# Database Integration

The integration of the ever-growing number of biological databases establishes the framework for the development of a very flexible and information rich virtual flora. Georeferenced specimen data offers the opportunity of creating custom checklists by comparing place name or map defined shape files with collection data. Checklists with a greater accuracy but less flexibility can also be obtained from a database of pre-compiled checklists. An image database storing metadata, urls, and taxonomic links for photos and images allows for an unlimited number of images to be accessed for any given taxon. Descriptive information for each species parsed into a relational database structure increases flexibility of how the data can be utilized. Coded descriptive information can be used for interactive identification as well as generating natural language descriptions in various formats or languages.

Integrating these databases through a Taxonomic Thesaurus allows for data to be searchable by accepted name, synonyms, or common names, and even more importantly overcomes the problems of integrating databases that differ in taxonomic option. As long as the Taxonomic Thesaurus is up-to-date, so will the taxonomic output of all the databases filtered through it.

Using a web interface to access such a network of databases via a taxonomic thesaurus allows for data presentation and search capabilities that rival any printed flora. It represents a flora whose pages could be rearranged to match any locality within its greater scope and character or taxonomic grouping defined by the user. It is a flora where the keys and descriptions are customized to fit the novice as well as the professional. Furthermore, output and tools are able to be customized to match various language with relatively little trouble. Finally, since the databases are 'live', there is no final edition and data can be revised and updated at any point in time. A good integration structure represents a flora with a solid foundation that is continuously built upon.
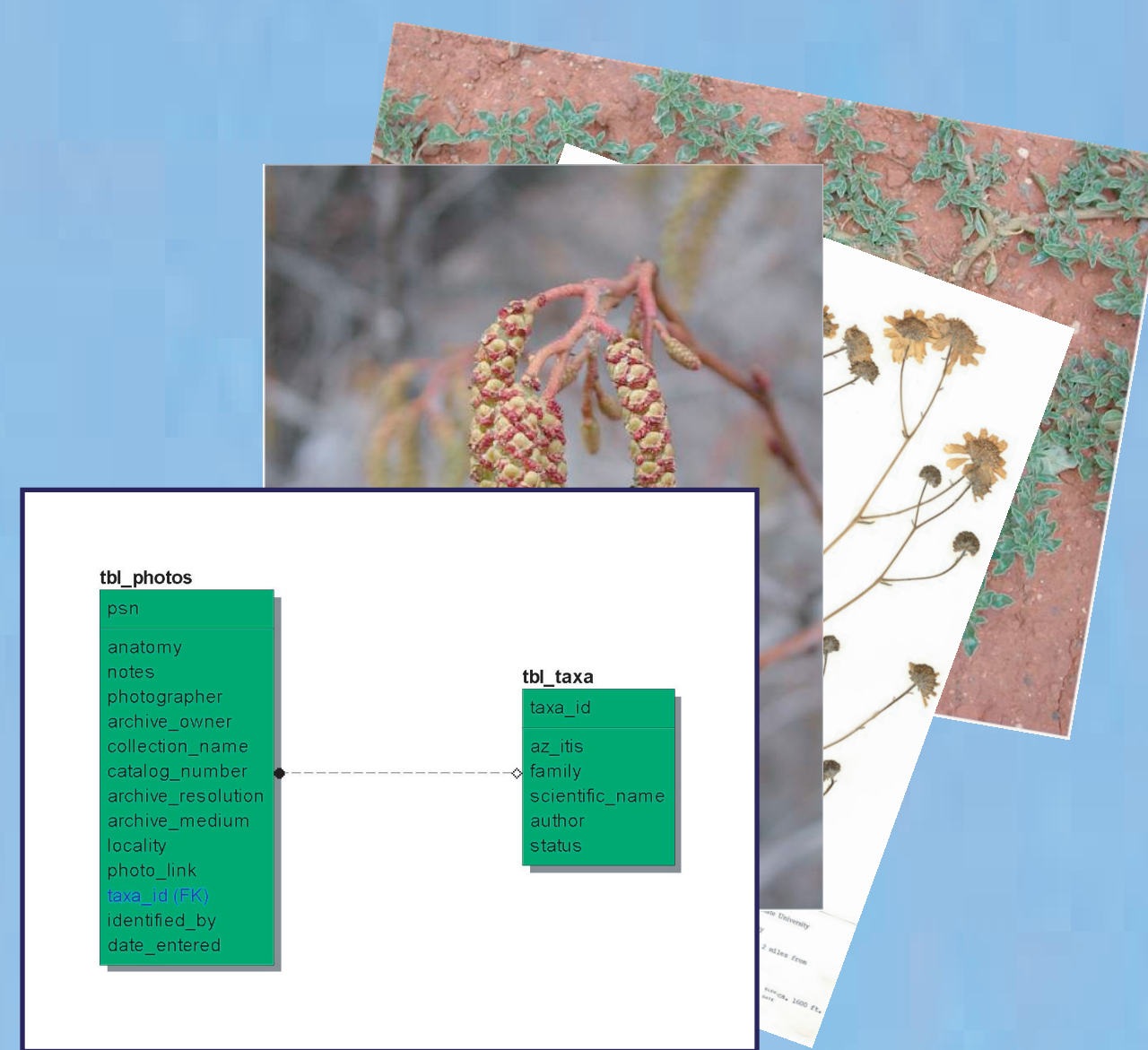
IMAGES

The image database stores metadata and URL addresses to JPEG images stored on a web server. Each image on the web server is linked to one or more taxonomic records within the taxonomic thesaurus. This allows for web display of a theoretically unlimited number of images per taxon.
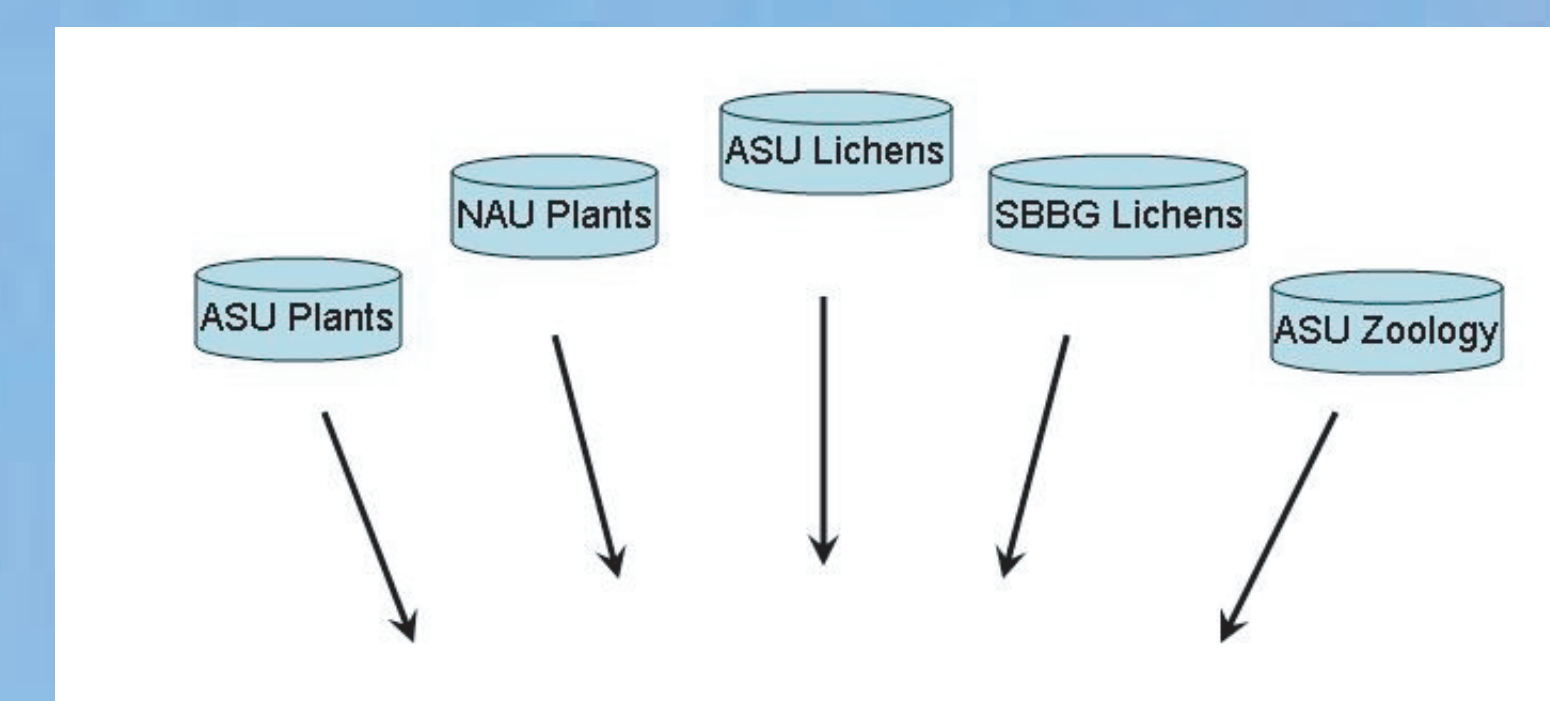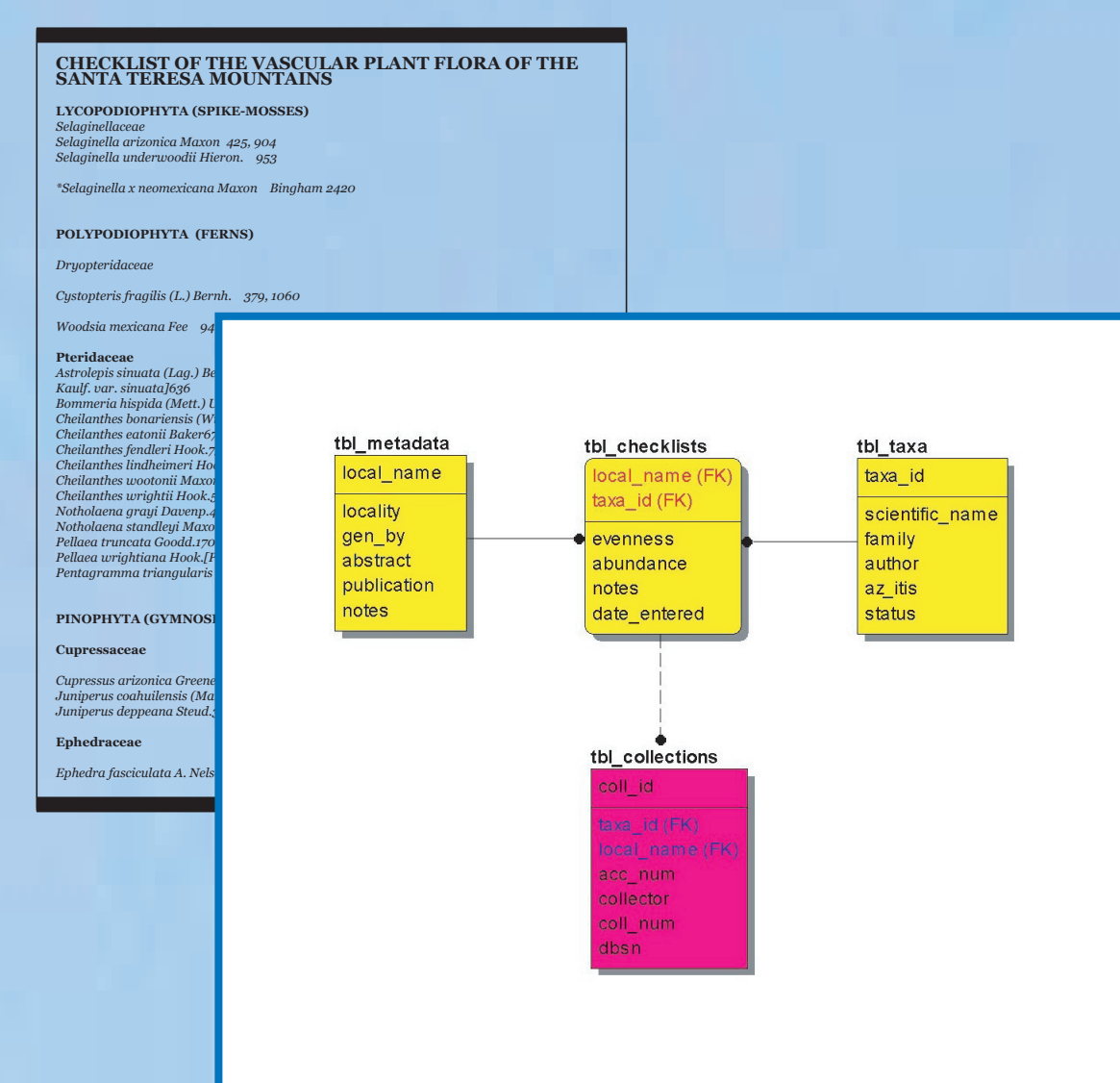
COLLECTIONS

Well documented biological collections are a rich source of geographical, morphological, temporal, and ecological information. The transfer of this data into a relational database structure brings much of this data out of the cabinets giving it a greater use potential.
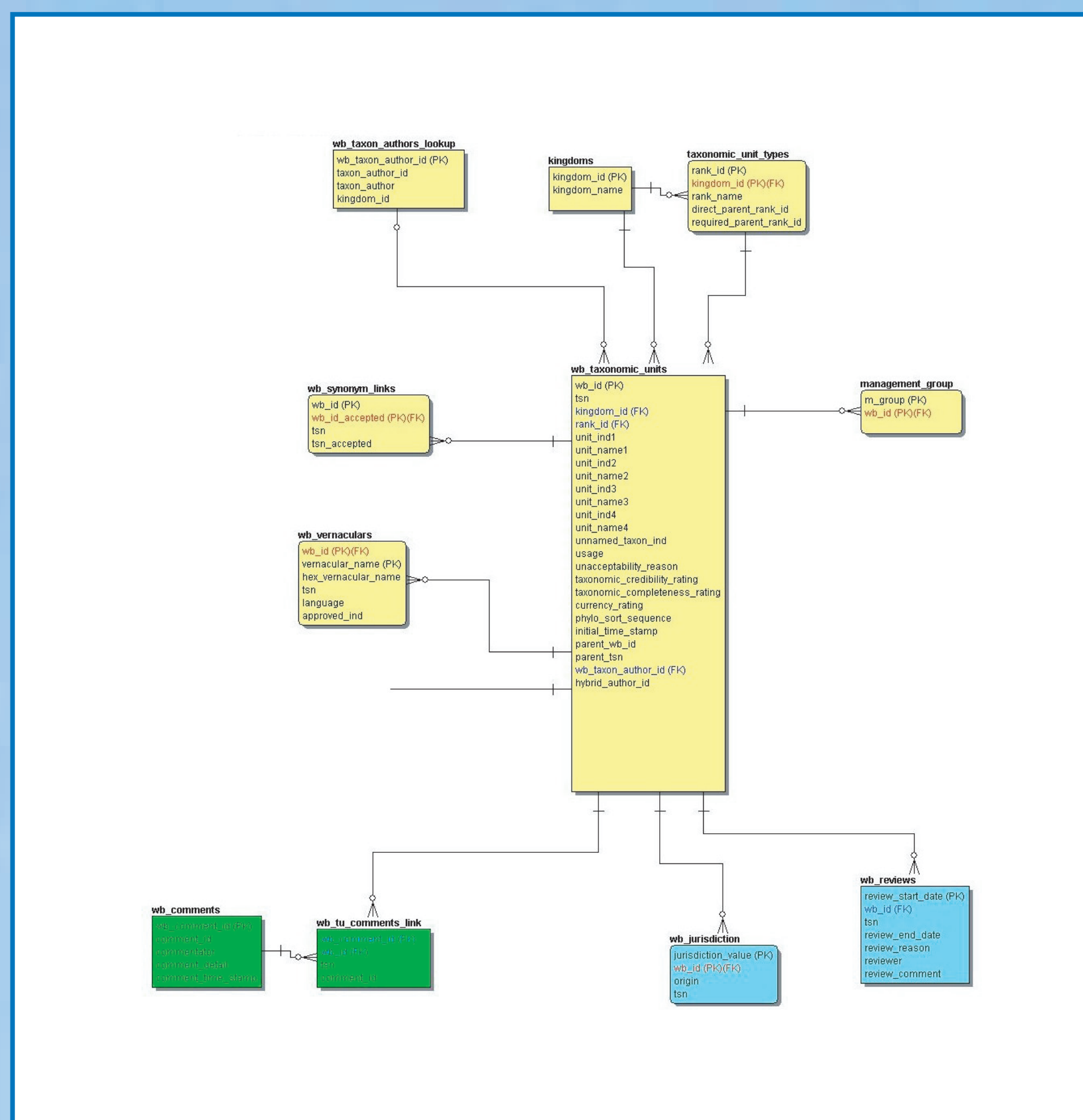
Geo-referenced collections makes possible the generatation of custom checklists based strictly on documented specimens. Once an area is defined using either place names in collaboration with a geographical thesaurus or a mapping wizard, the generated shape file can be compared with collection files. One of the limitations is that collections need to be complete, correctly identified and geo-referenced for the production of complete, accurate checklists.

For a more complete representation of Arizona, there is a need to integrate the various Arizona collection databases, which each differ in degree of coverage. Network access has permitted integrated access without removing the physical databases from the owners. Leaving control of data-entry, maintenance, and access with the owners better insures a quality database.
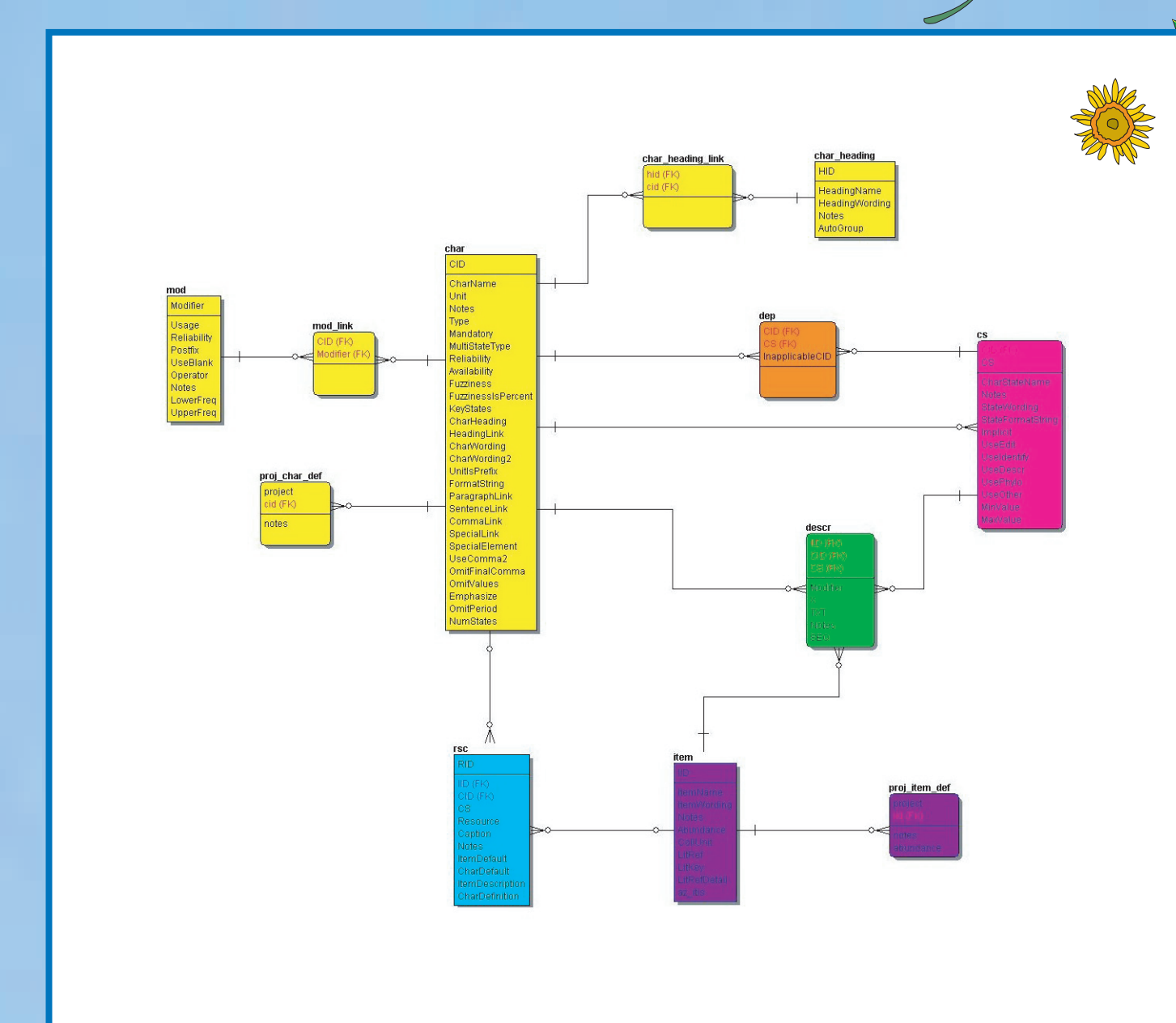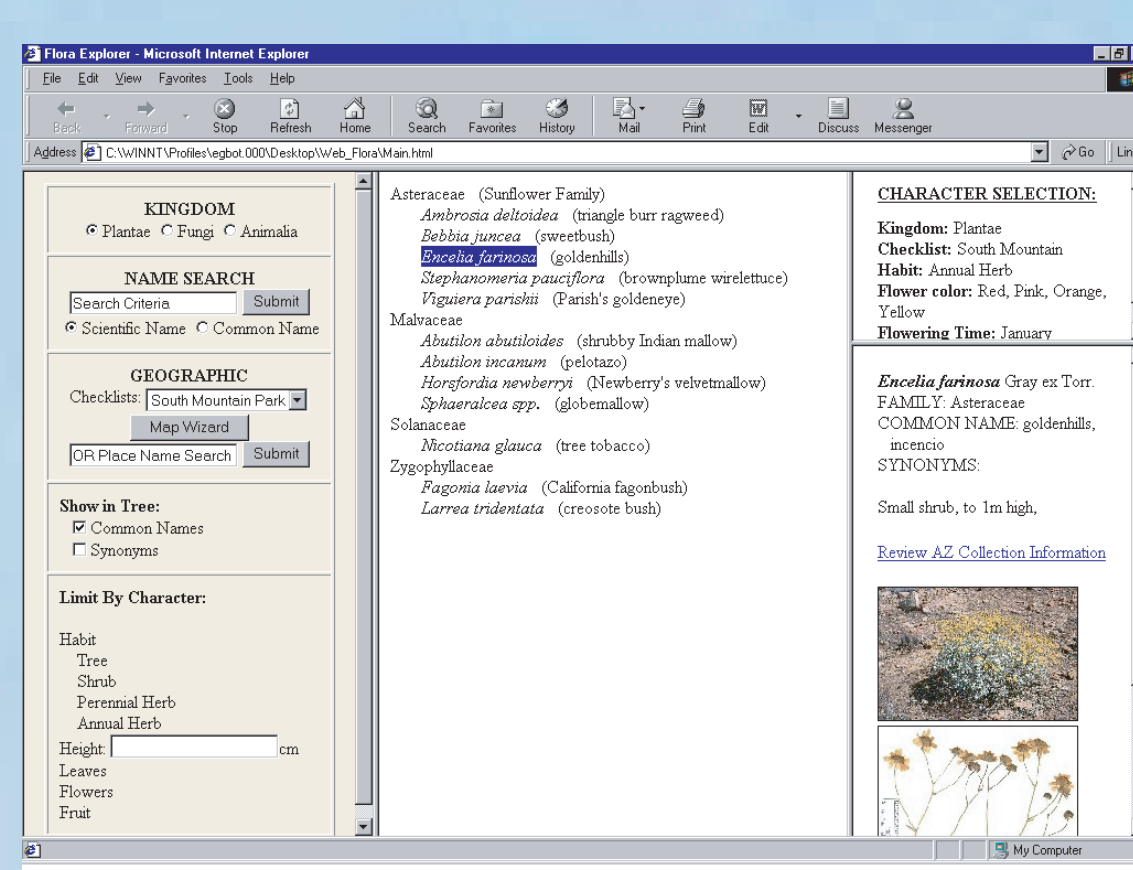
CHARACTERS

Descriptive information stored in a relational structure allows descriptive data to be used for the production of both custom keys and natural language descriptions. This database structure was initially based on DeltaAccess, a Microsoft Access 97 application. DeltaAccess is a relational interface to DELTA, a widely accepted standard for coding descriptive information. A slightly modified version of this database structure has been established on SQL Server 7.0 and the DeltaAccess application was configured to work as a front-end for data-entry and maintenance.

AZ TAXONOMIC THESAURUS

The USDA Integrated Taxonomic Information System (ITIS) served as the data and structure base for the AZ Taxonomic Thesaurus database. The database was then augmented with uploads from checklists extracted from the ASU Lichen Herbarium, ASU Vascular Plant Herbarium, ASU Zoology Department, and various professionals within the Arizona biological community giving the database a greater Arizona flavor of taxonomic option. Internal data links allow for database integration and search capabilities by accepted name, synonyms, and common names. A recursive hierarchical link in the central table allows for a parent-child relationship permitting capabilities of querying by taxonomic group (ie: all members of the Asteraceae family).

CHECKLISTS

Previously established checklists complied from floras, research, or other credible observations are stored in a simple relational database. In theory, such a database should not be necessary, for documented checklists of a higher caliber could be generated from geo-referenced collection data. Though to strictly rely on this, collections need to be significant in size, accurately identified, and geo-referenced. In time, collections will allow this but for now, some of the more complete and accurate checklists are those that are precompiled and edited by credible botanists.

Custom Floras, Guides, etc

Once a checklist is generated that fits the locality, taxonomic, and characteristic interests of the user, output can be generate to produce research datasets or hardcopies for use in the field. Printer output can range from a simple checklist to a more complete guide/flora with taxonomic data, descriptions, and pictures. This database integration would allow for the production of guides that could carry titles such as Flora of Northern Phoenix, Trees and Shrubs of the Verde Valley, or Spring Flowering Annual Herbs of the Lower Grand Canyon.

Problems / Difficulties

Output is only as good as the data stored. This stresses the importance of establishing quality database structures and techniques to ensures data quality and minimal error.

Collections: Some localities are poorly sampled. Furthermore, misidentified species are not uncommon in herbaria. This means that custom checklists generated from place names or a map wizard can be significantly incomplete or inaccurate. Mapping of geo-referenced collection data will make many of the errors and the poorly sampled areas standout. In time, this will aid in the increase of data quality.

At this time, the taxonomic thesaurus only allows for a single taxonomy to be presented which may open a door to conflict with the integration of collection databases with diverse taxonomic options. It is important to keep in mind that the main purpose of the Taxonomic Thesaurus is to establish relationships rather than present a taxonomic option.

Presently, a web-based flora in not the easiest thing to bring out to the field where it is the most useful. The abilities of generating and outputting a printed flora or guide customized to ones preferences in format style, taxonomic grouping, and/or character grouping overcomes this. Furthermore, practical web access in the field seems near at hand given the speed that wireless technology is growing
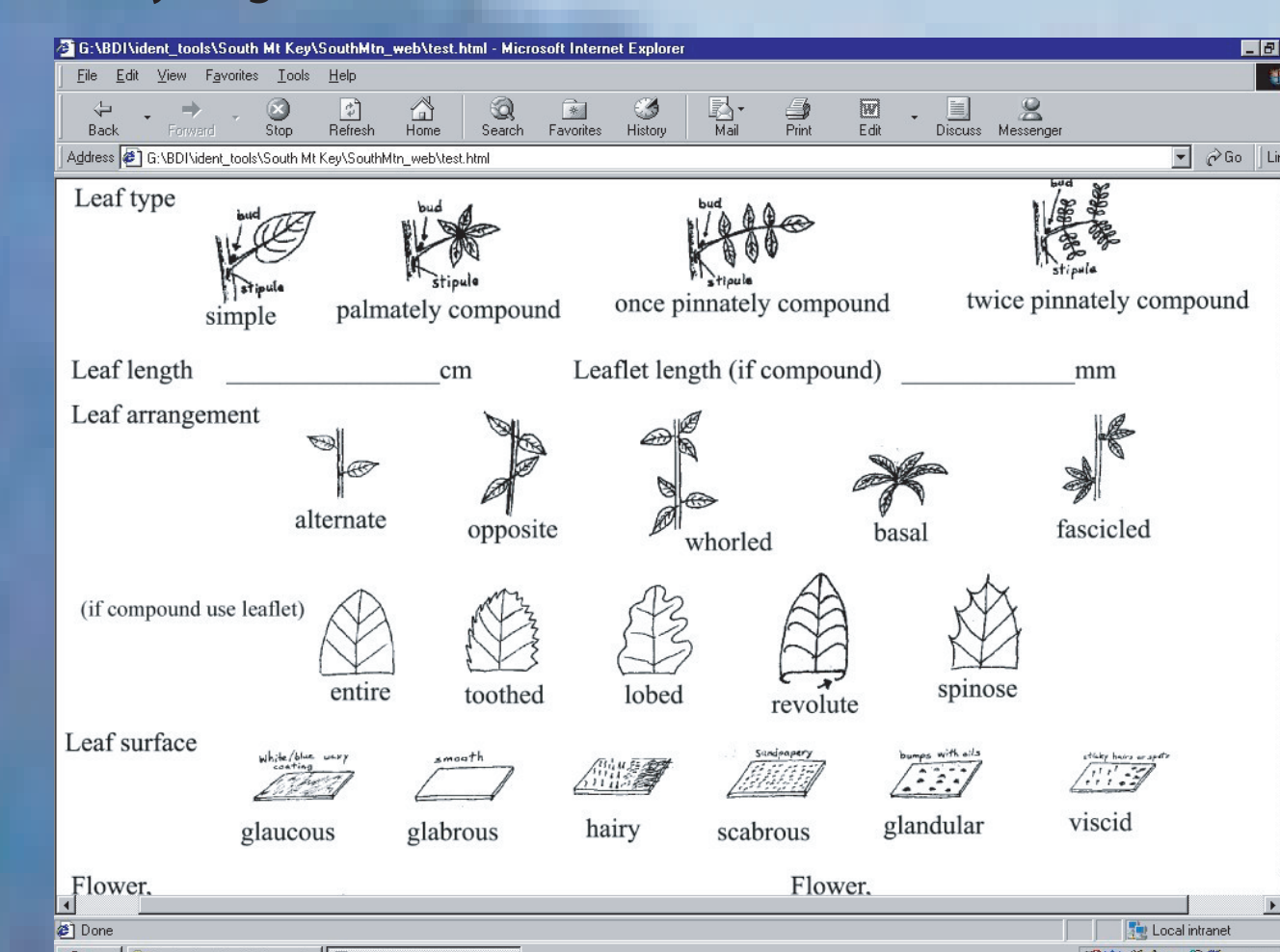
Flora Explora

The main purpose of this web interface is to generate a species checklist based on locality, taxonomic, and character information determined by the user. In the above example, the user has generated a list of annual herbs occurring in South Mountain Park with red, pink, orange, or yellow flowers that open in January. The user can now browse the checklist and access taxonomic, descriptive, and collection data along with photos for each species, or the user can export the checklist to an interactive key interface or more comprehensive output for research or field work.
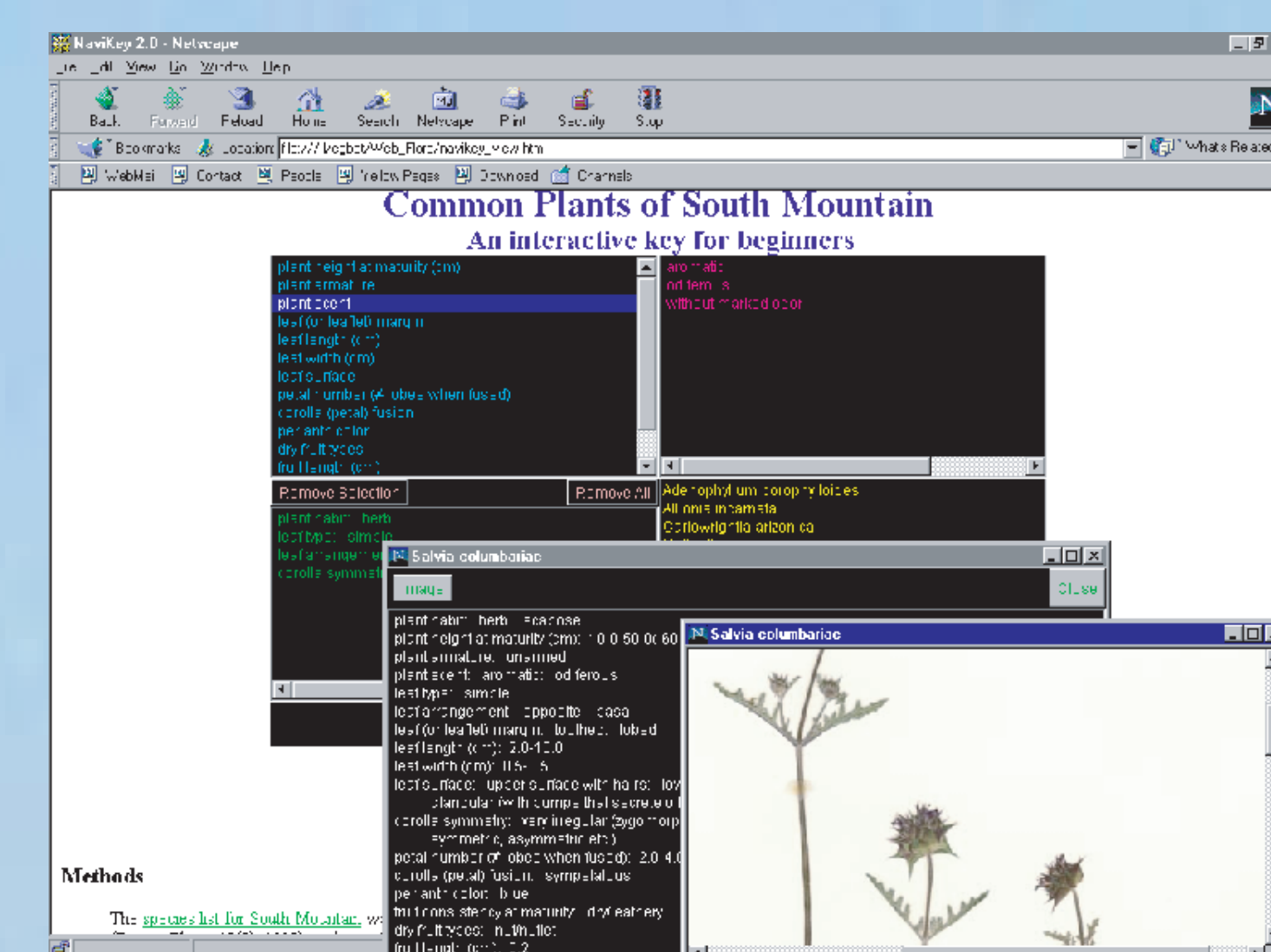
Web-based Interactive Keys

Supplied with a checklist generated from the Flora Explora, various interactive web-based identification tools, working in conjunction with the character database, aid the user to match an unknown to a scientific or common name. Identifications can then be verified with images or descriptions. The text based Navikey (above) may work well with participants who have a solid grasp of botanical terminology while a graphical key might better suit the novice.

Virtual Floras

Edward Gilbert[1,2], Jeremy Buegge[1], Peter McCartney[2], Corinna Gries[1,2], and Leslie Landrum[1].
[1]Department of Plant Biology and [2]Center for Environmental Studies

Interactive Web Key (Navikey)

Objectives:
1. Design and construct an interactive, web-based key to the plants of South Mountain Park for a general audience.

2. It will allow users to input what they know (rather than being constrained to specific characteristics typical of traditional dichotomous keys).

3. Plant names should be linked to descriptions and images.

Methods:

1. Create DeltaAccess database of characters & character states

2. Record specimen data and enter into database

3. Scan herbarium specimens

4. Link database records to images and NaviKey applet.

Results: Comparison of interactive key and traditional, dichotomous key to identify a plant.

Using the Interactive Key to South Mountain plants

Use worksheet — Take pictures — Take field notes

http://cochise.asu.edu/bdi/navikey/intro.htm

Select type of plant to identify

1. Character choice
2. Character state choice
3. Character and character states selected
4. List of possible species

Select known character state information about the plant until the list of possible species is sufficiently narrowed, then click on species names to view descriptions and images to confirm your identification.

All plants have descriptions.

All plants have a scanned herbarium specimen

Most plants have live images

Plant observed at South Mountain Park

Using traditional, dichotomous key (Arizona Flora)

Collect specimens (if allowed) — Take pictures — Take field notes

Go through the key in Arizona Flora to determine Phylum, then Class, then Family, then Genus, then species of the plant.

To key out *Olneya tesota*, from start to finish in Arizona Flora, you would need to know...

1. Stigma present; ovules and seeds in a closed cavity
2. Cotyledons usually 2; flower parts commonly not in threes; etc. (to page 626)
3. Petals united well above the base
4. Corolla distinctly irregular
5. Ovary superior
6. Stamens more or less united by the filaments; corolla not strongly sympetalous
7. Leaves compound; stamens usually 10. Petals usually 5; fruit a 2-valved or indehiscent pod; etc. (to page 395)
8. Corolla imbricate in bud; more or less irregular, the petals unlike in size or shape or both
9. Uppermost petal external in bud; the corolla commonly very irregular, with the 2 lowest petals often more or less united; etc. (to page 409)
10. Filaments, or 9 of them, united, at least near the base or if distinct, then attached at the base to the calyx, and the corolla wanting
11. Anthers not sharply differentiated, all of approximately the same size or if sharply differentiated, then the pods prickly, etc else the inflorescence conspicuously bracted; the corolla yellow, and the pods jointed or ending in a hook
12. Rachis of the leaf not ending in a tendril
13. Fruit not segmented, but sometimes constricted between the seeds
14. Leaves not glandular-punctate
15. Margins of the leaflets entire
16. Plants shrubby or arborescent
17. Leaves multifoliolate; corolla much less than 5 cm. long, not scarlet; banner petal not strongly keeled; etc.
18. Calyx without bracteoles; corolla purple, pink, or white; outer wall of the pod not separating from the inner wall
19. Herbage without dolabriform hairs; pods elongate, several-seeded; etc.
20. Inflorescence not glandular; leaflets not cuspidate. Plants small trees; corolla purple and white, about 12mm long

This is a fairly distinctive plant, but it is tedious to identify using a traditional dichotomous key. This is because technical characters are used, which are necessary to differentiate the 3438 species in this book.

All plants have a genus and a species description, but they are extremely variable.

Very few plant have pictures. These are at the beginning of the book in black and white.