Requirements for Electronic Field Guides

Emily Roseanne Schwarz
California Polytechnic State University
San Luis Obispo, Ca 93407

Abstract
Development in the United States and around the world is changing our lands in ways most people may not notice. Invasive species have literally changed the landscape. Development has destroyed or altered habitats far beyond the boundaries of the cities. Many species of native plants are extinct or on the verge of extinction right in our own backyards. Identification of specimens in the field is necessary for taxonomists to collect data on the characteristics and distribution of species. Field guides are the de-facto method of identification and source of information for professionals and amateurs in the field. Efforts are underway at many institutions to use advances in computer technology to create electronic field guides. These systems are improving the speed and accuracy of identification. Not only can they present taxonomic information but some will be able to collect scientifically useful information from users.

Our focus is primarily on the development of electronic field guides for plants; however, our work could be generalized to other taxons. We seek to answer to the question: how can computers best facilitate the identification of plants and further the related goals of presenting and collecting taxonomic data? To that end, this paper presents an analysis of existing electronic field guides and technologies available which could accomplish these goals. It is our hope this paper will inform the requirements and design decisions of future identification systems. This analysis forms a framework through which existing and future systems can be understood.

Introduction
Recording the characteristics and distributions of various plants is essential to their future conservation [Kress, 2004]. This is true not only in the tropics [Prance, 2001] but in the rest of the world as well. For instance, many of the endangered plants listed in [Tiber, 2001] would benefit from better data despite residing near some of the largest population centers in North America.

Identification is the first step needed to gather data on a specimen in wild. Identification resources are also critical in verifying the discovery of a new species. This data gathered on both new and existing species helps us better understand and protect the diversity of our natural world [Tiber, 2001]. However, the identification of an unknown specimen is currently time consuming and difficult [Agarwal et al., 2006]. It also requires specialized knowledge and skills that only highly trained experts possess. In the developing world this difficulty in identification can hinder our progress to discover new species before they become extinct. In the developed world better identification techniques could improve the cost effectiveness of field inventories and on site data collection both of which are important tools in effective land management and species conservation [Tazik et al, 1992]. Additionally, quality identification methods can increase public enthusiasm and support for conservation and enjoyment of our natural world [Junzen, 2003].

To help overcome the shortfalls within existing identification methods by improving their speed, accuracy, and completeness, many institutions are developing computer systems to aid in identification [Agarwal et al, 2006][Kirchoff et al., 2008]. These technologies require the interaction of the user with a computer program to help the user determine a correct identification for an unknown specimen. While identification of unknown specimens is the primary function of these systems, they can also be used to further other goals related to plant conservation and education. By providing a framework for the cataloging of existing information and allowing users to collect additional information, these systems have the potential to become indispensable tools for botanists and a variety of other users. In this paper, we will distinguish between two types of field guides. Electronic field guide describes a computer system that aids the user in identifying specimens and presents taxonomic knowledge to the user. Paper field guide refers to a book that accomplishes the same tasks. While the term field guide may imply portability, this is not a constraint on the systems thus classified. As we shall discuss later, portability is a desired characteristic but not included specifically within the definition, particularly because the nature of paper field guides is often far from portable.

The purpose of this paper is to evaluate existing systems and technologies by judging their progress in identification, presentation of knowledge, and other related functionality. In order for this evaluation to take place we
take a deep look at traditional field guides and their users. From this exploration we present a list of requirements by which electronic field guides can be judged. Some of these requirements will conflict because we are examining the field as a whole and particular implementations will serve only a subset of the requirements found herein. We hope that this analysis will provide a framework for future implementations and advances in the field. To that end, we also provide our thoughts on these areas that are most in need of research or have had little coverage thus far.

This remainder of this paper is divided into three parts. In “Background” we discuss the stakeholders, users, paper field guides, and current identification techniques. In “Requirements” we discuss the requirements for electronic field guides. This section is divided roughly into different concerns where we discuss the rational for the requirement(s) relating to the concern, existing implementations, and technologies. Finally in “Conclusions” we discuss our views on the requirements as whole as well as our outlook on future work.

**Background**

The first popular field guide was published in the late nineteenth century, but more esoteric taxonomic literature existed long before that time [Pearson and Shetterly, 2006]. In this section, we discuss the users of paper field guides and some of the challenges associated with their use.

**Stakeholders**

The beneficiaries of improved systems fall into two groups, organizations and users. It is useful for us to define and describe these stakeholders to better understand their goals and motivations and thus how current and future systems address their needs.

**Organizations**

Multiple types of organizations would benefit from improved methods of plant identification and increased accessibility to botanical knowledge. Governments, such as the United Kingdom, are investigating Internet-based taxonomy [Bishy et al., 2009]. This should come as little surprise since taxonomic data is seen as the foundation for more advanced ecological research [Shyamal, 2007]. Here in the United States the NSF is also funding programs to digitalize taxonomic information [Agarwal et al., 2006].

Academic institutions that do botanical research also see the benefit both from a scientific and educational standpoint. Herbaria and museums have much to gain from this technology as they are also a seat of botanical research. Additionally herbaria see added value in systems that provide a framework for the digitalization of their often huge collections [Agarwal et al., 2006].

**Users**

The consumers of botanical knowledge are a diverse group. [Hickman, 1993] lists possible users as “beginning students, self-taught amateurs, environmental consultants, native-plant gardeners, employees of diverse government agencies, ecologists, and academic systematics.” While far from complete, this gives a rough picture of the diverging motivations and interests of potential users of field guides. Regardless of their differences, all users can be roughly divided into two groups, experts and novices.

In this paper we will use the term **expert** to describe users with the following characteristics. They are skilled in botanical terminology and traditional identification methods [Kirchoff et al., 2008][Agarwal et al., 2006]. They have a high degree of visual expertise [Kirchoff et al., 2007]. They are able to provide systems with information based on their expertise. For example, while they may not know the species or genus of a specimen, they will likely know the family. Systems can then use the experts’ knowledge to narrow viable identifications [Agarwal et al., 2006].

**Novices** on the other hand lack knowledge of botanical terminology and are unfamiliar with traditional identification methods, such as the dichotomous keys discussed in the next section. Because novices do not usually work within plant taxonomy, they have little incentive to learn the difficult and arcane botanical terminology [Kirchoff et al., 2008]. They have some visual knowledge, but it is incomplete [Agarwal et al., 2006]. Novices might be botanical hobbyists, or they could be researchers in fields related to botany that do not require the training in botanical terminology or practice. It should also be noted that experts in one particular area of taxonomy might have the characteristics of a novice if that are working outside their area of expertise.

In some areas of taxonomy, novices may out number experts [Shyamal, 2007][Pearson & Shetterly, 2006]. This makes reaching novices a potential improvement in identification technology because the best identification systems are currently reserved for experts. Harnessing the power of the novice has the potential to revolutionize all areas of taxonomy [Janzen, 2009]. Public education is a common goal among conservation organizations as there are many opportunities for novices to make a difference in land management through volunteer work and supporting conservation policies [Tiber, 2001].

**Types of Paper Field Guides**

To understand the identification of plants we must first present the term **character**. Characters are discrete statements about a taxon. They relate a specific property to a specific state. For example, when describing a specimen we could say “leaves alternate.” This relates the property “leaves” to the state “alternate” which is defined as leaves appearing one per node on a stem.

There are two basic patterns of use in traditional plant identification. The first is the use of single (or few) character field guides. This approach is generally used by novices has received very little attention in the academic literature. This approach can be seen in [Spellenberg, 2001]. The user examines a single obvious portion of the plant for a small number of easy-to-recognize characters.
In [Spellenberg, 2001] these are the color and shape of the flowers. The information within the guide is indexed by these features. This kind of identification is problematic because it is difficult to include large numbers of species. If the user finds a yellow, radially symmetrical flower, the user must scan through all of the yellow, radially symmetrical flowers to find the one he or she is looking for.

Single character field guides are often filled with photographs or drawings to make searching easier [Spellenberg, 2001]. However, they can not present a very large number of species because of space needed for images. Moreover, as the number of images increases so does the time it takes the user to find a particular image. Guides make a trade off in covering large geographical areas at the expense of including less frequent species [Spellenberg, 2001] or in covering small geographical areas more thoroughly (see [Kiel, 2009] for a highly localized guide). This makes single character field guides a tool for those users interested in either identifying only the most common species or those willing to make approximate identifications to the nearest “similar” species. This approximate identification is not adequate for scientific or governmental processes because it is not uncommon for an endangered species to share habitat with a nearly identical common species. One such case would be *Layia jonesii* protected under the Endangered Species Act and *Layia platyglossa* its common cousin, both of which are found growing together in serpentine outcroppings [Tibor, 2001][Hickman, 1993]. Single character field guides tend to focus on a single group of popular species such as “wildflowers,” or trees. Because they are published privately and marketed to amateurs, single character field guides are rarely published on hard to identify or “ugly” plants such as grasses or scrappy gray bushes. Again this makes them of little use for the needs of professionals.

Because of the limitations with single character guides students and researchers use guides with large multi-character dichotomous keys. These keys are set up as a series of mutually exclusive pairs of statements, essentially a binary tree. The user determines the correct statement and proceeds to the next statement along the correct branch. Rarely both statements may be true; in this case the key will have two leaves that arrive at the same taxon. [Hickman, 1993] requires at most 28 decisions to key any plant in California to its proper family but on average requires about 9 or 10. Because of space considerations in guides that are very inclusive, abbreviations are common as well as technical terminology. This terminology also helps reduce ambiguity. For example, it is more efficient and more exact to refer to a “bract” than a “small, leaf- or scale-like structure associated with an inflorescence or cone.” While a few such terms might be picked up by a novice, the thirteen page glossary and two pages of abbreviations assures most descriptions in the manual can only be understood by an expert.

Space concerns are a major concern in professional field guides as much as popular ones. The need to include every species means that there are no photos, rather, line drawings which feature only a selection of the included plants and of those many include only a few characters (a seed, a leaf, a fruit). Since correct identification relies on written keys, the user’s ability to answer the questions asked with complete precision is necessary. Questions often ask about minute features that the user must have a microscope or field lens to see. Or about features that could be absent on a specimen, such as the fruit if the specimen is not currently fruiting. This lack of flexibility makes it easy to see why such keys are a huge barrier for novices. A single wrong answer precludes finding the correct identification, and even if a possible identification is found, the description may be difficult to understand and the picture may be missing or incomplete.

Not all field guides fit nicely into the divide between single- and multi-character types. Some field guides many use no characters at all. [Lightner, 2006] breaks his work into two groupings trees and shrubs and then herbs. The groups are broken into board categories of taxonomic classification, gymnosperms, dicots, and monocots. From there families are listed alphabetically with species alphabetical as well. Experts might be able to navigate this network, assuming they can get down to family and then browse, but amateurs will have little luck here. Interestingly this system allows [Lightner, 2006] to be free of the need for a single identifying character to unite the content which allows him to include an amazing variety of species that beyond the typical wildflowers and trees.

**Purpose of Field Guides**

To understand the proposed use of online field guides, we must understand the use of traditional ones. [Kress, 2004] states that paper floras have essentially four purposes: inventory, identification, description, and classification. Inventory is defined as knowing what taxa are present in a given area. This is important information for those wishing to study particular taxa as restricted to certain areas. Understanding what occurs where is crucial for more advanced research [Shyamal, 2007]. Identification is defined as knowing the names of the taxa we find or study. Description is defined as presenting us with knowledge of taxa and classification as knowing how the taxa are related to each other and those outside the region. While classification can be viewed as a subset of description, it is more helpful to see them as presenting two different types of information. Description is the information relating to the properties of the described entity, its characters. Classification presents information about taxonomic relationships. Classification is important because it gives the user a way to relate a new organism into an existing framework. This helps indemnification by enabling users to only remember how an organism differs from other closely related organisms, rather than memorizing all its characters. It can also provide missing information such as if the plant is likely to be poisonous, its possible pollinators, or economic value.
Identification

Identification is the core task for both paper and electronic field guides. Identification means assisting the user in obtaining the identity of a given unknown plant specimen to at least the species level and if available the subspecies or variety.

Because a human user is interacting with the system, it is not necessarily required to make the full identification itself; it could present the user with suggestions rather than a single determination [Agarwal et al., 2006]. This is appropriate since in many systems, the user will have much more information than the system as to the nature of a specimen. This is also reasonable; as both experts and novices have some degree of visual knowledge. If they are presented with visual information, they are likely to be able to choose the correct identification out of a set of possible species.

Requirements for Identification

Completeness of Species Coverage

To meet this functionality a system must include all or a significant portion of the species that occur or possibly occur its coverage area. Guides quickly lose their usefulness when there is not a reasonable expectation that the species of an unknown specimen will be found within the guide. If a guide is not expected to contain all species from the region it covers it should include those species which are most common and thus most likely to be encountered.

Although we must often settle to estimate completeness, there is several ways it could be calculated. Completeness could be calculated as the included species divided by the total species in the area covered. It could also be calculated by adding together all the percentages of the total flora each included plant within the guide makes up. For instance if three plants made up the flora of an area and there are 50, 25, and 25 individuals respectively their percentages would be 50, 25, and 25. A guide containing the first and second species would be 75% complete. However, a guide containing the second and third would only be 50% complete. While this is much more difficult to calculate than the first measure it is a better measure because it takes into account the chances of needing to identify each species. Completeness might be extended beyond 100%. For example, floras of specific locations could be augmented with neighboring floras, in case the exact distribution of a taxon is wider than previously believed. Distributions in certain fields or areas can be unreliable [Shyamal, 2007] or change over time so the overuse of localizing information in identification can be detrimental completeness.

It is tempting to calculate completeness as the number of species divided by the area covered. This would let you see the number of species per mile covered to show how densely the guide covers the area. However, different locations have varying numbers of species to begin with and this number varies widely. For instance the great plains are four times the size of California but contain only half the species [Holland and Keil, 1995]. Therefore comparisons in completeness calculated this way would be inaccurate.

Completeness is an important factor missing not only in single character field guides but also in existing electronic solutions [Pickering 2009]. Completeness is the most important factor in a field guide’s usefulness as it defines the outer boundary of how useful a system can be. Since most electronic field guides are in the development stage they have little coverage and are therefore of little use to most users. Not only is physical space a limit on coverage of paper fields guides but time to collect information is also a limiting factor. Botanical knowledge is far from stable even in North America and species may change their genus or distribution. This along with redefinition of species boundaries can cause keys to need to be revised and new taxonomic descriptions to be written.

An important distinction in completeness is between the current completeness and the level of effort required to make the system more complete. For example, the process to update a book is usually more effort than updating a website. Likewise, not all electronic systems are equally easy to make more complete.
Completeness plays a role in ascertaining if the specimen represents an undiscovered species. The ability to tell if a specimen represents and undiscovered species is suggested in [Agarwal et al., 2006]. However, they fail to discuss how this is different from the task of identification. Here completeness is important for a system to be confident that a species is undiscovered rather than just missing from the system’s coverage.

**Corroborating Information**

Another facet of identification is presenting information about the possible identification species. This information could be the characters of the species, descriptions of its habitat, or other facts about the species. This feature appears prominently in many systems especially paper field guides.

It may seem that this feature is more important from the perspective of education, but it is necessary to identification as well. Having corroborating information allows the user to test the identification by comparing the information given to the specimen they have at hand. This information could be provided as text, photographs, or other media. It is important because it gives the user the ability to verify the identification they have tentatively arrived at. This is especially important for a scientist out in the field doing research where reaching a correct answer is the most important and the most uncertain. The information presented to the user can also serve other purposes: to educate the user about the specimen and direct them towards additional resources.

The online identification website Discover Life [Pickering, 2009] focuses solely on identification but presents relatively little information on the species identified. Rather they present an amalgamation of varying online resources. This can be a problem because the pages contain some corroborating information it is difficult to navigate. Separating this information from the identification makes the process less efficient because reviewing corroborating information is an important part of the identification task.

**Portability**

Portability is another factor in developing an identification system. The system will be most useful if it is available to users in the field [Kress, 2004][Agarwal et al, 2006]. Portability is already becoming a reality with more powerful PDAs and smaller laptops. Portability is about more than just convenience. While a good set of photographs should be sufficient for the identification of a specimen later on, the best time to perform identification is in the field [Baskauf and Kirchoff, 2008].

The Discover Life website also embodies the idea of portability, since it can be accessed anywhere. However, its current frame-based design is not particularly well suited to mobile devices. In general, websites can be thought of as more portable than downloadable programs such as IntKey [Dallwitz et al, 2007] since they are able to run on more devices. This represents a trade off though as downloadable programs also have an edge because they do not require an internet connection to function.
Portability includes two factors. The first of which is the weight and size of the identification system. This is important because the specimen to be identified is often located in remote areas. Amateurs and professions alike will not carry more weight into the wilderness than they believe absolutely necessary. The second factor is usability in the wilderness. This can mean that a system that needs an internet connection may be less useful than one that can go without. Also if a system only runs on certain hardware and that hardware has a limitation such as low battery life that may make the system less usable in the wilderness.

Ease of Reaching Identification

Another consideration is how laborious it is to reach identification. This can be measured by time as in [Agarwal et al. 2006] or by fewest steps as in [Kirchoff et al. 2008]. Users of the system will likely identify many specimens over a single day in the field, so any savings over individual identifications will be compounded many times. Neither time nor number of steps is adequate by itself to fully represent the toll on the user. For example, a system could take more time than another, but if during that time the user is not required to interact with the system, then that system might be preferred. These tradeoffs are best studied more extensively in the context of user interface research. However, designers of such systems should keep these tradeoffs in mind to maximize their initial efforts.

[Kirchoff et al., 2008] argues that visual identification is the easiest and most natural form of identification. Because even novices have some visual expertise, keys that use drawing and photographs to aid in identification will have the most success. Their system displays images until their statistical framework calculates that they have reached a possible identification. Their vision is to build a complete expert system to decide which images to display to the users based upon inputs from experts in the field.

In most systems the ease of identification is inversely correlated with the total species in the system. For dichotomous keys, the fewer species means fewer character questions need to be asked. Likewise, [Kirchoff et al., 2008]’s solution will likely become more laborious as the number of species grows. One potential way to keep the number of questions down in any character based is to use the information gathered to inform what questions are most appropriate to the remainder of the search. [Pickering, 2009] does this by filtering out those questions which all the remaining candidates fall under a single answer. This approach is simple but effective. However, there seems to be room for improvement by adding additional methods to select which questions are the most appropriate.

Accuracy

Accuracy is another concern although it can be mitigated by providing the user with enough information to check the identification. There has been some work in measuring the accuracy of the systems. For example [Kirchoff et al., 2008] approached their system through mathematics. Their system by its very nature has a certain statistical reliability. [Agarwal et al., 2006] focused on user trails to judge the accuracy.

An important landmark will come when electronic field guides can demonstrate that users make more accurate identifications than using the best traditional paper guides such as [Hickmen, 1993]. This is a difficult measure to demonstrate though, as we have already said that systems that have fewer species are easier to use. Thus to properly compare accuracies the two systems may need to contain roughly the same number of species to avoid bias.

Flexibility

Users of multi character keys are completely at the mercy of the highly linear key. In most cases if they come to a statement pair for which they are unsure, they reduce their probability of finding the right answer by 50%. Rarely though when the answer to a particular question is particularly ambiguous the key author will include a "trap door" on the wrong side of the tree so to correctly identify a specimen given that wrong answer. This however can drastically increase the size of the key so it is used only in rare cases.

Electronic identification can mitigate these problems by allowing the user to select which character they would like to answer questions about [Dallwitz et al., 2007]. Some of the first efforts to improve electronic identification were along this line [Atkinson and Gemmerman, 1987]. Initial efforts used expert systems to provide the users a more flexible way of answering questions. The Discover Life [Pickering, 2009] and Int Key systems do not enforce an order on the ordering of questions.

The IntKey system also allows users to recover from errors by setting a tolerance level- the number of questions a possible identification is allowed to differ from the answers given by the user. The interaction of key writers with the system in [Pickering, 2009] is more similar to traditional keys in that it assumes the key writer will include as much flexibility as needed to anticipate wrong answers by the user. This is a large burden to put on the key writer and it contains the assumption that the key writer is an expert.

Interconnectedness

Electronic field guides have a potential to tap into resources online that paper field guides do not. This could allow electronic field guides to serve multiple purposes in a network of botanical resources.

When identifying a plant that might be new to science, the botanist must overcome numerous hurdles. Botanists must compare their specimen to known type specimens within herbaria. Herbarium plant collections are scattered all over the world and are not easily searchable [Agarwal et al., 2006]. In addition, botanists seeking to make comparisons between a specimen and herbarium specimens must know what they are looking for before hand, they must be specialists (or become specialists) to make sure their identification is valid. They must either travel to the herbarium collections or have the specimens shipped to them. This can be an extremely time consuming process.
These type specimens are far from the perfect means of identification they are touted as. Often they are incomplete and may not be sufficient for identification [Agarwal et al., 2006]. This highly manual process for identification can become more efficient if methods for identification are connected with other botanical resources.

**Identification Technologies**

There are two possible technologies that would revolutionize the task of identification if their full potentials were realized. Both of these technologies could potentially do more than assisting users in identification as they could completely identify organisms requiring minimal human interaction. The fact that neither has stopped researchers from developing methods for interactive identification points to the uncertainty about when such automated technologies will become practical. However, any future systems must be prepared to reevaluate these technologies and make determinations based on the current state of the art.

**Photographic Identification**

[Agarwal et al., 2006] is working on technology to identify plant specimens through the shape of their leaves. Users take a picture of the leaves and the system uses new shape recognizing algorithms to match them to the leaves in the system database. The system then displays the top matches for the user to choose from. This technology could be expanded by adding additional characters to recognize, such as flowers. Currently the leaf photos must be taken from a particular angle, essentially posed rather than attached to the plant. An improvement, possibly through 3-dimensional visualization would allow the system to identify the leaves anywhere in a photograph such as in a picture of the complete plant.

**DNA Barcoding**

DNA barcoding is the process of identifying an organism based upon a small segment of DNA from a particular location [Ausubel, 2009]. The location is determined by an area that is both easy to process and gives the best variation to allow for species level identification. Unlike the potentially inaccurate methods for photo identification, DNA barcoding has been shown to be highly reliable. Moreover, it has several advantages over methods that rely on morphology, including complete photographic identification as discussed above. These include not needing an intact or complete sample and the ability to differentiate species with few morphological differences [Packer et al., 2009].

The vision of DNA barcoding is presented in [Janzen, 2004], a world in which everyone has access to a “gadget” that can instantly identify any organism from a small amount of genetic material. In that world the users of such a gadget could be novices, hobbyists, everyday citizens, as well as experts. This would free the experts to pursue analysis while the novices could make useful contributions by collecting the data all around them. The author earnestly argues that, “better keys, more keys, more images on the Web, more Web sites, species pages, more descriptions, more phylogenies, more specimens, more anything,” is not truly sufficient to connect everyday people to the biodiversity all around them. Such ideas obviously have implications for the developers of keys, images, and websites.

And so we must ask another question: will there be any value in the keys, images, and websites developed for computer assisted identification once DNA barcoding becomes widely available? Judging from our earlier requirements, they retain much of their value. The part that is affected by barcoding is the identification method, which we have already stated must be as general as possible to encourage integration with new technologies. This does however mean that other identification methods might not be as fruitful as none hold the promise of such usability as DNA barcoding.

Progress in creating a DNA barcode for plants is lagging behind that of animals due to the difficulty of finding a DNA segment that meets all the criteria [Hollingsworth et al., 2009]. However, recent studies are working to build a consensus in this area, and it is likely that the DNA site for plant barcodes will be agreed on within the next year. Current methods for barcoding are making improvements in its speed and cost effectiveness. According to the Consortium for the Barcode of Life, the laboratory cost of processing a sample is a few dollars; however, such a laboratory requires a much larger initial investment for equipment [Cameron et al., 2006]. Generating a barcode can be done in as little as two hours [Ivanova et al., 2009]. However, this is a long way from the few minutes that researchers envision in the future. When that future will become reality is open to speculation.

[Cameron et al., 2006] points out that there is no reason to believe that barcoding will ever be as quick, portable, and cheap as proponents such as [Janzen, 2004] suggest. They also note that barcoding requires a tissue sample which makes it inappropriate for identifying certain kinds of specimens such as wild living animals. Depending on the size of tissue sample necessary it might be destructive to the plants sampled placing it at odds with conservation especially if it is available to everyone as [Janzen, 2004] suggests. Affordability is also an issue for its use by the general public. Researchers and government inspectors may be willing to pay a certain price for such a gadget but with out signification public interest in the product it may never reach general consumers. [Cameron et al., 2006] goes on to argue that such an interest just isn’t there, and the public would be better served by “well illustrated visual field guides.”

**Presentation and Collection of Knowledge**

An electronic field guide has an opportunity to become more than an online version of a paper field guide. A traditional field guide can only present users with information. Electronic field guides have the potential to become non-static, collecting new knowledge and
presenting the most current information. New DNA technologies are literally rewriting what we know about the relations between organisms every day; therefore, new systems must have the power to adapt to changing data sources least they become obsolete.

**Cataloging Existing Knowledge**

A wealth of information already exists in herbaria scattered around the world. This information needs to be digitalized so that it can be put to the most use [Agarwal et al., 2006]. A system with a strong means of identification could help in cataloging unidentified or misidentified specimens. Such a system would be useful if it could categorize or identify only a portion of the specimens autonomously and flag others for needing human processing.

[Baskauf and Kirchoff, 2008] make a strong point that plants dried and preserved in herbaria are not always in the best state for identification or teaching. They propose gathering sets of photographs to serve as digital specimens. These sets of photographs are better for identification and teaching because unlike single photographs, they can cover all important characters of a plant. They maintain their color unlike herbarium collections and because they are photographed in the wild they maintain information about their habitat and growth patterns. Digital specimens are a perfect marriage for electronic field guides allowing the system to display images in a scientifically rigorous manner.

**Enlisting Novices**

Field guides have a powerful effect on public interest in the subject they cover [Pearson, 2006]. Having a simple way to identify specimens allows novices to capture information that would normally have to come from experts. Enlisting novices to collect data frees up experts for higher level research. The concept of increasing novice participation in scientific endeavors is employed by the movement known as citizen science [Cooper et al., 2007].

Large scale information gathering over long periods of time is only practical for novices. [Shyamal, 2007] lists five factors which encourage novices to contribute in taxonomic data collection. While his work focuses on participation in India, these factors are unsurprisingly universal.

1. **Recognition**: ensuring credits for contributors.
2. **Opportunities for advancement**: enhancing the knowledge and skills of contributors by providing information and training.
3. **Demonstration of value**: demonstrating the value of individual data contribution is important. Computational systems can instantly compile new data and show the most up-to-date summaries.
4. **Rewards**: many contributors of earlier collaborative projects in India have been motivated by rewards such as free field guides (e.g. the Asian Wetland Bureau - Mid-winter waterfowl participants received reports and the complimentary field guides).
5. **Opportunities for social interaction and networking**: the possibilities of finding and interacting with contributors in the geographical vicinity.

Electronic field guides are in an excellent position to provide at least 4 of these incentives. Recognition can be achieved through linking of ones contributions to their personal profile. Opportunities for advancement can be met through creating and linking to online resources. Demonstration of value can be met by providing user information back to users. While the system itself cannot generate rewards as stated, the organizations, user communities, and system administrators are all potential sources of such rewards. Rewards would be most valuable in motivating actions outside the normal scope of the system use, such as using rewards to motivate users to gather data for a specific under-researched area.

Factor 5 has the most potential. As social networking becomes more common, the potential to unite users in a common cause becomes easier. While it is unknown whether amateur botanists would be drawn to such a community, the potential is high and the risk of developing a simple prototype is low.

[Agarwal et al., 2006]’s vision is clearly of experts talking to experts. However, the same social network mechanisms that could enable novice participation could be used to enable communication between experts, albeit in more restricted and enclosed areas within the system. Additionally novices could have the potential to become experts themselves and then participate in these communications as experts.

An interesting consequence of having a digital repository of the taxonomic information is that the system can know what information it lacks. This would be useful for both novices and professions gathering data. Such a system could alert users when a digital specimen for a particular species is lacking. Rules for the collection of images for a digital specimen varies by the type of specimen – a cactus requires different parts documented than a tree. Thus an intelligent system could prompt the user for what characteristics should be photographed to get a complete digital specimen.

Digital specimens also lend themselves to community participation. It is illegal to collect flora from federal parks; however, the Jepson Manual will not accept corrections to species distributions without a voucher specimen. This puts novices without special permissions at a disadvantage when collecting scientific data. Allowing digital specimens as vouchers would allow a much greater number of participants to revise distribution data.

**Conclusions**

**System Goals**

Through examination of the system requirements, we have discerned several high-level goals for an electronic field guide.

---

1. **Recognition**: ensuring credits for contributors.
2. **Opportunities for advancement**: enhancing the knowledge and skills of contributors by providing information and training.
3. **Demonstration of value**: demonstrating the value of individual data contribution is important. Computational systems can instantly compile new data and show the most up-to-date summaries.
4. **Rewards**: many contributors of earlier collaborative projects in India have been motivated by rewards such as free field guides (e.g. the Asian Wetland Bureau - Mid-winter waterfowl participants received reports and the complimentary field guides).
5. **Opportunities for social interaction and networking**: the possibilities of finding and interacting with contributors in the geographical vicinity.
**Flexibility**

The technologies used for identification are evolving rapidly. Any developer creating more than just a prototype for an identification method must seriously consider how their system will incorporate new technologies. Identification is likely to be the most evolving element, but there are concerns in other areas as well. Databases with large amounts of information on plants are being built or already exist. Future systems may choose to rely on outside data sources and focus development on the user-oriented features rather than the backend. Such development would be helped by standards in the way botanical information is stored and accessed. Standards such as [Bashauf and Kirchoff, 2008]’s digital specimens should be encouraged and integrated into the next generations of systems.

**Getting Results**

While papers taut advances in the field, most systems with advanced technologies are still not available to the public. Botanists and amateurs are still using paper field guides. Existing electron field guides have limited areas of use or low completeness. The field is ripe for new implementations especially those that have good coverage comparable to that of local guides. Only guides with this amount of coverage will be able to be compared analytically to paper field guides. Users of electronic field guides also need concise and accurate corroborating information to make the most of an identification instance.

[Pickering, 2009] acknowledges the need for users to participate in the creation of field guides. However their system is not built around user involvement. Individuals must write and maintain whole guides; guides cannot be merged and cannot be edited by individuals besides their author. Because of the lack of completeness their system is unusable outside their geographic area. The time needed for a single author to develop a countrywide key is prohibitive. None of the systems discussed have user involvement at there very heart. It is unlikely that large user communities will develop in systems that are not designed to explicitly engage them. Improvements in social networking technologies should be built into new systems if they wish to engage these communities.

**Future Work**

Advancements in photo identification are likely to continue. The kind of technology needed for identification is applicable to other computer vision problems.

Progress to create a system that engages novices would be helpful to test whether novice participation is a viable solution to the needs to large scale data collection. Novices can not only be helpful in providing auxiliary data but even primary taxonomic data. All systems will need to be populated with primary data and that data will need to be maintained. Enlisting novices to do as much of this work as possible may not only be cost effective but have other benefits as well. Allowing multiple users to collaborate together can add quality and ensure that no single individual will be burdened beyond their limits.

Most importantly, new systems that are available to users and have verifiable sources for their data are necessary to test requirements of identification systems used in the field. Technologies for these systems already exist. They should be implemented immediately with current technologies. They should be designed to be upgradable with different identification techniques. Such platforms will be critical to testing advancements in identification side by side.

**References**

[Agarwal et al., 2006] Agarwal, Gaurav; Belhumeur, Peter; Feiner, Steven; Jacobs, David; Kress, W. John; Ramamoorthi, Ravi; Bourg, Norman A.; Dixit, Nandan; Ling, Haibin; Mahajan, Dhruv; Russell, Rusty; Shirdhonkar, Sameer; Sunkavalli, Kalyan; White, Sean. First steps toward an electronic field guide for plants. *Taxon*, vol. 55, no. 3, pp. 597-610, August 2006.


[Bisby et al., 2009] Bisby, Frank; Scoble, Malcolm; MacLeod, Norman; Miles, Colin; Read, Amanda; Berendsohn, Walter; Costello, Mark; Orrell, Thomas. UK Roadmap for delivery of Internet-based Taxonomy. *Proceedings of TDWG 2009 Provisional*, p. 3, 2009.


[Ivanova et al., 2009] Ivanova, Natalia V.; Borisenko, Alex V.;
Herbert, Paul D. N.; “Express barcodes: racing from specimen to identification,” Molecular Ecology Resources, vol 9, no s1, pp. 35-41, April 2009


[Packer et al., 2009] Packer, Laurence; Gibbs, Jason; Sheffield, Cory; Hanner, Robert. DNA barcoding and the mediocrity of morphology. Molecular Ecology Resources, vol 9, no s1, pp. 42-50, April 2009


