 illustrate this type of link as it is displayed in a Multimedia tab of a collection. Note that these field recordings can also be associated with or linked to Taxon pages as is the case with #40202 and its link to the Taxon page *Ruellia foetida* and to the collection record 40202. Note that figure 30 shows a tool for finding the SQL reference for the field recording by typing in the collection number.

Figure 32: EAF file display and user configuration



The final innovation for online, line-by-line display of multimedia in DEMCA is access by the user and the inclusion of a mechanism by which users can configure and personalize the display. Figure 32 has a collapsable menu (orange circle to the top left) that is here expanded but can be collapsed with a click. Once expanded the user and select or deselect lines for display or non-display and can select whether they appear above or in the scroll-box. In the screenshot the two translations, Spanish and English, are displayed above the scrollbox. The user could eliminate one or the other, or place the translations in the scroll box and the transcription in the One-line display. In the above screenshot there are no lines that were uploaded but not displayed. If there had been the user could select them for display, perhaps replacing some other lines that in the default were displayed.

6. The DEMCA portal Taxon pages

The DEMCA portal, like all Symbiota portals, has what are called Taxon pages. These are pages that can be at the level of species, genus, or family. A Taxon page at the family level (figure 34) automatically generates thumbnails for all the species in that family at are in the DEMCA portal. These are clickable and doing so remits the user to the Taxon species page clicked. But a Taxon page can also be searched at the genus level. Doing so will take the user to a Taxon genus page (figure 35). As seen here there is a list of 5 species each with an iconic in situ species and a map. Clicking on the species takes the user to a Taxon species page such as that depicted in figures 36a-f. Taxon species pages will be discussed below. Clicking on the map will enlarge the map and show the distribution of the species clicked. Thus clicking on the map below Conostegia icosandra will open up a map with an orange dot for each collection of this species. The dot is also clickable and will take the user to a detailed summer of the species (see figure 1c).

One method for accessing any Taxon page at any level is directly through a Taxon Search box on the DEMCA home page (figure 33). Another way, not depicted here, is through a link that is available through any occurrence listing. The Taxon search box has an autofill function that will produce a drop-down list of possible taxa when at least four letters are typed in. The drop-down list in figure 33 was created as the user searched for *Canna tuerckheimii*. The Taxon page for this species, which is the most complete Taxon page at the present time, is despited in figure 36a-f. However, a Taxon search can also be at the level of genus and in this case will generate a link to a family or genus pages such as that for the family Melastomataceae (figure 34) or the genus *Conostegia* (figure 35).

A discussion of the a model Taxon page, that for Canna tuerckheimii, begins on page 34 after figures 36a–36g, which illustrate then entire content of a Taxon species page as it is relevant to ethnobiological studies of great importance to the humanities and social science.

Figure 33: DEMCA home page with Taxon Search box

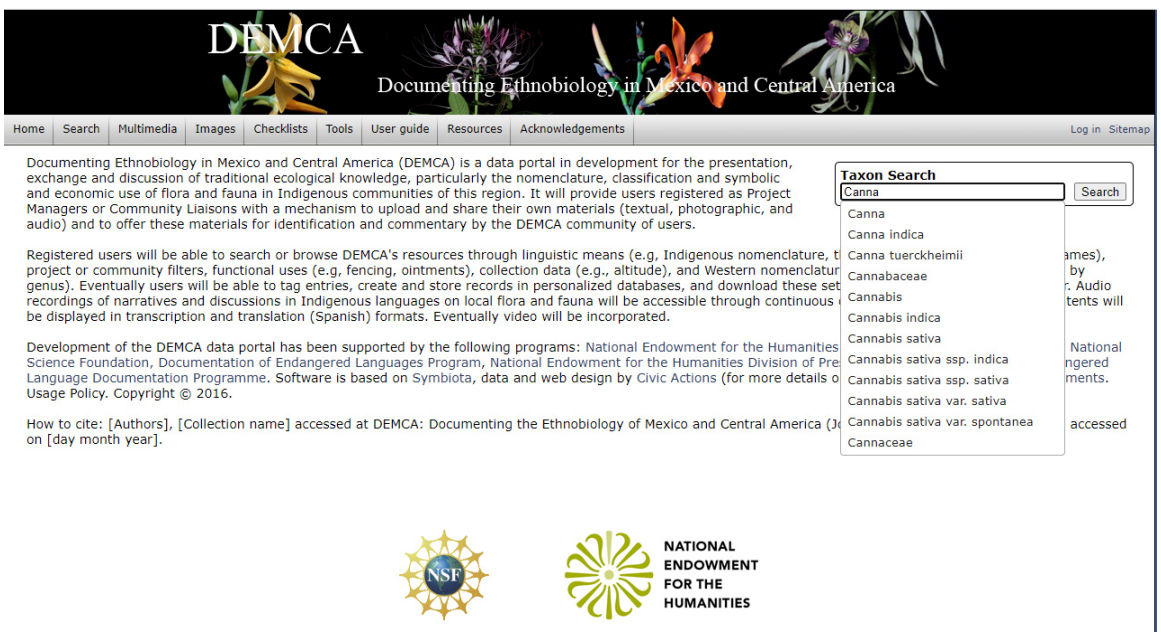


Figure 34: Taxon page at the family level: Melastomataceae

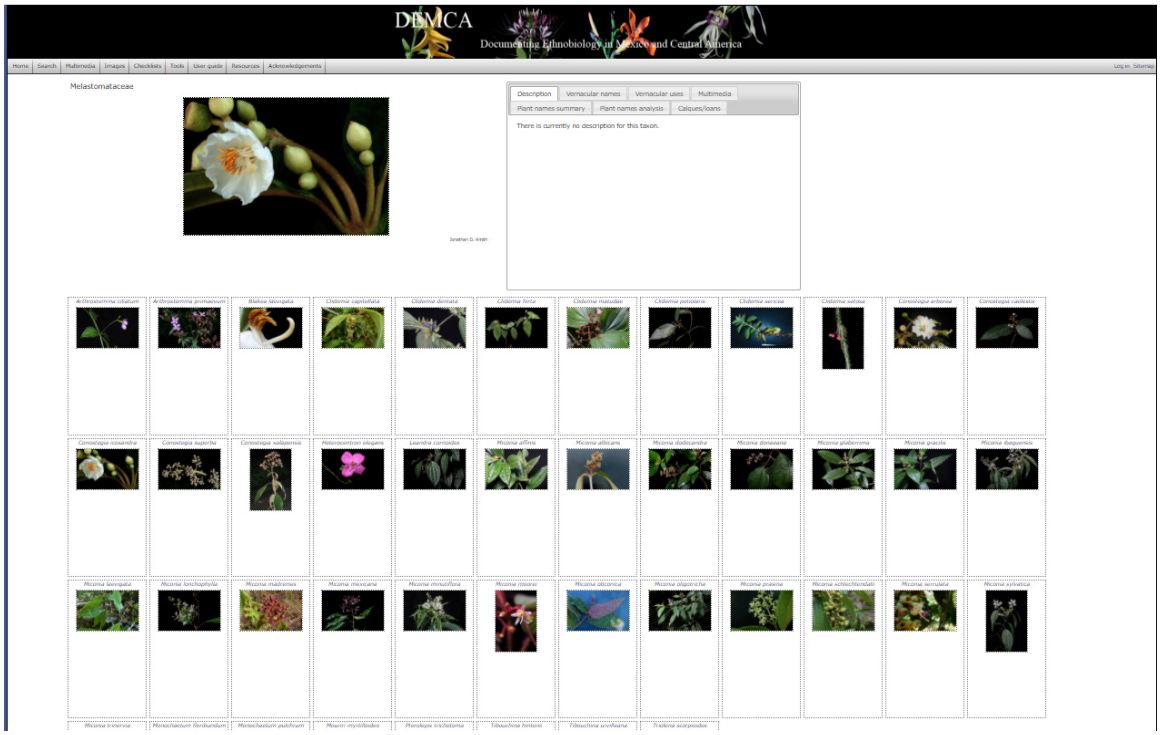


Figure 35: Taxon page at the genus level: *Conostegia* (family Melastomataceae)

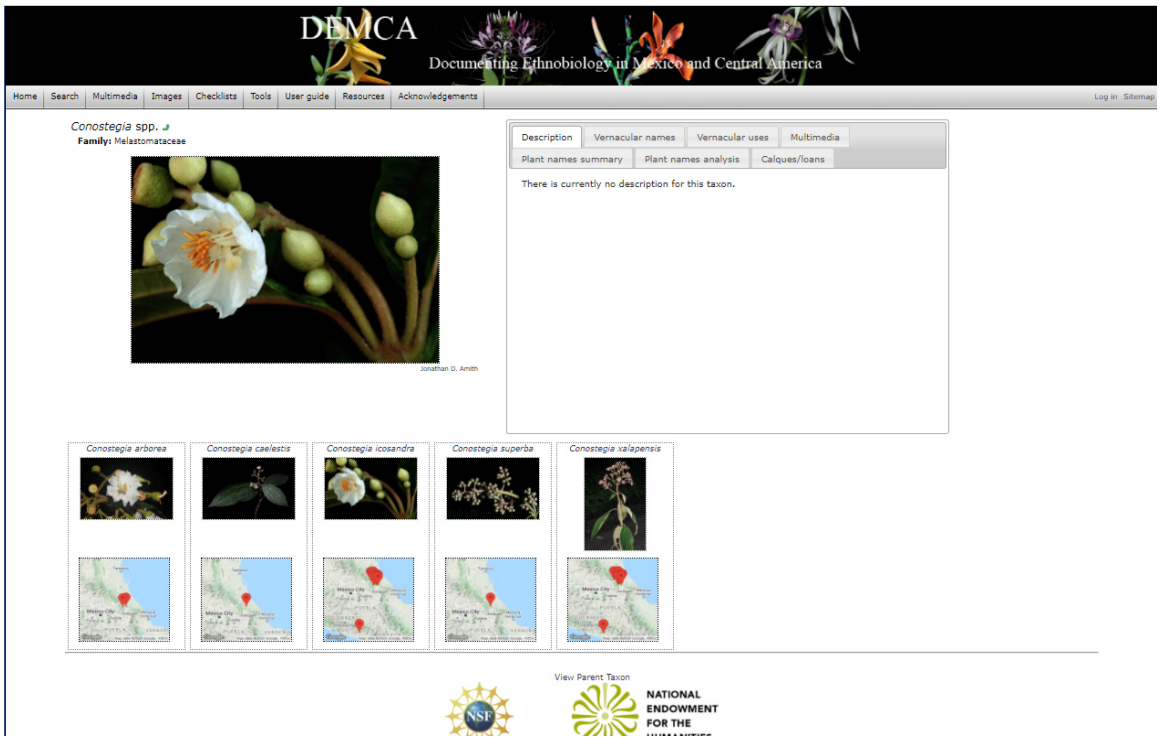


Figure 36a: Taxon page *Canna tuerckheimii*: Description
https://demca.mesolex.org/portal/taxa/index.php?taxon=Canna+tuerckheimii&formsubmit=Search+Terms

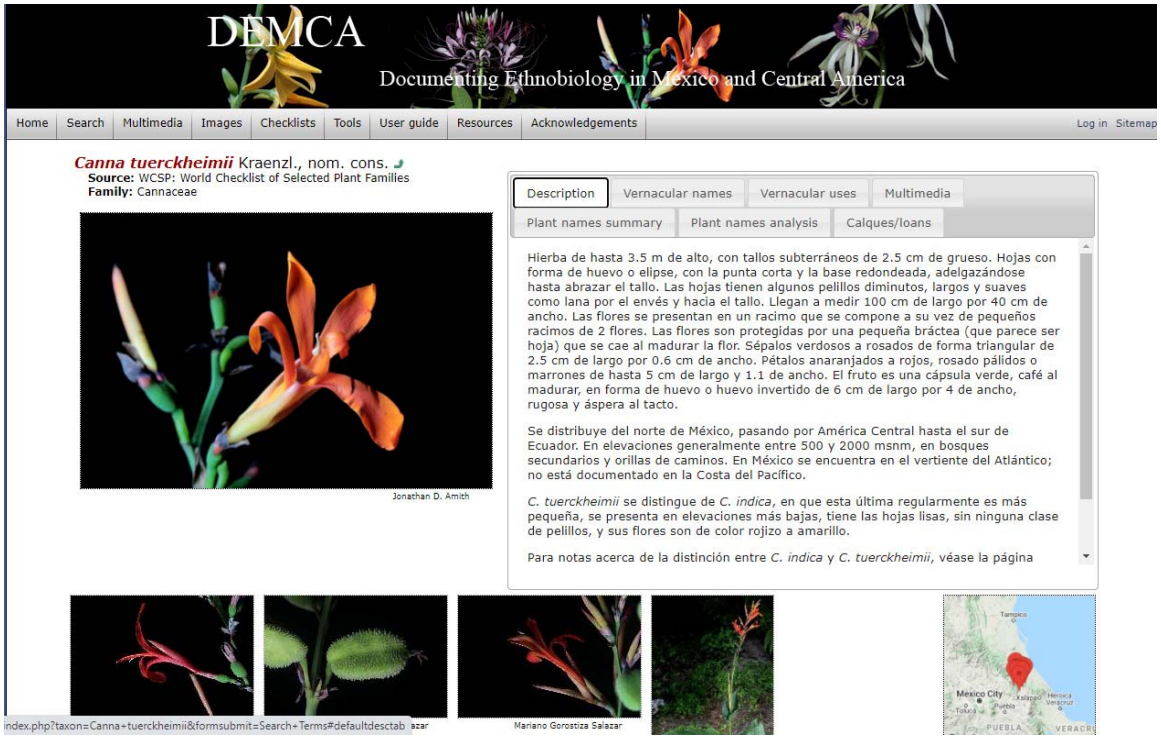


Figure 36b: Taxon page *Canna tuerckheimii*: Vernacular names

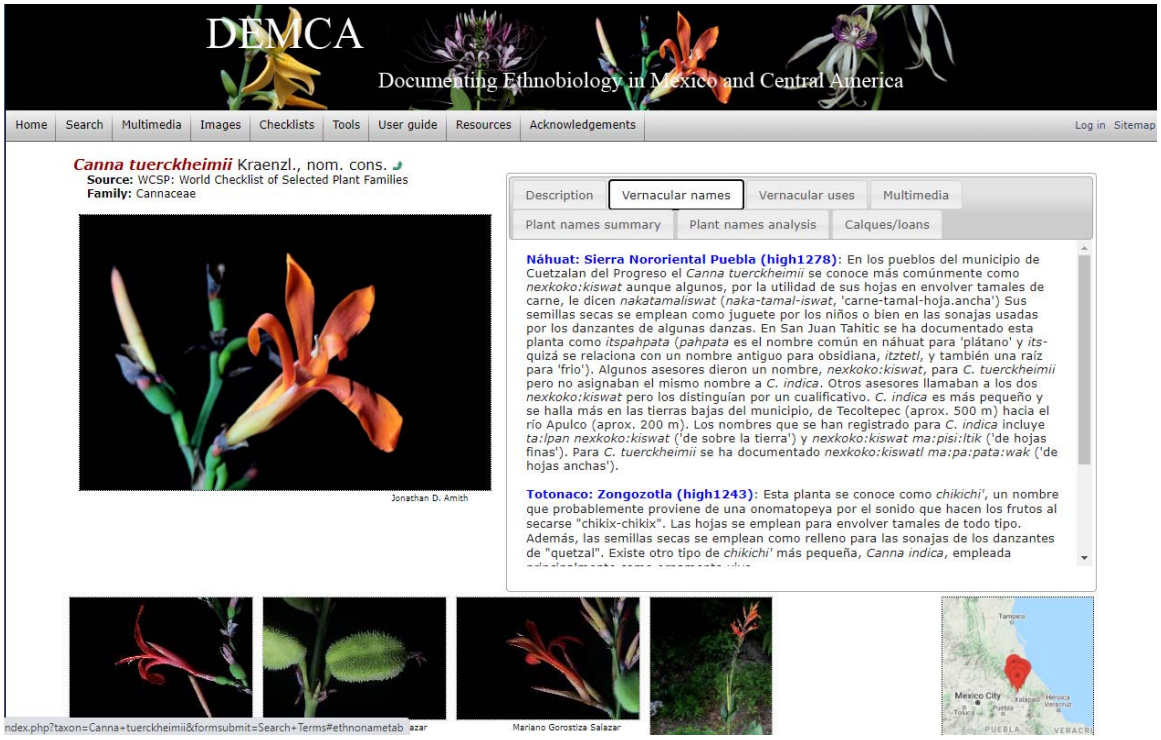


Figure 36c: Taxon page *Canna tuerckheimii*: Vernacular uses

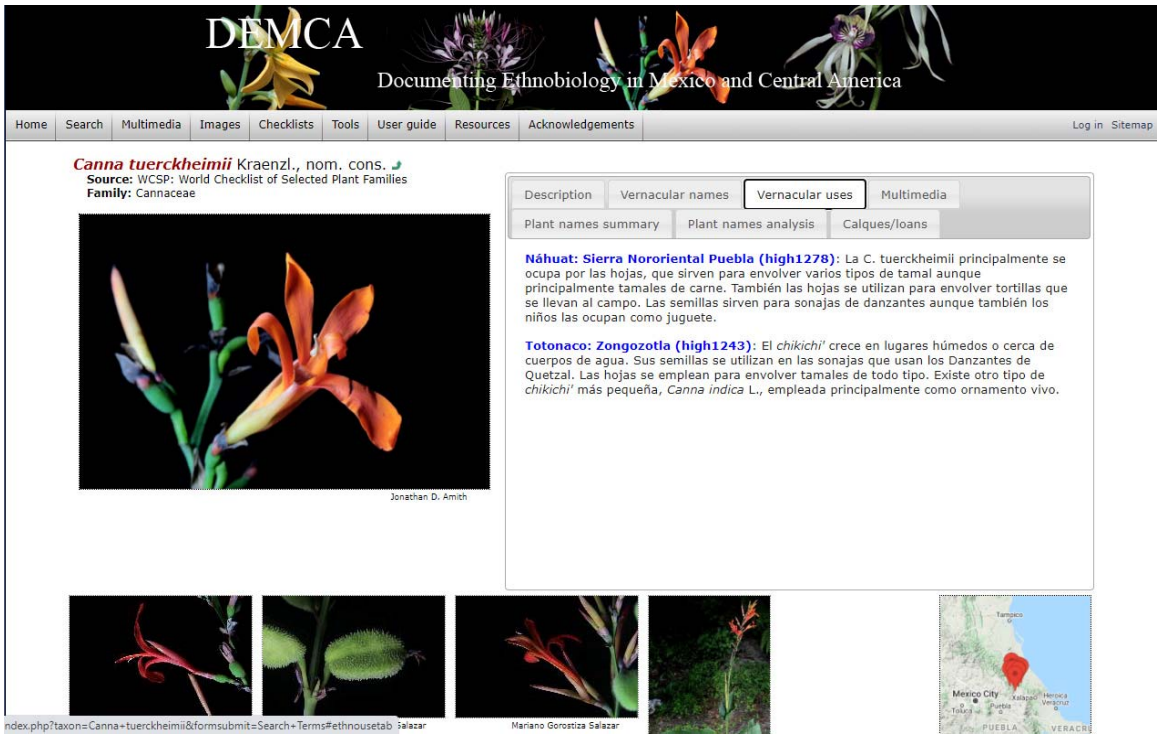


Figure 36d: Taxon page *Canna tuerckheimii*: Multimedia

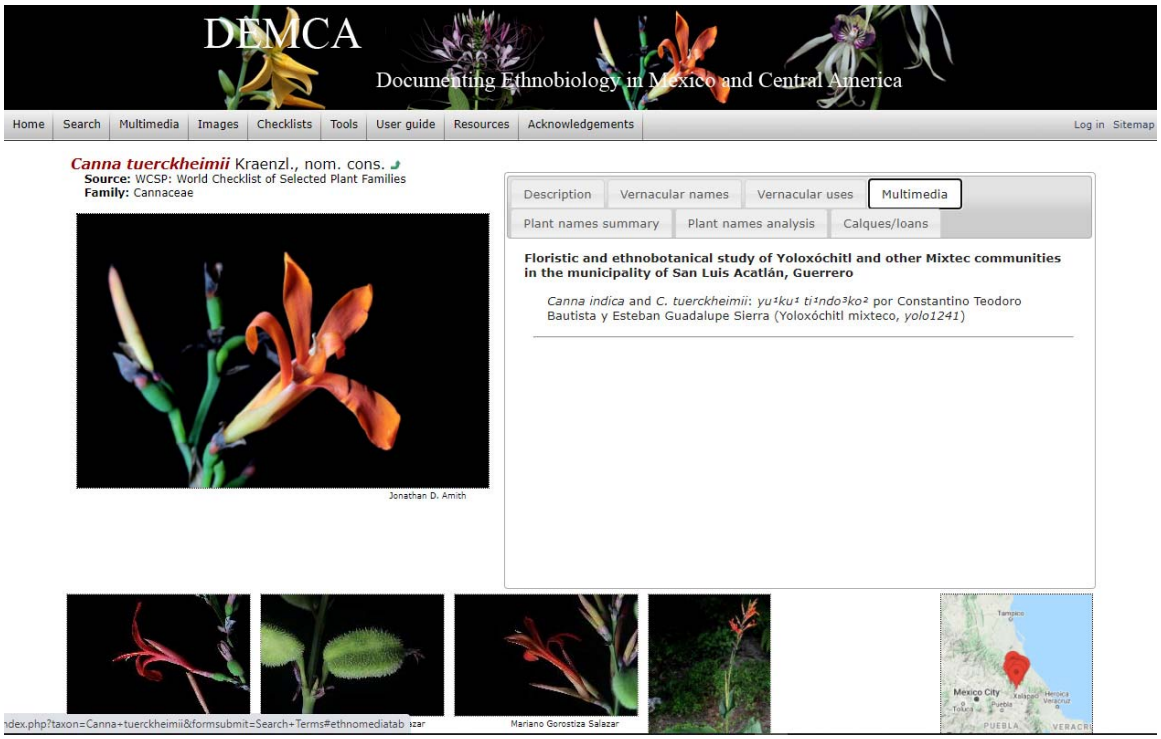


Figure 36e: Taxon page *Canna tuerckheimii*: Plant names summary

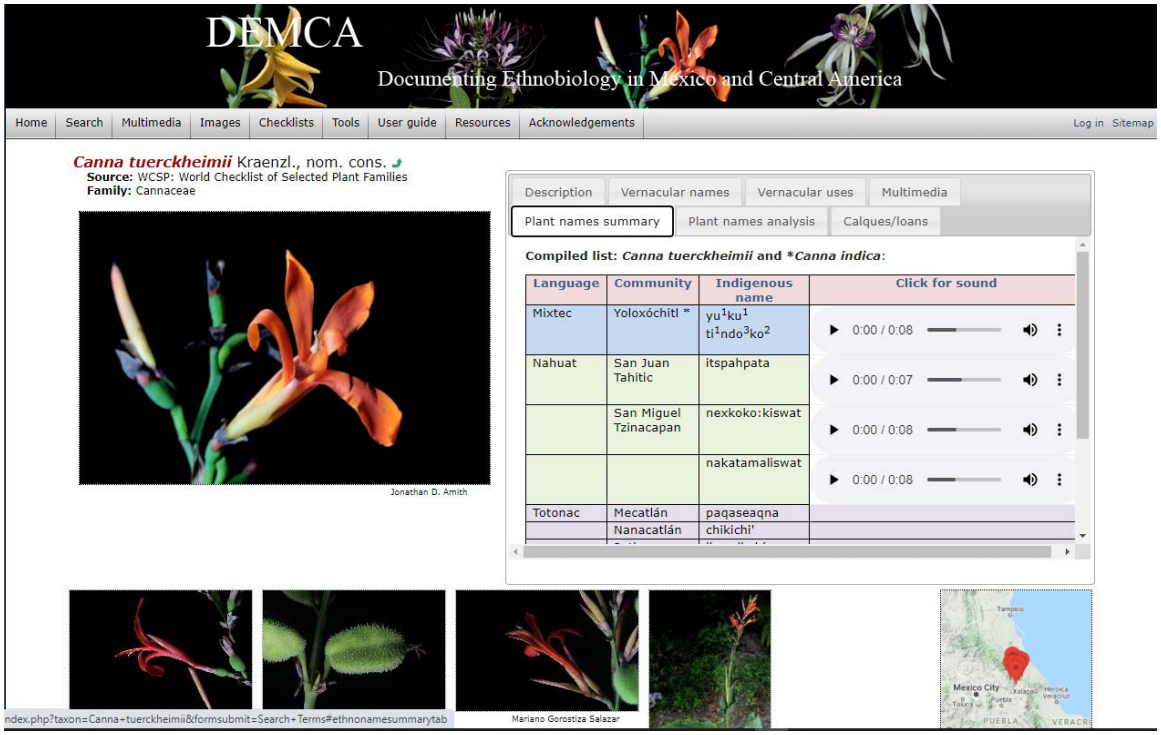


Figure 36f: Taxon page *Canna tuerckheimii*: Plant names analysis

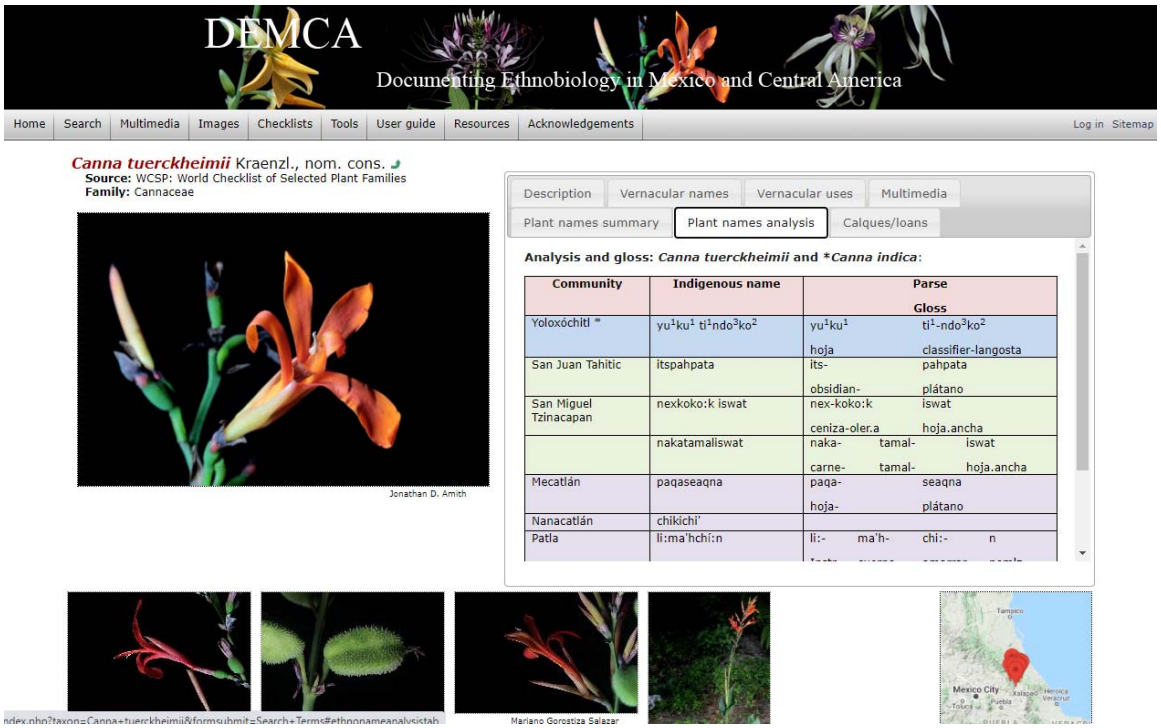
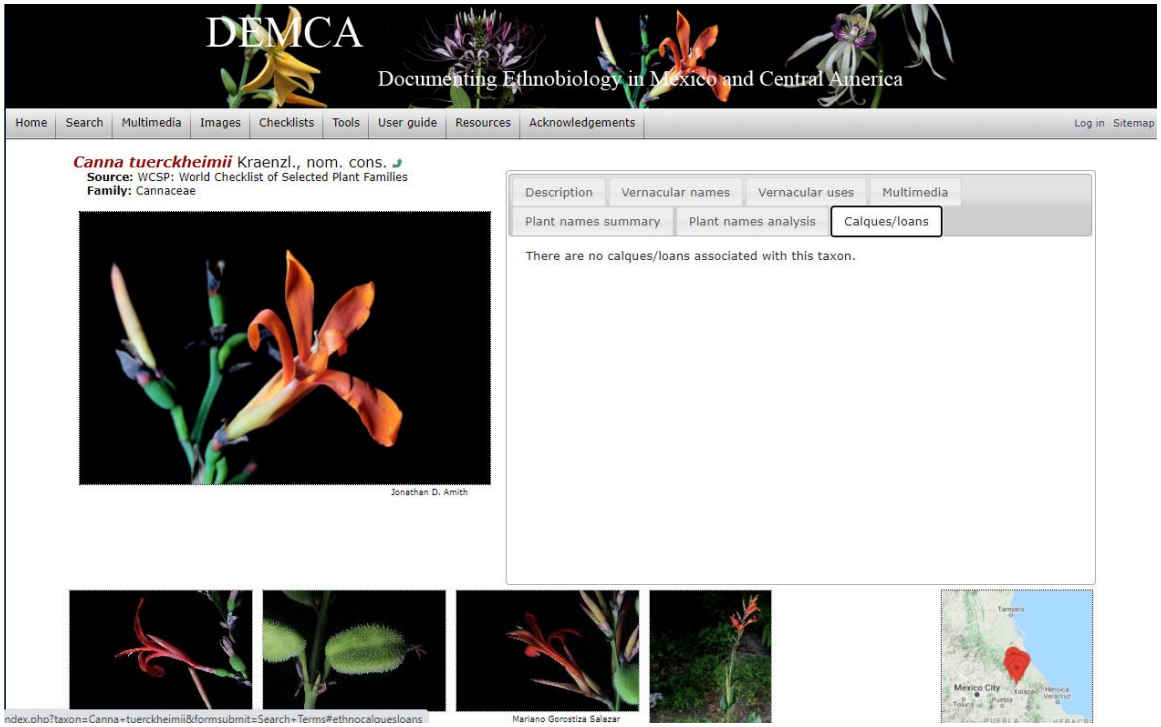


Figure 36g: Taxon page *Canna tuerckheimii*: Calques/loans

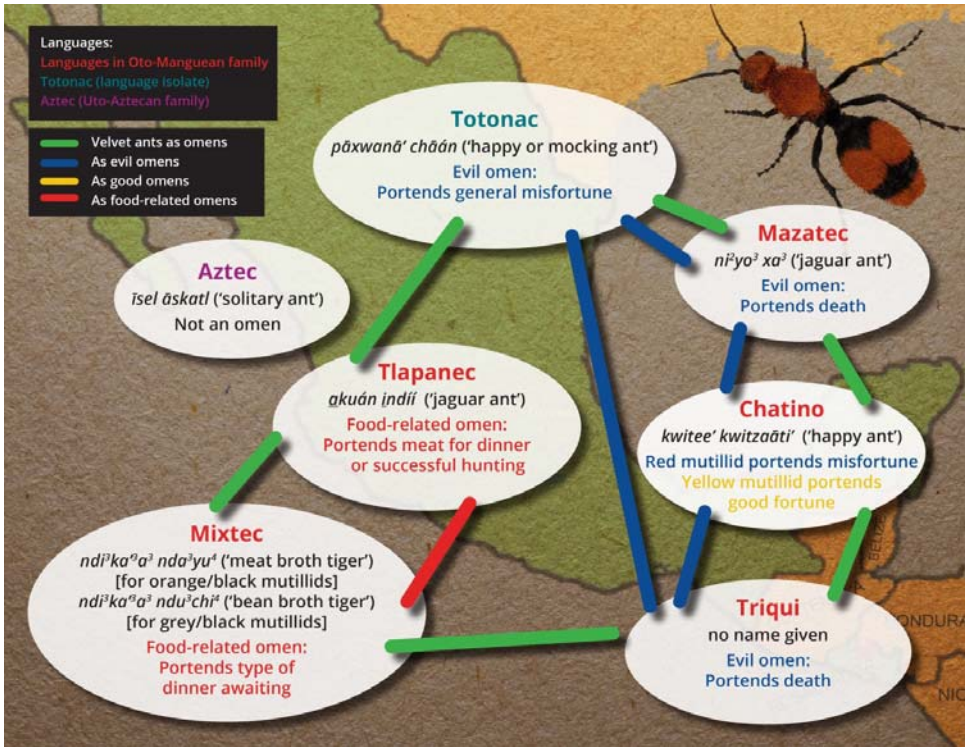


Taxon species page: Canna tuerckheimii. This page is representative of the pages that are automatically generated for every species in the portal although the content of each tab must be entered by hand. For those species for which there are no Indigenous names, classifications, or uses there will of course be no need for any of the tabs that deal with these sets of data. Indeed, most other Symbiota portals have very limiated information in the Taxon pages, often imited to a description and a list of web links that automatically remit the user to page about the same species in other resources (e.g., Encyclopedia of Life, Wikipedia, Encyclopedia of Life). The six extra tabs in DEMCA are a direct result of the ethnobiological focus of the present project. Each of the tabs will be discussed in turn. One tab, Public Comments, was in development when this project ended but will be added as soon as possible.

Description: This is a plant description that includes reference to its geographic range and basic characteristics of a morphological description. Most of the descriptions have been written by Kenia Velasco and Amith. At times experts in the relevant families will write descriptions. In almost all cases they have agreed to review the descriptions written by Velasco and Amith.
Vernacular names and Vernacular uses: At present these are summary accounts of all the data on these theme in the DEMCA portal. In the future, perhaps, some information from published sources (e.g., other ethnobiological studies) will be included but at present the goal is to integrate into one discussion, the ample information ingested into DEMCA.
Multimedia: As noted previously, there are two basic types of audio recordings. The first are field recordings associated with a single collection/observation event. The recording for collection #40202 was discussed in a previous section. The other type of recordings are made is a quiet setting and involve discussions among expert native speakers. Amith has over 500 of these transcribed and translated from Nahuat, Totonac, and Mixtec. Another 500 are in the process of being translated. These will be placed online as they are completed and additional funds become available.
Plant names summary: This tab gives access to a table of Indigenous names. The four columns of this table are: Language, Community, Indigenous name, and a link to play the recording of the plant name.

Plant names analysis: This is similar to the preceding but includes a parse (analysis) and gloss of the names, by community.
Calques/loans: This is a tab for all hypotheses about calques and loans across languages that have been found for the particular species. To date no calques or loans for C. tuerckheimii (e.g., Nahuat *nexkoko:k iswat*, among others) have been founded.
Public commentaries: A final wiki-like public commentaries tab will be integrated soon**.** The purpuse of this tab will be to offer registered users the opportunity to contribute or comment on the portal.

7. Conclusion: The project has completed its goals and has a functioning data portal presently at the University of Texas (demca.mesolex.org). Substantive content for the Taxon pages is presently being developed by Amith, Velasco, and native speaker colleagues. Funds to extend functionalities and add content are presently because sought from the NEH and NSF.

Addendum

The poster represented below demonstrates the types of contacts and sharing that can occur among languages and cultures. Velvet ants are frequently considered omens, sometimes good and sometimes bad. Note the shared concept (Totonac and Chatino) of mutillids being "happy ants" (happy because they bring misfortune and enjoy the bad luck of those so affected). The DEMCA portal will allow registered users to suggest links between words in one language or culture to those in another. Eventually (in its implementation stage) these links will be visible to users and mapped to show the geographic extent of contacts.