Next-Generation Field Guides

ELIZABETH J. FARNSWORTH, MIYOKO CHU, W. JOHN KRESS, AMANDA K. NEILL, JASON H. BEST, JOHN PICKERING, ROBERT D. STEVENSON, GREGORY W. COURTNEY, JOHN K. VANDYK, AND AARON M. ELLISON

To conserve species, we must first identify them. Field researchers, land managers, educators, and citizen scientists need up-to-date and accessible tools to identify organisms, organize data, and share observations. Emerging technologies complement traditional, book-form field guides by providing users with a wealth of multimedia data. We review technical innovations of next-generation field guides, including Web-based and stand-alone applications, interactive multiple-access keys, visual-recognition software adapted to identify organisms, species checklists that can be customized to particular sites, online communities in which people share species observations, and the use of crowdsourced data to refine machine-based identification algorithms. Next-generation field guides are user friendly; permit quality control and the revision of data; are scalable to accommodate burgeoning data; protect content and privacy while allowing broad public access; and are adaptable to ever-changing platforms and browsers. These tools have great potential to engage new audiences while fostering rigorous science and an appreciation for nature.

Keywords: species identification, taxonomy, citizen science, Web applications, digital data

Accurate species identification is a crucial prerequisite to documenting, managing, and sustaining the diversity of life on Earth. Basic identification begins with field guides: clearly presented syntheses of technical information about taxa, paired with identification keys for distinguishing among taxa. Close observation and illustrated, printed field guides are the traditional tools for organism identification. Now, myriad new electronic tools are emerging to help everyone from curious novices to seasoned biologists identify existing species; determine whether a species is new to science; and share observations, data, and discoveries with the wider world (Stevenson et al. 2003, Agarwal et al. 2006).

Here, we explore how field guides are overcoming the limitations of bound books by evolving in tandem with information technology. We review a broad range of applications (or *apps*) and detail several case studies to illustrate how next-generation field guides are facilitating the identification of organisms; being used to create customized guides to the flora and fauna of particular sites; promoting networking among a new generation of naturalists; enabling the collection and sharing of valuable scientific data; and encouraging interdisciplinary research in biology, computer science, education, and cognition. We discuss emerging innovations that are yielding especially successful apps in terms of their accuracy; ease of use; and ability to stimulate learning, particularly among nonscientists. We also highlight important challenges for ensuring that next-generation field guides and their associated resources can successfully reach their intended audiences, adapt to and evolve with new information and technologies, and be sustained over the long term. Although our focus is on field guides to organisms, classification systems are central to the practice of any science. The innovations and challenges that we discuss here can inform the creation of user-friendly, rigorous guides to items from molecules to Messier objects.

What is a field guide?

The term *field guide* broadly encompasses geographically restricted or taxonomically constrained (*pragmatic*) checklists, monographic treatments of particular taxa, comprehensive descriptions of regional natural communities, textbooks, nontechnical illustrated posters, flashcards, or brochures, as well as hybrids of these (Hawthorne and Harris 2006). The earliest field guides were created before the fourteenth century and were illustrated, utilitarian descriptions, such as herbals (Givens 2006), but until the invention of movable type, these works were reproduced rarely and saw only limited distribution. The expense of illustrations precluded widespread publication of biological compendia; illustrations were often eschewed in favor of technical text in the form of dichotomous keys (Scharf 2009). Books containing such keys burgeoned in the eighteenth century as the number of newly described species increased exponentially (Scharf 2007).

The first field guides with contemporary characteristics—detailed descriptions of species, with illustrations, clear taxonomic organization, and prose accessible to the lay public—were published in the early 1900s (for a review, see Dunlap 2005). Today, field guides in wide use by both professional and amateur scientists typically consist of two sections: (1) an overview of the broad group of organisms of interest, including tips for accurately observing them, their evolutionary relationships, and keys with which to identify them, and (2) species accounts, featuring descriptive

BioScience 63: 891–899. ISSN 0006-3568, electronic ISSN 1525-3244. doi:10.1525/bio.2013.63.11.8

images, range information, ecology, behavior, habitats, taxonomy, and synonymy (Stevenson et al. 2003).

Although many field guides are written for general audiences, they are also being used increasingly as authoritative references for species identification. For example, Schmidt (2006) found a nearly fivefold increase in citations of field guides—from 50 to 248 per year—in scientific publications between 1990 and 2004. Likewise, consumer demand for field guides continues to grow. In 2000, a search of Amazon.com book titles for the keywords *field guide* yielded 625 total entries (Stevenson et al. 2003); in 2012, a we ran a similar search, which yielded 1849 titles, with 81 forthcoming and published titles for 2012–2013 alone. Of these latter publications, 29 covered birds; 19 dealt with plants and fungi; 10 were on mammals and herpetofauna; 8 were about invertebrates; 7 were site-based treatments of multiple types of organisms; and 8 addressed gems, fossils, animal tracks, weather, or astronomical objects.

The classical bound field guide has well-known limitations. Books are relatively expensive, not always portable, and their information can be updated only when new editions are published. Crucial features for identifying many animals, including sounds (e.g., bird calls), movement patterns, or behavioral characteristics can only be approximated in print but are communicated much more easily with video or audio media. These and other multimedia features are increasingly included as online supplements to printed field guides or as stand-alone, entirely digital field guides.

Digital portals also invite active participation by users in ways that a book cannot offer. Near-universal inclusion of Global Positioning System (GPS) capabilities into handheld devices lets people instantly record where they sighted an organism and simultaneously contribute their observations to online communities, social networks, and repositories of scientific data. Educators have long recognized that K–12 and college students are comfortable with—and quickly master—such sophisticated hardware and software (Ellis 1984, Tapscott 1998). New technologies are frequently used by educators seeking to better engage students in studying science, technology, engineering, and mathematics and to involve volunteers in citizen-science projects (Kress 2004, Newman et al. 2012). In fact, a recent meta-analysis showed that using new instructional technologies and fostering student collaboration in scientific inquiry both have positive impacts on achievement (Schroeder et al. 2007).

The evolution of online field guides and identification keys

Digital identification tools have increased both in number and computational sophistication in the past several decades (Dallwitz et al. 2013). Early punch-card keys (*polyclaves*) allowed users to narrow down a set of species by physically aligning those that had matching character states (e.g., Simpson and Janos 1974). Basic computational features developed in the late 1960s included the ability for a user to choose the characters used for keying (i.e., relevant, observable characters, not a constrained series of steps); the ability to enter numeric values for character states; built-in ranges of data to account for uncertainty, user error, or polymorphisms in character states; and the retention of taxa during keying when data were missing for particular character states (Goodall 1968). Later, digital keys offered users guidance on the most informative questions to answer for particular taxa (Morse 1971) and also made it clear when certain character states were not applicable or were contingent on the states chosen for other characters (Pankhurst 1991). Many digital keys were—and are—types of *expert systems* (a term from artificial intelligence), in which background knowledge of particular taxa (the heuristic knowledge that many experienced taxonomists have) has been hard-coded into the programming, thus proffering the most informative questions first to the user in a semiguided keying process (Edwards and Morse 1995). Many programs also tend to choose species on the basis of positive matches across a range of character states rather than by eliminating taxa from the results set if one or more of their character states do not fit the data or if a full set of data is missing.

Today, the emphasis is on usability for novices and experts alike, the versatility of platforms and presentation, and accurate data. Stable technology supports Internet-based and stand-alone apps for smartphones, tablet computers, and other portable devices. Users can search online keys for specific taxa or characteristics. They can also follow *breadcrumb trails*—a navigation tool metaphorically similar to their namesake—to retrace their steps if they go astray. Online keys often use a variety of media, including text, drawings, photographs, audio, and video, set in visually appealing user interfaces that facilitate taxon identification with a minimum of steps (Leggett and Kirchoff 2011). Pop-up windows or hyperlinked glossaries define technical terms. The clever use of multimedia and machine-learning algorithms makes stand-alone apps and online guides simultaneously accessible to beginners and useful for specialists. Together, these features increase the accuracy of identifications, offer rewards for the user, and encourage learning.

Creating digital field guides

Several software packages and online resources allow people to create digital keys for organisms using a standard taxon × character-state data matrix (table 1). Some programs, including Intkey, IdentifyIt, Linneaus II, Lucid, MEKA, NaviKey, PollyClave, XID, and xPer2 are stand-alone freeware or proprietary software (Dallwitz 2011). Others, including ActKey (Brach and Song 2005), eFloras (Brach and Song 2006), and Stinger's Lightweight Interactive Key Software (SLIKS) run as Web-based apps. Technical reviews of many of these packages have been published elsewhere (Edwards and Morse 1995, Dallwitz 2011).

More recently, the Electronic Field Guide (EFG; *http://efg.cs.umb.edu/efg*; table 1) Project of the University of Massachusetts Boston has developed software that allows scientists and lay people to make their own Web-based,

Table 1. Software for creating customized field guides.

Software	Web address
The Electronic Field Guide Project	*http://efg.cs.umb.edu/efg*
Linnaeus II	*www.eti.uva.nl/products/linnaeus.php*
Horticopia	*www.horticopia.com*
Lucid Key Server	*www.lucidcentral.com/en-us/software/lucidkeyserver.aspx*
SLIKS (Stinger's Lightweight Interactive Key Software)	*www.stingersplace.com/SLIKS*
The XID Authoring System	*http://xidservices.com*
Xper2	*http://lis-upmc.snv.jussieu.fr/lis/?q=en/resources/software/xper2*
Intkey	*http://delta-intkey.com/www/programs.htm*

digital field guides without the restrictions imposed either by paper formats or by commercial considerations (Morris et al. 2007). Like most electronic field-guide generators, the initial efforts of the EFG Project were focused on developing pictorial or text-based keys that allow users to choose different views and ways of accessing them. The subsequent efforts were more flexible. In the Microsoft Windows– and Linux-compatible EFG2 software, a simple text file consisting of a taxon × character-state matrix is used to organize and store text-based information. This master text file also contains pointers to a folder that includes files in a variety of media, including illustrations, digital photographs, maps, video clips, and sounds. While creating or updating guides, the user can drag and drop files into the folder to import new content. The final product includes Web displays of taxon pages, configurable lists of taxa, and browse-and-search modes. End users have found this model (in which a master file points to a folder) easy to learn and sufficiently flexible to enable quick construction of custom field guides. Over 30 customized guides, covering a range of taxa and regions, have been produced to date using EFG2.

Customizing field guides for particular situations

EFG is a generic platform for making field guides. With an existing field guide in hand, users can easily adjust the species set to fit their locality or interest. The central concept of a customizable field guide is the *local list*: a subset of information extracted from a much larger database but restricted to, for example, a particular location, habitat, time of year, time of day, or observation method. The key to a workable local list is a large database with a flexible structure. For example, the New England Wild Flower Society's Go Botany database structure is generic and extensible (*http://gobotany.newenglandwild.org*; also see supplemental table S1, available online at *http://dx.doi.org/10.1525/bio.2013.63.11.8*, for links to all of the guides discussed here). The user interface and species list can potentially be customized for any flora, and the New England Wild Flower Society is working with five institutional partners to create floras for their regions.

Likewise, the Web-based, open-source Atrium biodiversity information system (*www.atrium-biodiversity.org*) includes a digital herbarium, a geographic information system data repository, a bibliographic reference-management system, a meteorological data module, and a module for managing and analyzing quantitative vegetation survey data. Users can browse or filter collection records by taxonomy, collector, project, or geographic region. Detailed collection data, high-resolution zoomable images of fresh plant material and preserved specimens, maps of collection localities, and multilevel taxon descriptions are all viewable. All data and images are downloadable and can be reused in customized guides; the Field Guide Generator utility is one of the most popular modules in Atrium.

Behind the scenes: The importance of database structures and semantics

The utility of any digital field guide is dependent not only on the availability and accuracy of species-level data but also on a common (and formal) language (*ontology*) used to specify a set of core concepts and ideas (e.g., Walls et al. 2012). Such formal structures are especially important for field guides covering many taxa with multiple-access keys (modern-day polyclaves). As taxon × character-state matrices get very large (more than 500 taxa [rows] or more than 50 character states [columns]), searching slows down, and errors are increasingly likely to occur in data input. Most developers shift to relational databases (rDBs) to handle large numbers of taxa, to increase the database query speed and to reduce the number of repetitive entries. For example, in an rDB, a genus name need only be entered once; species entries then point to the genus table to capture higher-level characters.

Relational databases (and the digital field guides based on them) use a *semantic* data model, in which instances (e.g., *red oak*) are defined for types (e.g., *tree*). Several such standardized semantic frameworks exist for taxonomy: the Description Language for Taxonomy (DELTA), Lucid, and Nexus (Dallwitz 2010). These frameworks work with many types of data: ranked or unranked multistate categorical variables, continuous or discrete numbers, or even free text. These data not only serve to record machine-readable character states for various taxa so that species can be distinguished but can also be parsed to provide natural-language descriptions of taxa. Finally, common semantics-based systems allow data to be exchanged among different apps (*interoperability*), which can allow end users to create new types of field guides that could, for example, combine the data for multiple higher-level taxa in a single habitat.

Relational databases also provide field-guide authors with easy-to-use interfaces for updating or correcting data as new information becomes available. For example, as taxonomic names are changed or new distribution records are discovered, a single new entry in an rDB can ensure that the field guide and keys remain definitive sources for accurate names and biogeography. Now that any taxon can be described or revised in online-only publications (Labeda and Oren 2008, Knapp et al. 2011, Zhang 2012), online identification keys, updated regularly and even in near real time, will supplant printed keys.

An important challenge for any digital field guide is its *scalability*: the capacity for a database both to accommodate and to rapidly deliver information—such as character states, images, and other linked data—for tens of thousands or even millions of diverse taxa. Physical space (memory) is rarely an issue for online (Web-based) guides, but it is for stand-alone apps loaded on small mobile devices, whose limited processor speeds also increase the time required to search through vast databases. Formal ontologies and semantics can help speed the searching and parsing of large databases. Graph databases, in which items in the database are linked together—like computers on the World Wide Web or neural nets (Edwards and Morse 1995)—using nodes and the paths between them instead of the look-up index of an rDB, have the as-yet-unexplored potential to further speed the search of digital field guides with multiple-access keys and very large databases (Angles and Gutierrez 2008).

Features of interactive Web-based identification keys

We compiled a list of 50 species-identification tools that are available online (table S1); all but 10 were Web based (i.e., not stand-alone apps). Although this list cannot be exhaustive (identification apps are burgeoning), the entries illustrate the range of features currently available in online field guides. After summarizing these features (see table S1 for a concise summary and Web addresses for all sites and apps discussed), we illustrate specific attributes with reference to the resources that we have developed.

All but four of the digital identification keys (96%) provide detailed data on specific taxa, including range maps, information on life histories, and distinguishing characteristics. Thirty-nine (76%) enable users to search or browse for a particular species of interest. Almost half (43%) offer a glossary of technical terms or a dedicated help page with tips for usage. Twenty (39%) allow users either to upload data to a central repository or to share data with selected other users. Of the 29 apps that feature identification keys, 18 have multiple-access keys, and 14 offer dichotomous keys.

Go Botany is one example of a resource that has all of these features (figure 1). Go Botany is a free suite of Internet-based interactive identification keys and learning tools that runs on desktop computers, laptops, tablets, and mobile devices. Dynamic multiple-access keys and more-technical, clickable dichotomous keys appeal to novices and experts, respectively. Both types of keys allow users to track and change their path to identification using breadcrumb trails. Several aspects of the display, including autoprompting search tools and virtual display cases showing thumbnail images of plants, are adapted from familiar formats originally developed for e-commerce (Matt Belge, Vision and Logic LLC, personal communication, 23 June 2013); therefore, even novice users learn how to use the app very quickly. Alternatively, users can select a candidate species set by directly choosing the plant family or genus; family and genus information pages help the users learn higher levels of taxonomic organization, which is useful in formal botany courses.

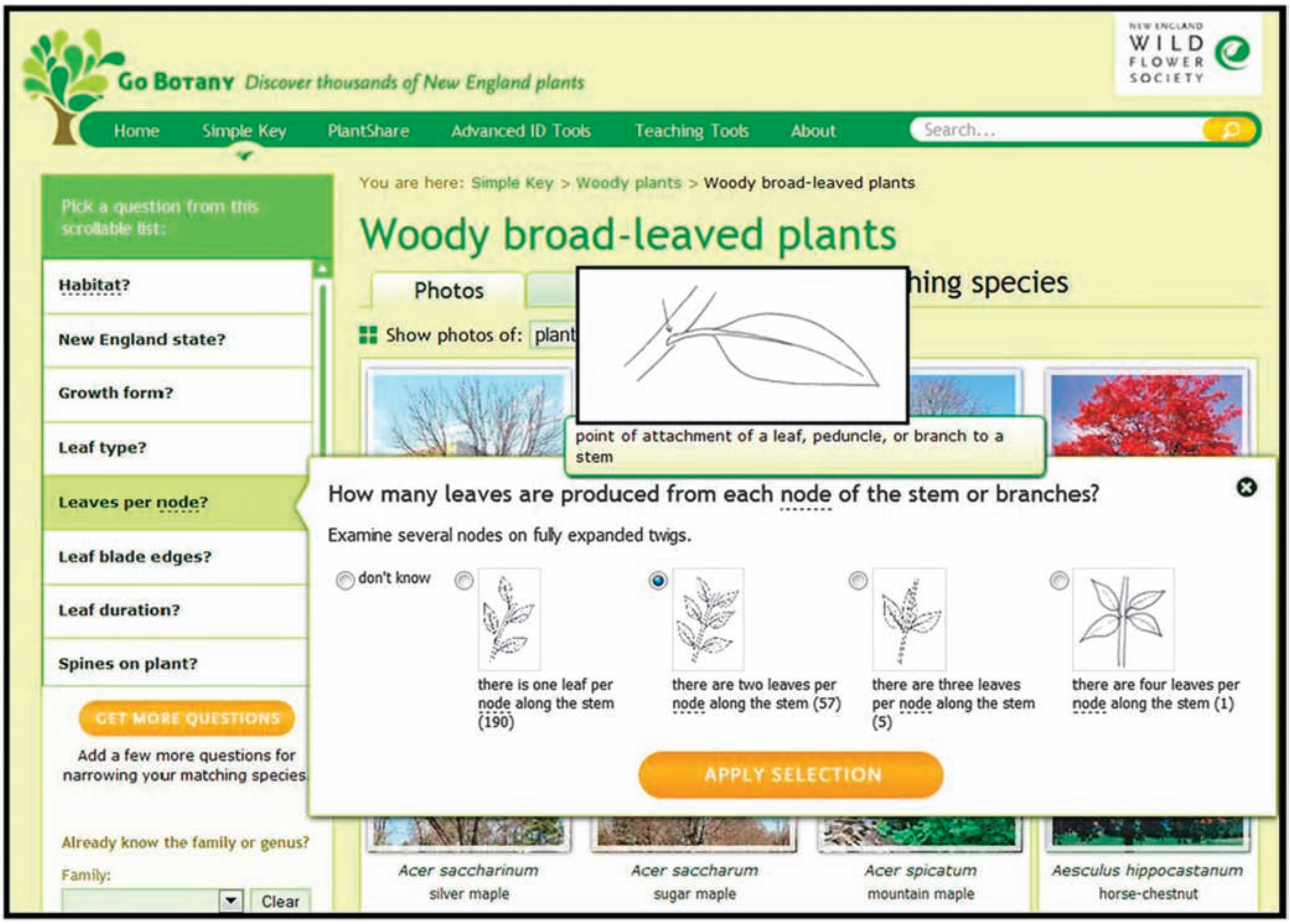
Go Botany includes more than 3500 plant taxa native and naturalized to New England (Haines 2011) and is built from an extensive database covering plant morphology, habitat affinities, synonymies, look-alike taxa, and species distributions. Over 37,000 photographs, technical drawings, and range maps illustrate the keys and species information pages. This wealth of detail presents a key challenge for Go Botany and other online multiple-access keys for many taxa: The application requires a connection to the Internet, because a stand-alone app would use most of the storage capacity currently available on handheld devices.

Stand-alone mobile apps

Smartphones and tablets are nearly ubiquitous. These portable devices have increasingly large amounts of storage; built-in, GPS-enabled cameras; and instant connectivity to social networks and networks of experts. Many field guides can be stored on a single smartphone and used as identification manuals, study guides, and data-collection devices. Hundreds of commercial apps are available for purchase or free download, covering a wide array of species; a recent (7 August 2013) search yielded 25 apps for plant identification alone on iTunes, 9 of which are free. Many apps exhibit a variety of features, including identification tools that use simple icons representing character states, information on each taxon, and the ability to network with others and instantly upload sightings (table S1).

The Guide to Texas Range and Pasture Plants, from the Botanical Research Institute of Texas (BRIT), (*www.brit.org/rangeplants*) is an example of a simple, image-based plant-identification system aimed at the general public. The BRIT guide is an inexpensive app with which rural students, farmers, ranchers, and naturalists can view and study images of herbarium specimens to identify 129 species of range plants. The guide provides images; nomenclature; pronunciation guides; and a brief description of each plant that includes its growth season, its value for wildlife and grazing animals, and data on whether it is native. Users can review species with a flashcard feature or can test themselves with identification quizzes.

Computer-based visual recognition is used to identify species in another way; a person sees an organism, photographs it, and queries a database for the identity of the resulting image (MacLeod 2008). Rapid advances in using visual-recognition software are yielding automated systems for identifying plants, insects, vertebrates, and benthic invertebrates (e.g., Gobi 2010, Lytle et al. 2010). Leafsnap (figure 2; *http://leafsnap.com*) is a widely used visual-recognition app for identifying trees in the northeastern United States; it was



Pick a question from this scrollable list:

- Habitat?
- New England state?
- Growth form?
- Leaf type?
- Leaves per node?
- Leaf blade edges?
- Leaf duration?
- Spines on plant?

Add a few more questions for narrowing your matching species

Already know the family or genus?

Family:

You are here: Simple Key > Woody plants > Woody broad-leaved plants

Woody broad-leaved plants

hing species

Show photos of: plant

point of attachment of a leaf, peduncle, or branch to a stem

How many leaves are produced from each node of the stem or branches?

Examine several nodes on fully expanded twigs.

○ don't know
○ there is one leaf per node along the stem (190)
● there are two leaves per node along the stem (57)
○ there are three leaves per node along the stem (5)
○ there are four leaves per node along the stem (1)

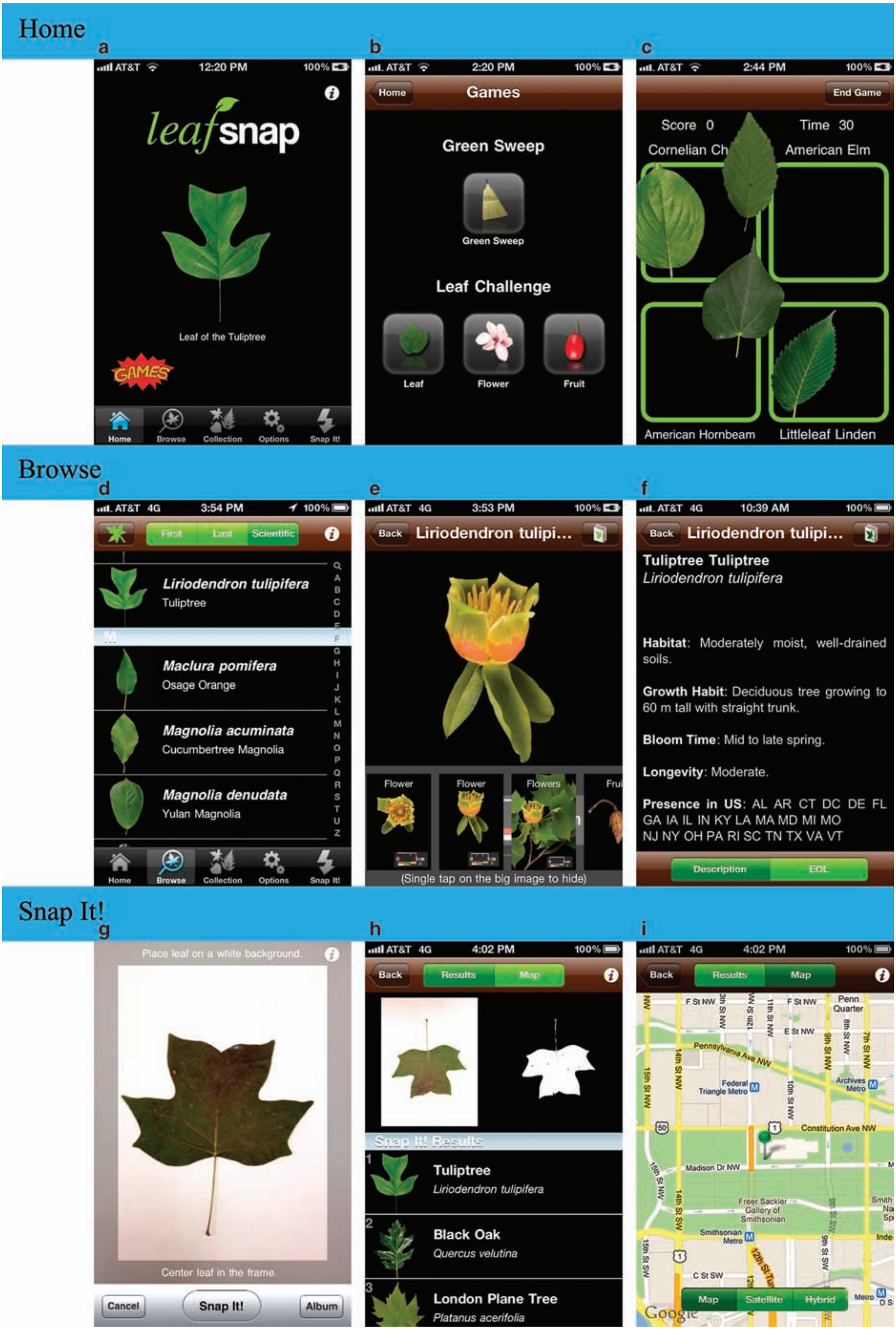
Acer saccharinum silver maple
Acer saccharum sugar maple
Acer spicatum mountain maple
Aesculus hippocastanum horse-chestnut

Figure 1. The species-identification page of Go Botany asks the user to answer simple questions about the plant that is to be identified. The user clicks on a question in the left frame, which opens a dialogue box with a question and a helpful hint for observing the specimen. All botanical terms (in this example, node*) are provided with an illustrated definition on rollover, and the choices of character states are simply illustrated with diagrammatic drawings. In the background are photos that show images of the species in the results set. The user can also click "Get more questions" to be provided with more questions about features the user can actually see on the plant of interest. Source: New England Wild Flower Society,* http://gobotany.newenglandwild.org.

developed by the Smithsonian Institution, in partnership with Columbia University and the University of Maryland. Leafsnap emphasizes interactivity: The user takes a photograph of a single leaf, using the built-in camera of their iPhone or iPad, and Leafsnap then compares the photograph to a central library of more than 9000 images stored in a remote database. Leafsnap automatically determines the contours of the leaf and uses visual-recognition software to find a match for it in the database (Agarwal et al. 2006, Belhumeur et al. 2008); results are returned to the user in 5–20 seconds, depending on the speed of the network connection. Next, Leafsnap brings up high-resolution images of the leaf, along with images of the species' flower, fruit, seeds, and bark. The app also supplies background information on the species and its geographic distribution. When the identification is not straightforward, Leafsnap users dig into other related images in its database, such as fruit shape or leaf venation patterns. In the end, it is up to the user to make the correct determination of the species, which reinforces botanical learning. Once a user successfully identifies a tree, his or her photograph and accompanying GPS location data are automatically uploaded into Leafsnap's database, contributing to the work of a community of scientists who are using the stream of data to map and monitor how the abundances and geographic ranges of different tree species are changing through time and as a function of climatic change.

Building online communities to enhance public engagement with science

Widespread popular interest in natural history and the availability of next-generation field guides is facilitating the growing engagement of citizen scientists with the professional scientific community (Newman et al. 2012). Apps such



Tuliptree Tuliptree
Liriodendron tulipifera

Habitat: Moderately moist, well-drained soils.

Growth Habit: Deciduous tree growing to 60 m tall with straight trunk.

Bloom Time: Mid to late spring.

Longevity: Moderate.

Presence in US: AL AR CT DC DE FL GA IA IL IN KY LA MA MD MI MO NJ NY OH PA RI SC TN TX VA VT

1 **Tuliptree**
Liriodendron tulipifera

2 **Black Oak**
Quercus velutina

3 **London Plane Tree**
Platanus acerifolia

as Leafsnap, nationwide efforts such as the USA National Phenology Network (Schwartz et al. 2012), and international projects such as the Tropical Ecology Assessment and Monitoring Network (Andelman 2011) are generating high-quality, georeferenced data on species' diversity, shifting distribution patterns, and life histories. Go Botany has a citizen-science portal, called PlantShare, where plant enthusiasts can share sightings, get crowdsourced or expert advice on identifying plants, and create checklists. The Cornell Lab of Ornithology's eBird Web site (*http://ebird.org/content/ebird*) encourages users to report bird sightings for use in a citizen-science project documenting changes in species ranges.

Similarly, BugGuide (*http://bugguide.net*), an online, taxonomically organized, and image-rich guide to the insects and arthropods of North America, is a resource for people who enjoy learning about and sharing observations of insects and other arthropods. From BugGuide's launch in 2003 to the end of 2012, it reached more than 4 million unique visitors; image submissions have increased even faster. BugGuide currently includes more than 26,000 species on more than 42,000 pages and nearly 580,000 images that have been contributed by nearly 22,000 individuals. A distinctive aspect of BugGuide is the ability of a user to request an identification of an unknown specimen; it is 1 of only 2 of the 16 applications that we reviewed that does so (table S1). Volunteer editors and taxonomic experts monitor the request queue and identify the specimens (i.e., images); most are identified within 1 day of submission. After identification, the images are moved out of the request queue and into BugGuide itself, ending up on individual-taxon information pages created by a BugGuide editor. Each taxon is arranged within a taxonomic hierarchy, and each level of that hierarchy has its own information page. These pages contain contributed images and information on a range of topics, including species diversity, key characteristics, distribution, and ecological characteristics, which can be used to generate distribution maps and summaries of phenological information. BugGuide has become a popular online resource for enthusiasts of the study of insects, not only because of its content and ease of use but also because of its welcoming atmosphere for both scientists and citizen scientists.

Discover Life (*www.discoverlife.org*) also links next-generation field guides with citizen science to yield high-quality scientific data. This integrated science and education platform, currently used by more than 350,000 people every month, enables users to collect and analyze data on the identity, distribution, and abundance of organisms; to conduct original research; and to learn science. Discover Life's integrated tools include more than 600 multiple-access identification guides to, and checklists for, groups including plants, vertebrates, fungi, and many arthropods; a global mapper that displays the distribution of more than 480,000 species; and quantitative tools to assess changes in phenology. User-created albums enable contributors to manage the data associated with photographs, to map where they photographed a species, and to maintain a digital list of their contributions. Since March 2010, for example, through Discover Life's moth project, phenological data on more than 1200 species from 130,000 photographs from North America and Costa Rica have been identified and analyzed.

With any data-collection effort, ensuring verifiable contributions is paramount. Discover Life research protocols require participants to include photographs of the time and date on their cell phone, of a GPS display, and of landmarks to confirm that the time and location are correct. Novices, experts, and computer algorithms work in concert to name specimens and correct errors associated with observations. Discover Life, eBird, and BugGuide, among many others, employ both professional and citizen scientists as moderators or gatekeepers for new data. As the number of users and the volume of their contributed data increases, more moderation will be needed. It is especially crucial to ensure that sensitive data on rare species (such as the locations of taxa vulnerable to poaching) and information that personally identifies specific users are protected.

Improving online tools with machine learning and crowdsourcing

Rather than simply consulting a printed field guide for help with identification, many people now attempt online searches. For example, more than 8 million people annually visit the Cornell Lab of Ornithology's All About Birds Web site (*http://allaboutbirds.org*); many use the search box in

Figure 2. The mobile application Leafsnap consists of a number of interactive screens that provide the user with information about tree species and with the ability to automatically identify a tree by taking a photograph of an isolated leaf. (a) Leafsnap Home screen. (b) Leafsnap provides two types of games to hone the skills of the user in identifying species from leaves, flowers, or fruits. (c) Green Sweep challenges the user to place a moving leaf into the correct species box. (d) The Browse mode allows the user to scroll through thumbnails of leaves, flowers, or fruits of the species included in the application; species can be sorted by common or scientific name. (e) For each species selected in the Browse mode, high-resolution photographs of all parts of the plant are illustrated to help in identifications. (f) The Browse mode also provides a short text description of the plant. (g) The Snap It! mode allows users to take a photo on a white background of a leaf of the unidentified species. (h) The shape of the leaf is automatically separated from the background and sent to the home server; within a short time, a list of prioritized identifications is sent back to the user. (i) Once the proper identification is selected by the user, the name of the species and the location coordinates are recorded in the user's own collection page in Leafsnap and on the central Leafsnap server. Source: W. John Kress, Leafsnap.

All About Birds as an identification tool, typing descriptors such as “small bird with black stripe.” But search engines are not built as identification tools (the semantics differ) and often return incomplete or misleading results. Therefore, the Cornell Lab of Ornithology is developing the Merlin bird-identification tool. Merlin asks people questions about their sightings and uses citizen-science data, crowdsourced descriptors, and artificial intelligence algorithms to enhance the semantic definitions within All About Birds, thereby improving the search results. Merlin first asks users when and where they saw the bird, then taps into the eBird citizen-science database, which contains more than 100 million observations from birders. Merlin thus narrows the number of possible species seen by 75% or more to the set of taxa most likely to be encountered at any given location and time of year. Merlin then helps users further refine their identifications on the basis of attributes they saw, such as color, size, and behavior. However, people notice, recall, and describe the same details many different ways, including ways that do not match the descriptors in a database provided by experts (e.g., Kempen and Tredoux 2012); the end result can be a misidentification. To improve the returns of searches based on variable responses, Merlin uses artificial intelligence algorithms to consider a user’s prior responses to inform the next question it asks, in much the same way that the Go Botany algorithm does. It uses probabilities to tap into a database populated with expert descriptors and crowdsourced descriptors gathered through online activities such as Mark My Bird, in which people describe the traits of birds on the basis of photos. Similar to how they do with Leafsnap, machine-learning algorithms enable Merlin to get “smarter” and return more-accurate identifications as more people use it.

Merlin uses visual-recognition software developed by the Visipedia project (*www.vision.caltech.edu/visipedia*), which analyzes crowdsourced data to help computers recognize objects in images. Massive amounts of data can be gathered in a short time by engaging the online communities of citizen-science projects such as eBird, social media, and other Web sites. Photographers have contributed more than 80,000 annotated photographs of birds, which were used to develop the visual recognition system. Mark My Bird and other online activities have gathered more than 250,000 rounds of data in 6 months from volunteers who are “teaching” Merlin about the color, size, and shape of birds as the public perceives and describes them. If Merlin proves successful, the techniques used to develop it will be adapted for other taxa, providing a new generation of online identification tools.

Bringing your field guide to the public

Field guide developers may think that they have produced the perfect app, but people will use it in unpredictable ways (such as using search entry fields to look for “black ants with large gasters”). Ultimately, the process of creating a next-generation field guide that communicates reliable information and that people will want to use depends on four fundamental steps: (1) clearly identifying the target audience; (2) conducting iterative user testing with that audience during the design and development of the app; (3) ensuring that the data are protected from inappropriate reuse; and (4) building long-term resources to maintain the app, update the data, and respond to changing technology.

Although user testing is a *sine qua non* of software and Web-site development, opaque and difficult to navigate Web sites continue to crowd the Internet. In clearly articulating the characteristics of the archetypal user of a next-generation field guide, it is useful to develop a persona (*sensu* Cooper et al. 2007) based on interviews of prospective users or empirical evidence that describes the user’s motivation, computer expertise, and level of experience with the taxonomic group and terminology. This persona should be a portrait sufficiently rich that the design team can determine whether a user matching the persona would use a certain feature. Iterative, objective user testing must be conducted to help refine and improve preliminary designs (*wireframes*) for a field guide. In such tests, the wireframes are shown to a set of users—not affiliated with the project—who broadly represent the personas identified by the design team. User tests are reality checks in the design process: They help the design and programming team overcome internal biases and ensure that the application will be user friendly and widely adopted.

Copyright protection for intellectual property, including photographs and illustrations, should be in place, and image contributors must have mechanisms for permitting the reuse of their work. Creative Commons licenses are frequently used as a means to control the reuse of proprietary data, and sets of best practices for noncommercial uses are available (Hagedorn et al. 2011). Before building any new Web app, the developers should research existing patents to make certain that unintentional infringement does not occur, especially if the product is intended for sale.

The Web can be a place where good ideas go viral and flourish or a burying ground for obsolete information or Web sites. Developing versatile apps that are adaptable to changing platforms, browsers, operating systems, and other software is an expensive, long-term enterprise. Long-term support is crucial to ensure the longevity of a next-generation field guide; it is also crucial early in the life of a project to articulate a vision for generating income in order to guarantee a long life for the product. Start-up funding, such as grants provided by the National Science Foundation, can jump-start a new initiative, but it is typically temporary. Users can be engaged, Wikipedia style, to contribute time and money to sustaining Web sites, but donations can decline as interest wanes. Citizen scientists and other users are most likely to stay engaged when the scientists who developed the application communicate clearly how user-generated data improve science. Bearing these considerations in mind, electronic field guides are essential, evolving tools that will transform how professional and amateur biologists collaborate to identify new species and move science forward.

Acknowledgments

Funding for this work was provided by National Science Foundation grants no. 0840186 to EJF; no. 1010818 to MC; no. 0325867 to WJK; no. 9808462, no. 0111540, and no. 0416835 to RDS; and no. 1136646 to AME. We also acknowledge funding from the Smithsonian Institution, Columbia University, and the University of Maryland to WJK; the Ambrose Monell Foundation, the Vetlesen Foundation, and the US Geological Survey to JP; and US Department of Energy grant no. DE-FG02-08ER64510 to AME. MC thanks Jessie Barry at the Cornell Lab of Ornithology for additional information about the Merlin project. WJK thanks the Leafsnap team, including Peter Belhumeur, David Jacobs, Sean White, Ida Lopez, Steve Finer, the people at Finding Species, and many students and volunteers.

References cited

Agarwal G, et al. 2006. First steps toward an electronic field guide for plants. Taxon 55: 597–610.

Andelman SJ. 2011. Conservation science outside the comfort zone. Nature 475: 290–291.

Angles R, Gutierrez C. 2008. Survey of graph database models. ACM Computing Surveys 40 (art. 1).

Belhumeur PN, et al. 2008. Searching the world’s herbaria: A system for visual identification of plant species. Pages 116–129 in Forsyth D, Torr P, Zisserman A, eds. Computer Vision—ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12–18, 2008, part IV. Springer.

Brach AR, Song H. 2005. ActKey: A Web-based interactive identification key program. Taxon 54: 1041–1046.

———. 2006. eFloras: New directions for online floras exemplified by the Flora of China project. Taxon 55: 188–192.

Cooper A, Reimann R, Cronin D. 2007. About Face 3: The Essentials of Interaction Design. Wiley.

Dallwitz MJ. 2010. A Comparison of Formats for Descriptive Data. Institute of Botany, Chinese Academy of Sciences. (2 August 2013; *http://delta-intkey.com/www/compdata.htm*)

———. 2011. A Comparison of Interactive Identification Programs. Institute of Botany, Chinese Academy of Sciences. (2 August 2013; *http://delta-intkey.com/www/comparison.htm*)

Dallwitz MJ, Paine TA, Zurcher EJ. 2013. Principles of Interactive Keys. Institute of Botany, Chinese Academy of Sciences. (2 August 2013; *http://delta-intkey.com/www/interactivekeys.htm*)

Dunlap T. 2005. Tom Dunlap on early bird guides. Environmental History 10: 110–118.

Edwards M, Morse DR. 1995. The potential for computer-aided identification in biodiversity research. Trends in Ecology and Evolution 10: 153–158.

Ellis JD. 1984. A rationale for using computers in science education. American Biology Teacher 46: 200–206.

Givens JA. 2006. Reading and writing the illustrated *Tractatus de herbis*, 1280–1526. Pages 115–145 in Givens JA, Reeds KM, Touwaide A, eds. Visualizing Medieval Medicine and Natural History, 1200–1500. Ashgate.

Gobi AF. 2010. Towards generalized benthic species recognition and quantification using computer vision. Pages 94–100 in the Proceedings of the Fourth Pacific-Rim Symposium on Image and Video Technology: PSIVT 2010. IEEE Computer Society. doi:10.1109/PSIVT.2010.23

Goodall DW. 1968. Identification by computer. BioScience 18: 485–488.

Hagedorn G, Mietchen D, Morris RA, Agosti D, Penev L, Berendsohn WG, Hobern D. 2011. Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity information. ZooKeys 150: 127–149.

Haines A. 2011. Flora Novae Angliae: A Manual for the Identification of Native and Naturalized Higher Vascular Plants of New England. Yale University Press.

Hawthorne W, Harris S. 2006. Plant names and botanical publications. Pages 61–90 in Lawrence A, Hawthorne W, eds. Plant Identification: Creating User-Friendly Field Guides for Biodiversity Management. Earthscan.

Kempen K, Tredoux CG. 2012. “Seeing is believing”: The effect of viewing and constructing a composite on identification performance. South African Journal of Psychology 42: 434–444.

Knapp S, McNeill J, Turland NJ. 2011. Changes to publication requirements made at the XVIII International Botanical Congress in Melbourne—What does e-publication mean for you? BMC Evolutionary Biology 11 (art. 250).

Kress WJ. 2004. Paper floras: How long will they last? American Journal of Botany 91: 2124–2127.

Labeda DP, Oren A. 2008. International Committee on Systematics of Prokaryotes; XIth International (IUMS) Congress of Bacteriology and Applied Microbiology: Minutes of the meetings, 23, 24, 26 and 28 July 2005, San Francisco, CA, USA. International Journal of Systematic and Evolutionary Microbiology 58: 1746–1752.

Leggett R, Kirchoff BK. 2011. Image use in field guides and identification keys: Review and recommendations. AoB Plants 2011 (art. plr004). doi:10.1093/aobpla/plr004

Lytle DA, Martínez-Muñoz G, Zhang W, Larios N, Shapiro L, Paasch R, Moldenke A, Mortensen EA, Todorovic S, Dietterich TG. 2010. Automated processing and identification of benthic invertebrate samples. Journal of the North American Benthological Society 29: 867–874.

MacLeod N, ed. 2008. Automated Taxon Identification in Systematics: Theory, Approaches and Applications. CRC Press.

Morris RA, Stevenson RD, Haber W. 2007. An architecture for electronic field guides. Journal of Intelligent Information Systems 29: 97–110.

Morse LE. 1971. Specimen identification and key construction with time-sharing computers. Taxon 20: 269–282.

Newman G, Wiggins A, Crall A, Graham E, Newman S, Crowston K. 2012. The future of citizen science: Emerging technologies and shifting paradigms. Frontiers in Ecology and the Environment 10: 298–304.

Pankhurst RJ. 1991. Practical Taxonomic Computing. Cambridge University Press.

Scharf ST. 2007. Identification Keys and the Natural Method: The Development of Text-Based Information Management Tools in Botany in the Long Eighteenth Century. PhD dissertation. University of Toronto, Toronto, Canada.

———. 2009. Identification keys, the “Natural Method,” and the development of plant identification manuals. Journal of the History of Biology 42: 73–117.

Schmidt D. 2006. Field guides in academe: A citation study. Journal of Academic Librarianship 32: 274–285.

Schroeder CM, Scott TP, Tolson H, Huang T-Y, Lee Y-H. 2007. A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. Journal of Research in Science Teaching 44: 1436–1460.

Schwartz MD, Betancourt JL, Weltzin JF. 2012. From Caprio’s lilacs to the USA National Phenology Network. Frontiers in Ecology and the Environment 10: 324–327.

Simpson DR, Janos D. 1974. A Punch Card Key to the Families of Dicotyledons of the Western Hemisphere South of the United States. Field Museum of Natural History.

Stevenson RD, Haber WA, Morris RA. 2003. Electronic field guides and user communities in the eco-informatics revolution. Conservation Ecology 7 (art. 3). (5 August 2013; *www.consecol.org/vol7/iss1/art3*)

Tapscott D. 1998. Growing Up Digital: The Rise of the Net Generation. McGraw-Hill.

Walls RL, et al. 2012. Ontologies as integrative tools for plant science. American Journal of Botany 99: 1263–1275.

Zhang Z-Q. 2012. A new era in zoological nomenclature and taxonomy: ICZN accepts e-publication and launches ZooBank. Zootaxa 3450: 8. (5 August 2013; *www.mapress.com/zootaxa/2012/f/zt03450p008.pdf*)

Elizabeth J. Farnsworth (efarnswo@mtholyoke.edu) is the senior research ecologist with the New England Wild Flower Society, in Framingham, Massachusetts. Miyoko Chu is the senior director of communications at the Cornell Lab of Ornithology, at Cornell University, in Ithaca, New York. W. John Kress is the director of the Consortium for Understanding and Sustaining a Biodiverse Planet, in the Office of the Under Secretary for Science, and a curator and research scientist in the Department of Botany at the Smithsonian Institution, in Washington, DC. Amanda K. Neill is the director of the Herbarium, and Jason H. Best is a biodiversity informaticist, at the Botanical Research Institute of Texas, in Fort Worth. John Pickering is an associate professor of biology at the Odum School of Ecology, at the University of Georgia, in Athens, and the founder of Discover Life. Robert D. Stevenson is an associate professor of biology in the Department of Biology at the University of Massachusetts Boston. Gregory W. Courtney is a professor and the curator of the Iowa State Insect Collection, and John K. VanDyk is a systems analyst and adjunct assistant professor, in the Department of Entomology at Iowa State University, in Ames. Aaron M. Ellison is the senior research fellow in ecology at Harvard University’s Harvard Forest, in Petersham, Massachusetts.