

## ebDB - Filling the gap for an International Ethnobotany Database

ebDB - Filling the gap for an International Ethnobotany Database

Steven Skoczen<sup>1</sup> & Rainer W Bussmann<sup>2</sup>

<sup>1</sup>  
23 Monkey Run Rd. Ithaca, NY 14850, email: [steven@quantumimagery.com](mailto:steven@quantumimagery.com);  
<sup>2</sup> 508 El Paso St., Austin, TX 78704, email:  
[rbussmann@natureandculture.org](mailto:rbussmann@natureandculture.org)

December 2006

Download at: <http://www.lyonia.org/downloadPDF.php?pdfID=2.492.1>

# ebDB - Filling the gap for an International Ethnobotany Database

## Abstract

Today, there are a few primary repositories for Ethnobotanical data, including the University of Riverside's Ethnobotany Database, Dr. Duke's Databases, and NAPRALERT, and several more specialized databases. Each of these systems fills a valuable role, but each also has shortcomings in accessibility and use in international applications. The International Ethnobotany Database (ebDB) is a new, public database that fills in the existing gaps in functionality, and provides a standardized, secure, independent, and non-commercial repository for ethnobotanical data. In this system, individual researchers maintain complete control over the access to their data, and can choose to keep it completely private, or securely share the information with selected colleagues. The database is fully multilingual, and supports translation, data entry, and viewing for any language. Most importantly, the database is open to any researcher to add new data, datasets, and languages. The ebDB is managed by the Foundation for Open EthnoBotany Research (FOER), an organization dedicated solely to managing the database, and providing an international repository for ethnobotanical data. As a result, it is insulated from institutional political turmoil and commercial interests, and provides a safe home for data for those concerned about bio-prospecting. Built over four years of field research in Peru, the ebDB contains a broad feature set, and is designed specifically for ethnobotanical research. It is fully multilingual, has a glossary, more than 20 categories of data, complete location information, strong searching, and data export features. Currently, the database houses data from Ecuador, Peru, Kenya and Hawai'i. It is the hope of this project that the ebDB becomes an international standard for storing ethnobotanical data.

### Background

There are several widely-known ethnobotanical databases already in existence. While each of these repositories has its own strength, none has become a standard for international research. The reason for this is quite simple: none of these databases allow researchers to add their own data. There are also several other weaknesses and strengths of each of these applications. To their credit, none but NAPRALERT were designed to be an international standard, so it is not surprising that they do not meet this need. We shall examine each of the largest in turn.

#### Dr. Duke's Phytochemical and Ethnobotanical databases

<http://www.ars-grin.gov/duke/> This database houses research compiled by Dr. Jim Duke of the Green Farmacy Garden and the United States Department of Agriculture (USDA). The USDA's Agricultural Research Service provides support for the database. Dr. Duke's database has several key strengths: a large number of plants, complete data on most database entries, and chemical information on a variety of plants. It also has several critical weaknesses, most notably the lack of public input, and limited support for a variety of ethnobotany data. The "EthnobotDB" itself only has fields for botanical description and use. While this makes the database an effective tool for disseminating Dr. Duke's work, it does nothing to assist other researchers to effectively detail the fruits of their research. With the USDA backing, Dr. Duke's database also raises a thorny question: if it were possible for researchers to add their data, to whom would it belong?

#### UC Riverside's Ethnobotany Database

<http://maya.ucr.edu/pril/ethnobotany/database/database.html>. This database provides a much more generalized approach to ethnobotany. The University of California, Riverside (UCR), provides support. UCR's database is well-designed in the breadth of fields and information types it covers. However, it too suffers shortcomings as a central, international database. First, it is a completely off-line tool. While this has advantages in portability, a non-internet based tool can not function effectively as a central storehouse for ethnobotanical data. Second, like Dr. Duke's database, UCR Ethnobotany Database is a closed system: we are unable to add our own data to a shared repository. Finally, the database interface is in English only. Importantly, the database is not freely available for download online. While the team at UCR has developed an excellent tool, it does not meet the needs of international researchers.

#### NAPRALERT <http://www.cas.org/ONLINE/DBSS/napralertss.html>

NAPRALERT is in many ways, the antithesis of what we believe an Ethnobotany database should be. To its favor, it does contain information on a wide variety of plants and their uses, but its use to researchers begins and ends there. Unlike the other databases, NAPRALERT is a pay-per-use system, with a cost at writing of 8.90USD per search. While the University of Illinois

at Chicago is entitled to fund its research, clearly this is no generalized storehouse for ethnobotanical knowledge. It's important to also note that NAPRALERT focuses on the phytochemical information for the plants in its database. While phytochemistry is an important science, with the concerns of bio-prospecting it is more responsible to keep phytochemical data separate from pure ethnobotanical data. Finally, like the aforementioned two databases, NAPRALERT is a closed, monolingual system.

#### Specialized Databases

In addition to the three mentioned above, there are also a handful of specialized ethnobotanical databases. Dan Moerman of the University of Michigan has an excellent Native American Ethnobotany database located at <http://herb.umd.umich.edu/> <http://herb.umd.umich.edu/>. Moerman's database references USDA plant descriptions, and provides preparation method, plant part, and use data. All told, it contains a staggering 44,000 items from 4,029 species. Fort Lewis College also provides an accessible database

<http://anthro.fortlewis.edu/ethnobotany/database.htm>, <http://anthro.fortlewis.edu/ethnobotany/database.htm> that focuses on the Southwestern American Indians. While smaller than Moerman's database, the Fort Lewis database still contains use and preparation information for 293 species. Both of these databases (and others) are valuable tools, but fall short in providing services to the larger research community. Once again, the missing keys of public input, international language support, and clean intellectual property rights are still wanting. It should of course be noted that the main reason these smaller databases do not fill this role is that they were never designed with that purpose in mind!

#### International Ethnobotany Database (ebDB)

The ebDB was created out of a need for an independent, public, open database that provided data security and support for many languages. It was completed in July of 2004, and is currently open to the public. The database is located at: <https://olorien.org/ebDB>. The database was developed with private funding, and the researchers who submitted it own the data it contains. The ebDB took shape over four years of field research in Peru. As a result, its feature set has been shaped by the real problems faced in field research, data storage, and retrieval. The result is a robust, scalable system that can easily accommodate data from anywhere in the world.

#### Database and data ownership

The ebDB is owned and managed by the Foundation for Open Ethno Botanical Research (FOER). FOER is an organization formed solely to maintain and ensure access to the database. At the time of writing, the Foundation is still being formed, but should be finalized by the end of 2006. Details will be available at <https://olorien.org/ebDB> as soon as they are finalized. Steven Skoczen, who will release the copyrights to FOER upon its formation, wrote the database itself. There are no other corporate or institutional interests in the creation or maintenance of the database. Data submitted to the ebDB remains under the copyright of the dataset's owner, and any relevant informants. FOER serves only as a storehouse for the information, and retains no publication rights, copyrights, or even access to submitted data. Dataset owners can specify global and user-by-user permissions for data access, and protect or share any data as appropriate.

#### Database features

What it is and What it's not:

First and foremost, the ebDB is an Ethnobotanical database. It is not a botany database, and does not have a focus on traditional botanical tools for plant identification. It is also not a phytochemical database, and as such, does not contain any chemical information for plants. The goal of the ebDB is simply to provide a complete solution for ethnobotanical data storage, and as a result, neither of those two tertiary ideas is ever planned for the database. With the intellectual property complications arising from phytochemical data, and completely different scope and feature set of a botany database, this seems a reasonable and good idea.

#### Features

The ebDB has several unique features that make it a particularly good choice as an international repository.

#### Multilingual

First, it is fully multilingual, supporting an unlimited number of languages for both the data itself and the web site. For data storage, the database can accept any language contained in the UTF-8 character set. This includes over 650 languages including English, German, Chinese, Russian, and less well-known languages like Oromo and Dzongkha. Leveraging internet standards, users can input their data in their native language without any extra software or changes - just as they'd type in any

word processor. After data has been entered, it can then be translated into any or several other languages. 1 The site itself can also be displayed in any language. Currently, users can view the site in English, Spanish, and Hawaiian. While data input is possible in any language, we currently need user support in translating the interface to languages other than the aforementioned. If you are interested in translating the ebDB interface into your language, please contact any of the authors. The ebDB also allows for the *translation* of all data in the database into any number of other languages. As entries are translated, users are able to view the data within the database in an ever-broadening range of languages. This also allows international researchers to easily provide their findings to the informants if the publication and informant's languages differ.

#### Datasets

The second key feature that differentiates the ebDB from other solutions is the idea of *datasets*. A *Dataset* is the term used in the ebDB nomenclature to denote a group of plant samples that share a common owner and copyright. The database uses *datasets* to define access permissions, and allow users to control the visibility and availability of the data they submit. When a dataset is created, users specify a name, description, and the language in which data will be entered. Dataset owners also define which users of the ebDB should be able to access their dataset, and what privileges those users should have. Owners can choose to allow users one or more of the following permissions: No Access, Read, Print, Add, Edit, Translate, Delete, and Own. In addition, owners can set a default set of permissions for everyone not specifically given permissions. Since these permissions can be changed at any time, there is a great deal of flexibility for dataset owners in security and publication.

#### Broad Ethnographic Information

An Ethnobotany database would be of little use if it did not provide fields for capturing all the relevant data for a particular sample. With that in mind, the ebDB contains fields for a wide variety of information, including the following:

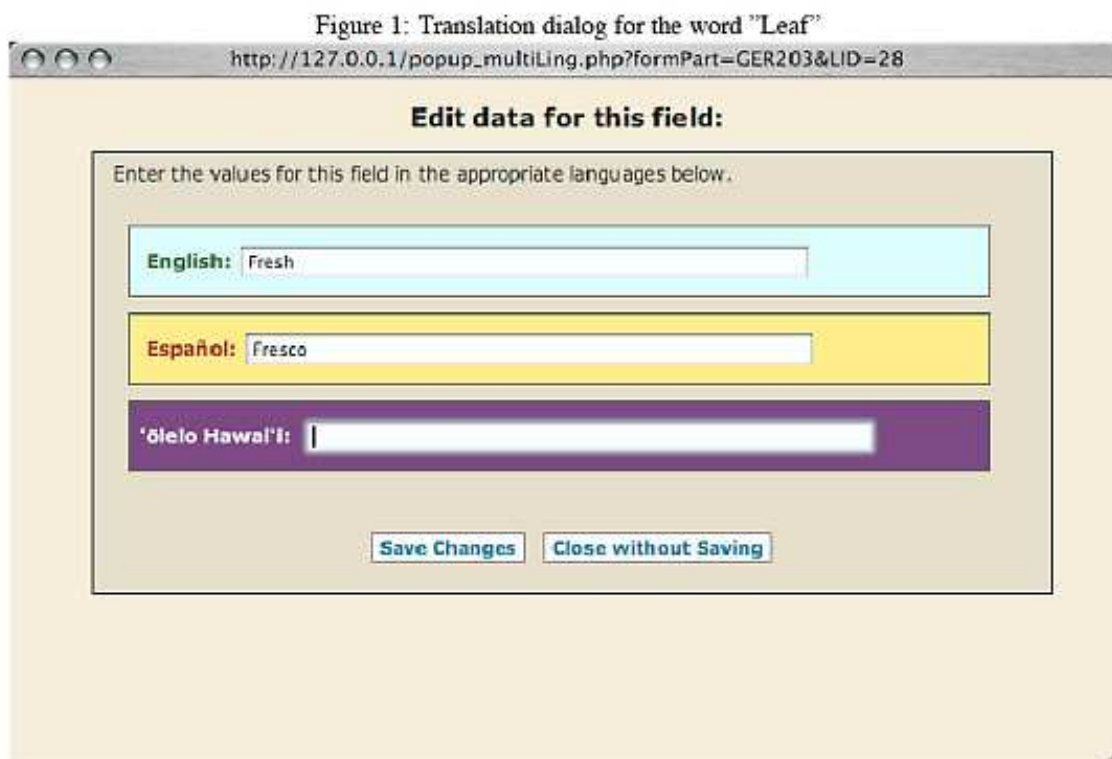


Figure 1: Translation dialog for the word "Leaf"

Figure 2: Dataset Management screen

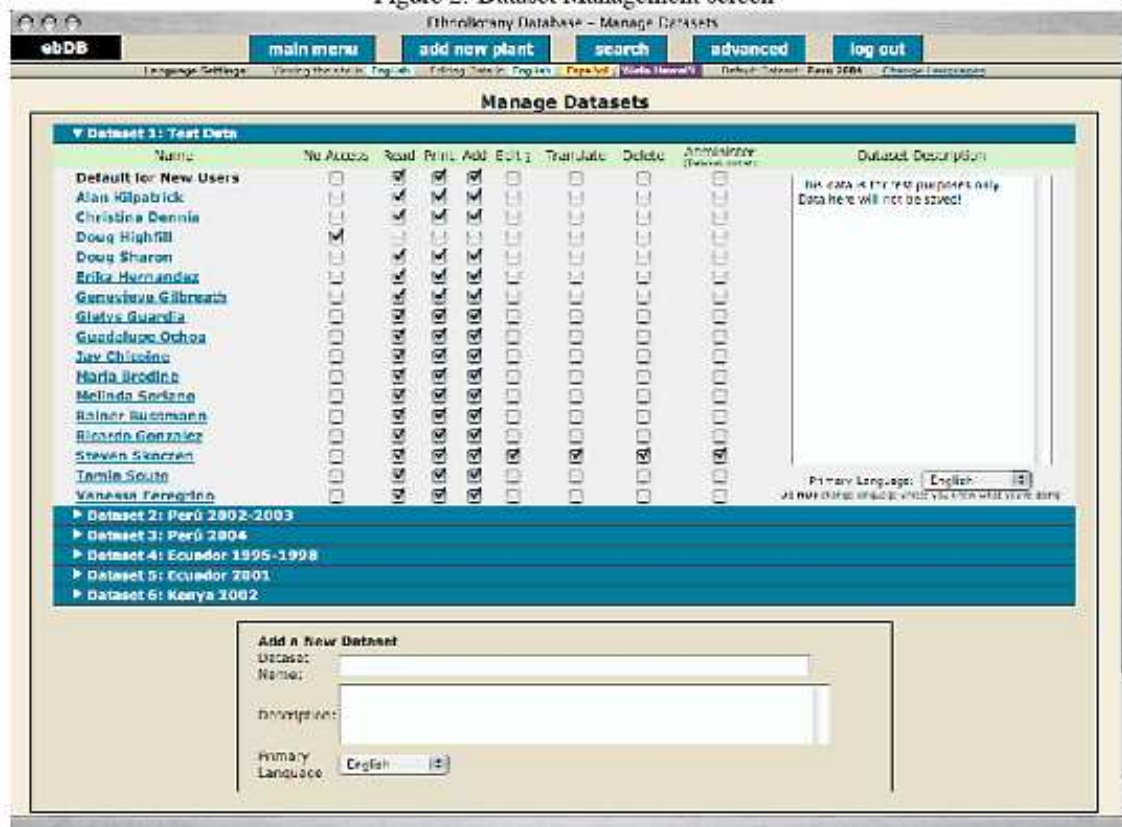


Figure 2: Dataset Management screen

- Plant Information
  - Kingdom
  - Family
  - Genus
  - Species
  - Description
  - Origin(s)
  - Other scientific names
- Sample Information
  - Picture(s) (An unlimited number of high-resolution pictures are allowed)
  - Collection Number (User-specified to match researchers' own numbering methods)
  - Collection Date
  - Informant (with an associated mini-biography about the informant)
  - Indigenous Name(s)
  - Location researcher collected the sample
  - Location(s) the informant collects or purchases the sample
  - Location(s) where the sample grows
  - Time of day (AM/PM/Both) the sample is generally collected
  - Temperature classification (Four Humors)
  - Uses:
    - Part(s) of plant used
    - Used fresh, dry, or both
    - General preparation method
    - Preparation details
    - General administration method
    - Administration details
    - Dosage

What thing(s) the sample (prepared with the above method) is used for

#### General use notes

As indicated by the list above, the Plant, Sample, Use, and Used For data are nested inside each other. Any plant species can contain an unlimited number of samples from any number of individual researchers, and any Sample can contain an unlimited number of Uses, etcetera.

#### Glossary

Also important in the ebDB's applicability is the glossary. For many ethnobotanic projects (or any cross-cultural project, for that matter), there are often words or phrases that are difficult to translate into another language. For example, in coastal Peru there is a condition called "*Mal Aire*". Translated literally into English, this would mean "bad air" or "bad wind". Unfortunately, neither definition has really much to do with *Mal Aire*, since its definition is grounded in Latin American culture (Calderon & Sharon). In these instances, it is desirable to be able to define a term more fully in a linked glossary. To that end, the ebDB has a built-in glossary that can have an entry attached to any translated word or phrase. Glossary definitions can be defined globally, per dataset, or per informant. For example, if two informants both specify that they prepare a particular plant as a tea, but have greatly differing views on how to make a proper tea, the glossary categorically captures those ethnographic differences for each informant.

#### Locations

Location information is important for many aspects of ethnobotany, from market flow analysis to retracing cultivation history. Consequentially, the ebDB has strong support for location data. Each sample can have an unlimited number of locations for the place it originates, the place the informant collected the sample, and the place the researcher collected the sample. Any location in the database can contain information for any or all of these fields:

- Continent
- Country
- Region
- City
- Location
- GPS Longitude and Latitude
- Location Notes

The variety and specificity of location information greatly improves the ease of analysis for location-based data. With detailed data in each of these categories, it is easy to determine trends from the continent down to the city. In addition, GPS-based data provides a precise long-term solution for location tracking. "Used for" categorization The ebDB also allows dataset owners to categorize what various plants are used for into broad categories. With this technique, the original ethnographic data is retained, but the researcher can easily analyze broad trends in what plants are used for. The ebDB provides a default set of categories that fit most ethnobotanical data, or dataset owners can specify their own. There is an easy-to-use interface for dataset owners to categorize the uses in their datasets.

**Figure 3: Location fields**

The screenshot shows a web browser window with the URL `http://127.0.0.1/popup_location.php?formPart=sample1InformantCollect&multiple=true`. The page title is "Location". Under the heading "Political Geography:", there are several dropdown menus: "Continent" (South America), "Country" (Peru), "Region/State/Province" (Lambayeque), and "City" (Morrope). Below these is a text input field for "Location (business name, etc):" containing "Dos Palos Chacra". The "GPS:" section has two text input fields: "GPS Longitude" and "GPS Latitude", each with a directional dropdown menu (-E/W- and -N/S- respectively). A "Notes:" section has a large text area. At the bottom, there are two buttons: "Save Changes" and "Close without Saving".

**Figure 3: Location fields**

#### Data export

As might be expected, a database is of limited use unless users can export relevant data for analysis. To that end, the ebDB supports the export of an entire datasets by the dataset owner. Datasets are exported as .csv (comma separated value) files that can be opened with most spreadsheet and statistical packages.

#### Field research

In keeping with its field-based development, the ebDB has printing features designed specifically for in the field research. The database has a page to "Print Field Sheets" that allows authorized users to enter collection numbers and print out pre-formatted data collection sheets that contain all the current data and pictures, and blank lines for any missing data. The ebDB also has a "Print Manuscript" page that allows authorized users to print a formatted manuscript draft of all the data in a particular dataset. Since the output is in HTML, it serves as a very good starting point for a book or other publication.

#### Standard tools

As should be expected, the ebDB also has all the standard tools and features of most databases. This includes internal data integrity checks, regular backups, a strong, flexible search capable of searching any field in the database, user management, and strong security. It is important to note that the ebDB can only be accessed over a secure HTTP connection (`https://`), so users with very old browsers will need to upgrade. The secure HTTP connection ensures that all data sent to the database is encrypted, and can not be read while it is being transmitted from the user's computer to the server.

Figure 4: Search Screen: any field in the database is searchable.

**Search for Plants**

---

**Step 1:** Choose whether you want to search for plants or samples.  
 Search by:  Plants  Sample  
[What does this mean? \(Show/Hide help\)](#)

---

**Step 2:** Select your search criteria. To add several items, use the [+ ] button below.  
 Look in (category) for things that (search method) have (this text) anywhere in them.  
 Genus is exactly equal to Asteraceae  
 Lived For contains pain  
 [+ ] Add another field [- ] Remove the last field

---

**Step 3:** Advanced options:  Show advanced options

---

**Step 4: Search**  
 Use the buttons below to search.

Figure 4: Search Screen: any field in the database is searchable.

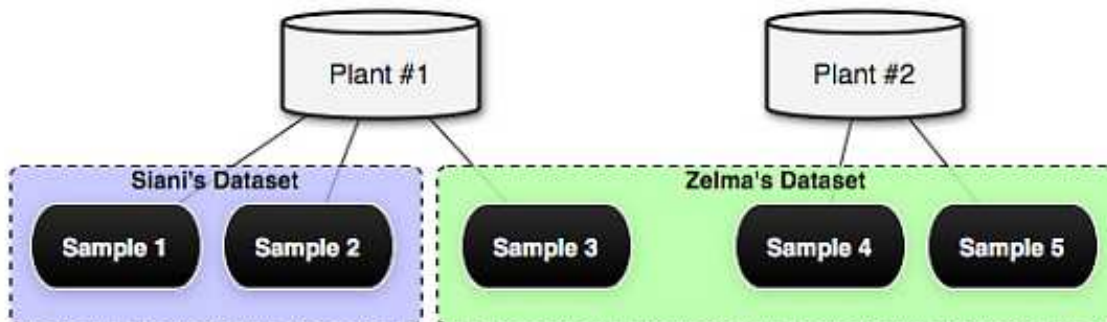
#### Database structure

In the database, the hierarchy begins with unique plant species. Each species can then contain an unlimited number of samples, in which specific ethnographic information is stored. Samples are grouped as *datasets*, where each dataset has an owner and specific set of copyrights. Figure 5 shows the overall hierarchy of data. As shown, by default users Siani and Zelma have access to only the plants in their datasets, even though Siani's "Sample 2" and Zelma's "Sample 3" are botanically the same plant. Dataset owners can completely control the level and type of access to the data in their dataset.

#### Technical details

The International Ethnobotany Database has been designed from the ground up to be reliable and scalable. As a result, it is built on the best of established, tested web technologies. The database itself is built in MySQL <http://www.mysql.org>, a freely available database developed by MySQL AB. MySQL has been established as a premiere database for large projects, both in reliability and speed. The site is built in PHP <http://www.php.net>, an open source web scripting language that draws its structure from Perl and C.

PHP has been proven to be reliable, is well documented, and is one of the most actively maintained projects on the internet. Finally, to host the site, the ebDB relies on Apache, the most widely-used and stable web sever available. Apache is also stable, fast, and open. As would also be expected from any critical database, regular backups are made of the ebDB both on and off-site.





**Figure 5: Datasets, Plants, and Samples: Each user has control of only their samples, even if another user has samples that are botanically the same plant.**

A unique solution for internationalization

In building the multilingual features of the ebDB, we ran into a common problem, and have developed a unique solution that works well. This problem has been the subject of much discussion in the internationalization and localization communities, so the result may be of some interest. A challenge arises when trying to create a database structure that supports multiple languages for any given field. Two solutions come quickly to mind, and have been covered in detail online. However, the consensus is that neither solution is truly a "fix". Both seem to be "hacks" to work around databases that are not inherently multilingual. In this project, we have come upon a solution that seems to be a real "fix", is elegant, and works well.

The first generally accepted solution is to add multiple fields for each field's language within one table.

Part of Plant Used Table			
PartID	Partname_en	Partname_es	Partname_??
1	leaf	hoja	??
2	apple	manzana	??

**Figure 6: The first generally accepted localization solution. New languages must be hard-coded as new columns into the table structure.**

While this idea clearly works, it becomes quite difficult to manage once more than a few languages are required. Adding new languages requires manual changes to the database *structure*, when language is best served as an *attribute* of a well-defined structure. This solution requires the database table to have explicit knowledge of the language of its data, an idea of poor design.

The second obvious idea is to use a delimiter to separate data within a given field.

Part of Plant Used Table	
PartID	Partname
1	leaf;hoja;????
2	apple;manzana;?????

**Figure 7: The second generally accepted localization solution. New languages are concatenated onto the list string with a delimiter.**

While this idea does a good job of keeping the database structure clean (the table structure is now agnostic to the languages it contains), it has two key flaws. First, adding another language is a difficult task that requires an update of every piece of internationalized data with the new delimiter sets. While this is not impossible, it is highly inefficient. Second, the operation of viewing and writing

data to these fields is fairly expensive. Each data read requires parsing of the data, and each writes a concatenation.

Both these solutions also have another flaw: neither can inherently take translation knowledge from one table and apply it to another table. For example, if a user has translated the English *apple* to the Spanish *manzana* in "Part of Plant Used" table, there is no good reason that an internationalized field with the data *apple* in another table shouldn't be able to take advantage of that knowledge.

A new solution: Internationalized Reference Tables. The solution we've devised produces an elegant database, is scalable, and when used in a database with good indexing, is exceptionally fast. Three tables are used: the data table and two reference tables for language and localized text. The result can be seen in figure 8.

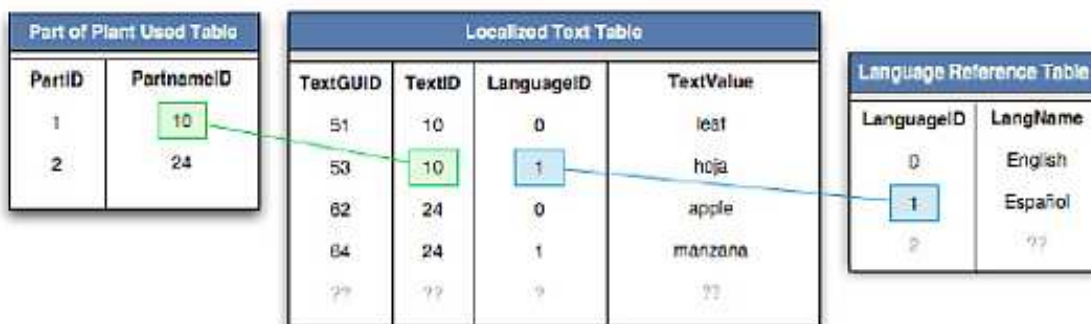


Figure 8: The localization scheme used in the ebDB. The database is completely agnostic to the language of any given data point.

At first glance, this schema seems more complex. However, looking closer, it is apparent that the database structure accommodates any number of languages without a structural change, and without any real hits in performance. The system does require a procedure to link the tables and languages as relevant to the context, but since this is a consistent operation, a generalized function with low performance cost can be easily created. Perhaps more importantly, the reference table solution above allows other tables in the database to take advantage of the "translations" offered by the localized text table. For instance, with the data listed in the figure, any future entries by any user of the English *apple* are automatically translated to the Spanish *manzana* without any work (or knowledge of Spanish) on the part of the user. Overall, this schema produces a system that can easily support an unlimited number of languages, leaves the data tables agnostic to the language of the data they contain, and allows for translation information to be applied across tables. For many applications that require internationalized data, this would seem an excellent solution.

#### Database Access

The ebDB is open for any member of the public to use. As discussed above, access to data is limited by dataset, where the dataset owners have complete control of their data. Consequently, any member of the public with an account is authorized to create their own dataset and add ethnobotanical data for their projects.

To obtain an account, email your full name and desired username to [accounts@ebdb.org](mailto:accounts@ebdb.org). Also specify if you would like an account with dataset creation privileges. Please note that though accounts are freely available, an account is needed to log into the database. No member of the public can browse the database or create datasets without an account.

How to use it, and how it should be used

How to use it

To access the database, simply visit <http://www.olorien.org/ebDB>. The application has been built using web standards, so a recent, standards-compatible browser is preferred. Mozilla Firefox, Konqueror, Safari, and Opera are all good choices. Microsoft Internet Explorer does not correctly support web standards, so one of the other browsers should be used if available.

For information on how to use the database itself, the reader is referred to the official documentation:

- The complete online manual located at <http://ebdb.org/manual> <http://olorien.org/ebDB/manual>.
- A PDF can also be downloaded from <http://ebdb.org/manual/manual.pdf>

<http://olorien.org/ebDB/manual/manual.pdf>.

How it should be used

A researcher's data is theirs to control, but sharing helps the larger ethnobotany community. Depending on the specifics of a given project, researchers are encouraged to share their data with selected colleagues, or possibly even all users of the database. The FOER is simply another user on in the database; you can choose to share your data with the foundation or keep it private, just as you would with any other colleague. In general, we suggest that dataset owners provide read-only access to all users, and reserve other access permissions to specified, trusted users. This allows all database users to share the wealth of knowledge across the variety of projects, and gives owners the greatest flexibility. It bears repeating that the database is not viewable by anyone without an account, and that ultimately, access control for a given dataset resides exclusively with its owner.

Future Research

While the ebDB as it currently exists is an excellent solution, there is always room for improvement. Our plans for the future include:

- Reaching out to ethnobotanical researchers to make them aware of the International Ethnobotany Database, and how it can work and help with their projects.
- Providing a generic framework for dataset owners to add their own cultural or project-specific fields.
- Continual improvements in database speed, responsiveness, and search functions.
- Possibly integrating the scientific name descriptions with the Standard of the International Taxonomic Databases Working Group.
- Translation of the interface into as many languages as possible. We will need the support of users to make this happen, so please contact any of the authors if you are interested in help translate text into your language.

## Conclusions

While several ethnobotany databases exist, none are adequate solutions for international work, and none allow researchers to add their own data. A need clearly exists for a centralized, independent database that facilitates ethnobotanical research. The International Ethnobotany Database is a proposed solution to this need. It is supported by an independent organization dedicated to maintaining it as a neutral, secure repository, free from commercial or political influences. The ebDB provides complete control to researchers over their own data, and natively supports almost every language used in the world today. It provides a system for user translation of all data into any language, and leverages those translations for future data entry. In addition, it has a set of features directed specifically for ethnobotany research, including a glossary of culturally-specific terms, detailed location data on the sample's origin and market flow, and field sheet printouts to assist research in the field. With its feature set, organizational backing, and robust function, the ebDB is a solution of use to anyone doing ethnobotany anywhere in the world.

## References

Dan Moerman's Native American Ethnobotany Database, <http://wwwherb.umd.umich.edu/>

Dr. Duke's Phytochemical and Ethnobotanical databases, <http://www.ars-grin.gov/duke/>

Calderón E & D Sharon. 1978. *Terapia de la Curandía*, Lima.

Fort Lewis College's Ethnoecology Database, <http://wwwanthro.fortlewis.edu/ethnobotany/database.htm>

NAPRALERT, <http://www.cas.org/ONLINE/DBSS/napralertss.html>

UCR's Ethnobotany Database [http://www.maya.ucr.edu/pril/ethnobotanydatabase/ database.html](http://www.maya.ucr.edu/pril/ethnobotanydatabase/database.html)

Unicode Language List, <http://www.unicode.org/onlinedat/languages-scripts.html>

US Department of Agriculture, Agricultural Research Service, <http://www.ars.usda.gov>