FLORISTIC INFORMATION SYSTEM FOR CALIFORNIA TRACHEOPHYTES (FISCT)

THOMAS J. ROSATTI and THOMAS DUNCAN
Jepson Herbarium and University Herbarium
University of California, Berkeley, CA 94720

ABSTRACT

In this paper a plan is outlined for the long-term development of an electronic, network-accessible Floristic Information System for California Tracheophytes (FISCT). FISCT is both an extension and an application of the Specimen Management System for California Herbaria (SMASCH), and will serve as an interface between SMASCH and future, electronic versions of *The Jepson Manual.* It is an extension of SMASCH in the sense that its initial development is occurring within the SMASCH project at UC Berkeley, in conjunction with efforts to provide—through the construction of electronic, multiple-entry keys—identifications of California plant material to be included in the SMASCH database. In the future, development of MEKA keys for FISCT will be undertaken mostly by contributors to *The Jepson Manual,* in connection with their efforts to annotate specimens belonging to and revise treatments of the groups for which they are authors. Thus, FISCT will comprise MEKA keys to the taxa of vascular plants occurring without cultivation in California, and eventually will include for each, via electronic versions of *The Jepson Manual,* dichotomous keys, morphological as well as ecological descriptions, and digitized illustrations. FISCT is an application of SMASCH in that it will rely on that system for accession-level data, including but not limited to nomenclatural histories, the citation of relevant literature, elevational ranges and geographic distributions. Ultimately, each accession record in the SMASCH database will be connected by its geographic coordinates to digitized maps of California, enabling nearly instantaneous display of distributions as well as checklists, based on whatever annotations users wish to specify. FISCT will serve as an interface between SMASCH and *The Jepson Manual* once the latter is put into machine-readable form and connected electronically to the SMASCH database.

Botanists and others have been speculating, at least informally, about the potential advantages of electronic floras for some time now. Among other things, most observers agree that they would enhance the general flow of information about plants and that they would be much more readily corrected and updated than their paper-based counterparts have been. Nevertheless, it was apparent to us in attending the session "Flora of the World, Current Knowledge and Future Prospects" at the recent International Botanical Congress in Yokohama that little has been done to implement the ideas and to make use of the technology involved; participants discussed various floristic projects throughout the world, including a world flora (*Species Plantarum*), yet mentioned the application of networked information technology only in passing or did not consider the subject at all.

MADROÑO, Vol. 42, No. 2, pp. 189–196, 1995

In this paper a plan is outlined by which data about the flora of California will be distributed electronically over computer networks. It involves the SMASCH project and its accession-level data on one hand (Bartholomew and Duncan 1992; Duncan et al. 1993; Rosatti et al. 1994), and *The Jepson Manual* and its information about taxa on the other (Hickman 1993). The plan will require cooperation and collaboration among the SMASCH project, the Museum Informatics Project (MIP), the Jepson Herbarium and Trustees, the University of California Press, the University Library, and the editors and contributors who will be involved in future editions of *The Jepson Manual.* In the electronic format that we propose, the text of *The Jepson Manual* will be much more easily searched and revised, while it and related resources will be available to a larger number of users for a wider range of purposes.

Background: MEKA Keys

As implementation of SMASCH proceeded, we began to develop electronic, multiple-entry identification keys because of practical needs related to our commitment to identify all California accessions in JEPS and UC for the database. In the process, we formulated a plan that would encompass all of the elements that could be regarded as comprising a truly complete, electronic flora, and that could serve as a prototype for others as well.

We currently are using treatments in *The Jepson Manual* to help identify accessions of California plants in JEPS and UC. For genera in which only one species is known to occur in the state, we are using these treatments directly, but for genera in which more than one species is involved, we are using data primarily from *The Jepson Manual* as a starting point in the construction of electronic, multiple-entry identification keys, using the interactive, Multiple-Entry Key Algorithm (MEKA and MEKAEDIT), developed by Duncan and Meacham (1986a, b) and Meacham (1994).

In the development of these MEKA keys, data primarily from *The Jepson Manual* are being used to fill out data matrices in which values of "+" (present in all members of a taxon), "−" (absent from all members of a taxon), "*" (present in at least one but not all members of a taxon), and "?" (condition unknown for a taxon) are entered for each of a string of character states for each taxon. A specimen then may be identified by clicking on character states that apply until the list of possible taxa is reduced to one.

For our purposes, multiple-entry keys are preferable to dichotomous ones in several ways (Tables 1 and 2): they do not require that each of a fixed series of choices be made correctly in order to arrive at an identification; they do not often involve bottlenecks (e.g., steps in dichotomous keys that rely on single structures, such as flowers

TABLE 1. DICHOTOMOUS KEY PATH IN *THE JEPSON MANUAL* AND MINIMUM DIAGNOSIS THROUGH MEKA FOR *CHAENACTIS SANTOLINOIDES* GREENE. In this example, the character states needed to completely distinguish *C. santolinoides* using a minimum diagnosis from MEKA are fewer in number, more easily determined, and more frequently present than those encountered in the key path to this taxon in this dichotomous key to the 20 taxa of *Chaenactis* treated in *The Jepson Manual* (see Hickman 1993, p. 223).

The Jepson Manual (5 key leads, 18 character states):

1. Per (bien?) to subshrubs, rarely fl 1st year; pappus scales (6)8–20 per fr in distinct series, ± equal; outer corollas radial, not strongly enlarged (sect. *Macrocarphus*).
   2. Pls ± scapose, cespitose to ± matted; heads 1(–3) per st; gen subalpine to alpine.
      3. Phyllaries glandular-hairy.
         4. Lf blades linear to elliptic, longest lobe near middle, tips curled to twisted.
            5. Largest lf blades linear to narrowly elliptic, lobes gen > 9 pairs, densely crowed ..... *Chaenactis santolinoides*

MEKA (2 character states):

Longest primary leaf lobes near blade middle, primary lobes of largest leaves gen > 9 pairs ..... *Chaenactis santolinoides*

or fruits, that are not always present); they allow for the determination of minimum diagnoses, which are lists or sets of lists of the fewest numbers of character states needed to completely distinguish a taxon (e.g., a species in a genus); they facilitate the determination of unique character states, which are those that by themselves completely distinguish a taxon; and they may be expanded with respect to both taxa and character states quite readily, with no need for restructuring of any kind.

The advantages of MEKA keys discussed above are of tremendous value to us in meeting our commitment to identify over the next 5 years or so approximately 300,000 accessions of California plants in JEPS and UC. It is much faster and easier to identify a stack of herbarium specimens by determining only one or a few character states for each than it is to perform the same task by having to read through and choose from many more character states in a dichotomous key; of course, the severity of this limitation depends on the length of the key and the position of the taxa within it. In addition, the minimum diagnoses and unique character states in some cases exclude habit, reproductive structures, underground parts, and other features commonly found in dichotomous keys that often are absent or difficult to determine.

Preliminary MEKA keys, which are direct and unedited translations of some of the data from dichotomous keys and descriptions

TABLE 2. KEY LEADS AND CHARACTER STATES ENCOUNTERED IN DICHOTOMOUS KEY PATHS IN *THE JEPSON MANUAL* AND MINIMUM DIAGNOSES THROUGH MEKA FOR EXAMPLES SELECTED FROM GENERA A THROUGH C IN CALIFORNIA. In most of these examples, the number of character states needed to completely distinguish a taxon using a minimum diagnosis from MEKA is fewer than both the number of character states and the number of key leads encountered in the key paths to these taxa in the dichotomous keys in *The Jepson Manual* (see Hickman 1993, pp. 202–203, 215–216, 223–224, 229).

Taxon	Dichotomous key leads in *The Jepson Manual*	Dichotomous key character states in *The Jepson Manual*	MEKA character states (minimum diagnosis)
Artemisia tridentata subsp. *tridentata*	11	34	3
Artemisia norvegica subsp. *saxatilis*	6	16	2
Artemisia spinescens	2	8	1
Brickellia arguta var. *odontolepis*	7	14	7
Brickellia frutescens	6	13	3
Chaenactis glabriuscula var. *megacephala*	7	21	3
Chaenactis stevioides	6	27	3
Chrysothamnus nauseosus subsp. *mohavensis*	7	16	3
Chrysothamnus viscidiflorus subsp. *viscidiflorus*	8	16	3

into MEKA format, have been prepared mostly by Data Entry Assistants in the SMASCH project. Once in MEKA format, expressions of the character states in English as well as the values for each taxon with respect to each of these character states are modified, when necessary, as a result of observations made in the process of examining the plant material to be identified, yielding a final MEKA key. In some cases, additional character states are added, from the plants themselves or from various sources in the literature, to complete or strengthen the distinctions among taxa. Thus far we have produced, mostly for Asteraceae in California, preliminary keys for about 80 genera and 450 species and infraspecies, and "final" keys for about 32 genera and 212 species and infraspecies. In the future, others will be involved in this process in that contributors revising treatments for subsequent editions of *The Jepson Manual* will annotate the pertinent accessions in JEPS and UC and produce MEKA keys for their groups as well.

THE PLAN: FISCT

Over the long-term we will develop a complete floristic information system for the vascular plants of California, of which an

electronic version of *The Jepson Manual* will be a most important part, to be accessible over Internet and to be known as the Floristic Information System for California Tracheophytes (FISCT); we expect that within the next five years a functional prototype of this system will be in place According to our plan, FISCT eventually will comprise electronic, multiple-entry identification keys to the taxa of vascular plants occurring without cultivation in California and will include for each, via electronic versions of *The Jepson Manual,* dichotomous keys, morphological as well as ecological descriptions, horticultural information, and digitized illustrations; each taxon will be represented by digitized photographic slides as well, from the Jepson Slide Collection as well as other sources, of entire plants in their habitats as well as of individual parts critical in identification. For type specimens of these (and other) taxa, we will build an archive within SMASCH of images, including both magnifications and dissections showing critical features.

FISCT will serve as an interface between an electronic *Jepson Manual* and the SMASCH database in that it will rely on *The Jepson Manual* for the elements listed above, in addition to the glossary and other introductory material, while it will rely on SMASCH for images of accession sheets as well as accession-level data including but not limited to nomenclatural and annotation histories, the citation of relevant literature, elevational ranges, and geographic distributions. Ultimately, each accession record in the SMASCH database will be connected by its geographic coordinates to digitized maps of California, enabling nearly instantaneous display of the distributions of taxa in the state as well as checklists of taxa for any area of interest, based on whatever annotations users wish to specify.

Regarding our own annotations, we have kept in fairly standardized form notes regarding the ways individual accessions do not conform to the taxa to which we have assigned them. Thus, for example, users eventually will be able to obtain a point-by-point distribution on a digitized map of all accessions we have assigned to a particular taxon, as well as of all accessions we have assigned to that taxon that have a character state that does not conform to that taxon. In these and other ways, SMASCH will serve as a tool in the evaluation of taxa and patterns of varation for a wide range of purposes.

Staff of MIP have loaded from tape and placed in Interleaf format the entire text of *The Jepson Manual,* and portions of it have been used in collaboration with the staff of the University Library as part of a more general effort to explore the use of Standard Generalized Markup Language (SGML) for processing electronic text. SGML is an emerging standard for encoding text, an important consideration for ensuring long-term viability as well as interoperability via computer networks with systems that will be developed elsewhere in the

world. The development of network applications and the use of standard editing tools for SGML also will facilitate the ongoing correction and revision of *The Jepson Manual.*

In the context of FISCT, the use of SGML markup (which describes the structure or content of a text, not its appearance) in conjunction with a document type definition (DTD) designed for floristic information will allow for the extraction of sections of *The Jepson Manual* based on defined criteria as well as context-based searching of the text. For example, habitat descriptions for specific taxa could be extracted, or occurrences of an idea such as "red flowers" could be identified by searching contexts defined as "flowers," "petals," "corollas," and "stamens" for the word "red" (thus excluding other contexts in which "red" would be irrelevant).

An initial prototype of an electronic version of *The Jepson Manual* has been created using as sample text the family description and treatments of the first nine genera of Asteraceae (*Acamptopappus* through *Agoseris*). The user interface employed in this prototype is the Dynatext Browser (Electronic Book Technologies, Inc.), a system for electronic publishing that provides formatting as well as searching capabilities based on SGML markup. For the prototype, contexts including but not limited to description, chromosome number, commonness and rarity, habitat, elevation, geographic range, and horticultural value have been encoded, making it possible to search this entire document or specified parts of it for values of interest within them. For example, the command in Dynatext's query language "⟨habitat⟩ cont juniper woodlands" executed on the entire sample document indicates that only the habitat statement for *Acamptopappus sphaerocephalus* contains the words "juniper woodlands." Although it is generally possible to search an electronic version of a text for occurrences of something like "juniper woodlands" or any other string of characters, it is desirable to define contexts (such as "habitat" in this case) for purposes of interoperability and in order to exclude occurrences of a given string of characters that are considered irrelevant to the question being asked.

The Dynatext prototype also includes a "table of contents," which lists the taxa included hierarchically and allows a user to go instantly to any desired treatment by simply clicking on the appropriate taxonomic name; the results of searches are also displayed numerically, by taxon, in this list. Another feature built into the prototype is a mechanism whereby clicking on a scientific name in a dichotomous key sends the user to the treatment of the corresponding taxon. Finally, some of the smaller illustrations have been embedded in the text in appropriate places, while the larger ones, in addition to digitized photographic slides, are brought onto the screen by clicking on icons.

Eventually, we will connect our MEKA keys and the text of *The Jepson Manual* to a fully illustrated glossary, so that by merely

clicking on a word of unknown or uncertain meaning a user will be provided with a pop-up definition and accompanying picture. We believe that such a tool, in addition to other mentioned advantages of electronic keys and floras in general, will result in more accurate identifications of plant material, particularly by non-specialists, and might even render the prospect of identifying plants less formidable to those with budding interests in the field. Of course, we would caution users that in critical cases, as with any floristic work, other types and sources of information should be consulted for confirmation.

Conclusions

Our goal is to make various kinds of floristic information available in electronic form, for use in herbaria, libraries, offices, homes, and even in the field (using hand-held computers). We also expect that in the near future, in addition to coordinates for collection localities determined by a geographic positioning system (GPS), a person in the field will be able to enter other collection data into hand-held computers for later input into databases and production of collection labels.

The plan discussed here was not developed in isolation, and its implementation will not proceed without cooperation and collaboration with others. With regard to accession-level data, SMASCH was developed with input from throughout the community of expected contributors and users, and its implementation will continue to proceed in close association with the California Academy of Sciences and the Rancho Santa Ana Botanic Garden. Representatives from major institutions in New Mexico, Arizona, Oregon, Washington, Nevada, and Hawaii are preparing grant proposals involving the use of SMASCH for their collections, and have held discussions with us about forming a Western United States Botanical Accession Database Federation. On the floristic side, collaborations will occur that involve the people and institutions mentioned above, and we have been and continue to be in close contact with the Biota of North America Program (BONAP) at the University of North Carolina about how FISCT will mesh with their plans for a floristic information system for North America north of Mexico.

The system we envision will not only store, manage, and distribute information about the flora of California, it will also serve as a research tool in the generation of such data by monographers and floristicians. Accession-level data, a major source of the information on which taxonomic and floristic information is based, will be more accessible and better managed in the future because of SMASCH, while the ability to evaluate patterns of variation will be dramatically enhanced by digitized mapping and other geographic tools. These and other improvements should result in sounder, more efficiently

produced taxonomic and floristic treatments of California plants in the future.

Acknowledgments

The plan discussed in this paper would not have been developed without support provided to the SMASCH project by the National Science Foundation, through NSF-BSR-9104156. Wee thank many people for the contributions they have made thus far. Randy Ballew of MIP and Mickey Ellinger of SMASCH have been responsible for making the hardware and software do what we all think it should. Elizabeth Neese, formerly of SMASCH, provided important insights from her perspective as a field botanist. Courtney Terry and Liza Lewis, both formerly of SMASCH, constructed most of the preliminary MEKA keys. Glen Jackson of MIP resurrected the back-up copies of *The Jepson Manual* and placed them in Interleaf on the MIP server for use in the SGML prototype. Susan Stone of MIP marked-up in SGML the part of *The Jepson Manual* discussed, and has been very helpful in explaining to us what she has learned, in collaboration with the University Library, about this standard for encoding text. Peter Rauch has experimented with Kodak photo CD technology as a way to have photographic slides quickly and inexpensively digitized, and has located substantial sources of such material outside the University for inclusion in FISCT; Brother Eric Vogel of St. Mary's College has offered us the use of about 20,000 slides of California plants from the collections at that institution. We are also grateful to Chris Meacham of MIP for modifying and improving MEKA and to John Kartesz of the Biota of North America Program and Buck Reed of the U.S. Fish and Wildlife Service for discussions about how FISCT will fit into larger plans for an electronic floristic information system for the entire continent.

Literature Cited

Bartholomew, V. and T. Duncan. 1992. The Specimen Management System of California Herbaria as a model for an inter-institutional distributed database system. Pp. 85–91 in Ching-I Peng (ed.), Phytogeography and botanical inventory of Taiwan. Institute of Botany, Academia Sinica Monograph Series No. 12. Tapei, Republic of China.

Duncan, T. and C. A. Meacham. 1986a. Multiple-entry keys for the identification of angiosperm families using a microcomputer. Taxon 35:492–494.

——— and ———. 1986b. MEKA (Version 1.3)/MEKAEDIT (Version 1.3): a general purpose multiple-entry key algorithm and editor. Unpublished manuscript available through the University Herbarium and the Museum Informatics Project, University of California at Berkeley.

———, T. J. Rosatti, J. D. Morefield, J. Beach, S. D. Jacobson, R. Ballew, and D. Kelley. 1993. A relational data model for botanical collections. Unpublished manuscript available through the University Herbarium and the SMASCH project, University of California at Berkeley.

Hickman, J. C. (ed.). 1993. The Jepson manual: higher plants of California. University of California Press, Berkeley.

Meacham, C. A. 1994. MEKA (Version 2.0): a general-purpose multiple-entry key algorithm. Unpublished manuscript available through the University Herbarium and the Museum Informatics Project, University of California at Berkeley.

Rosatti, T. J., E. C. Neese, A. N. Kikuchi, C. S. Terry, and L. M. Lewis. 1994. Specimen Management System for California Herbaria (SMASCH): data entry. Unpublished manuscript available through the University Herbarium and the SMASCH project, University of California at Berkeley.

(Received 29 Sep 1994; accepted 1 Feb 1995)

Rosatti, Thomas J and Duncan, Thomas. 1995. "FLORISTIC INFORMATION SYSTEM FOR CALIFORNIA TRACHEOPHYTES (FISCT)." *Madroño; a West American journal of botany* 42, 189–196.

View This Item Online: https://www.biodiversitylibrary.org/item/185854
Permalink: https://www.biodiversitylibrary.org/partpdf/171270

Holding Institution
Smithsonian Libraries and Archives

Sponsored by
Biodiversity Heritage Library

Copyright & Reuse
Copyright Status: In Copyright. Digitized with the permission of the rights holder
Rights Holder: California Botanical Society
License: http://creativecommons.org/licenses/by-nc/3.0/
Rights: https://www.biodiversitylibrary.org/permissions/

This document was created from content at the **Biodiversity Heritage Library**, the world's largest open access digital library for biodiversity literature and archives. Visit BHL at https://www.biodiversitylibrary.org.

This file was generated 1 February 2024 at 01:55 UTC